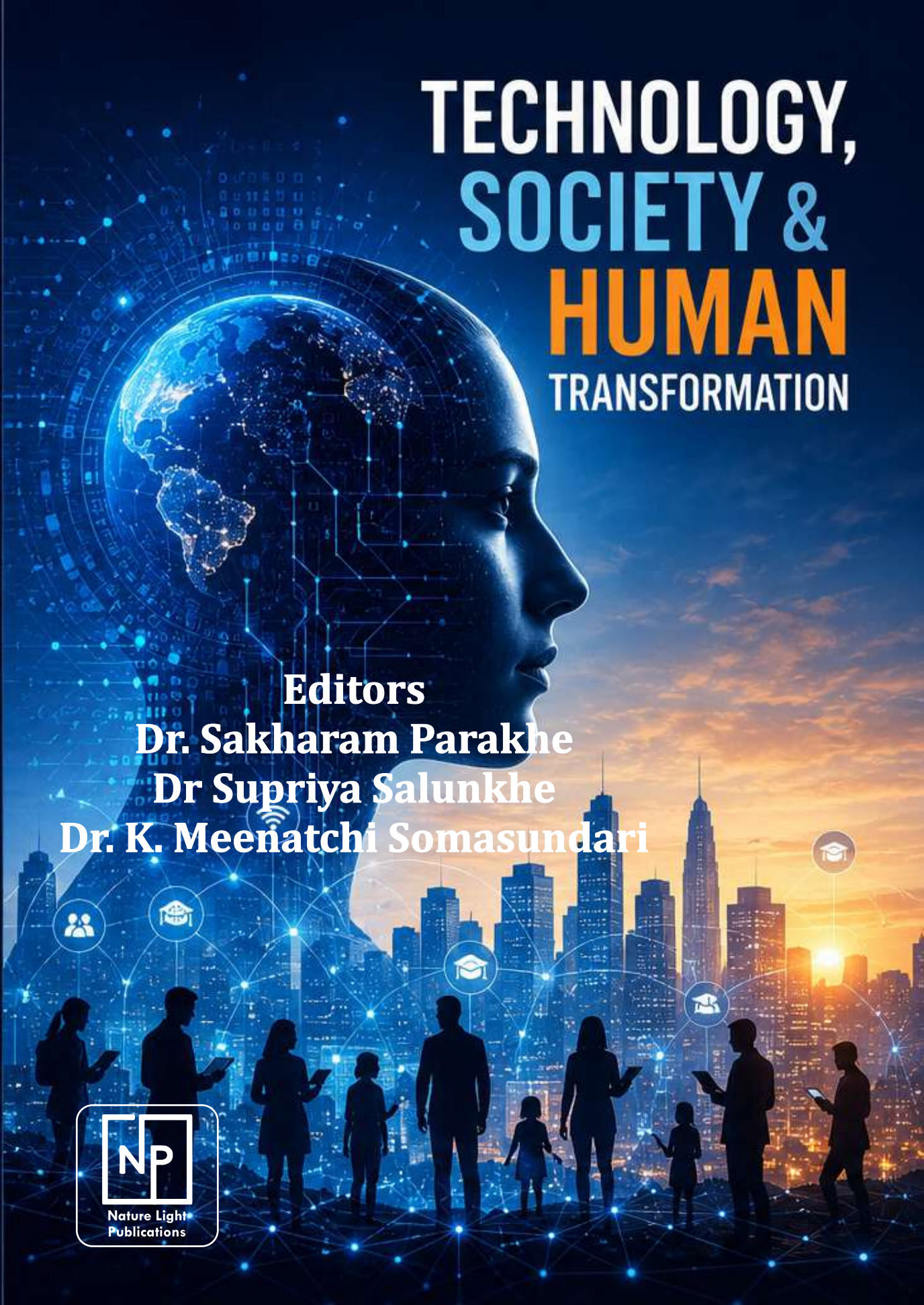
TECHNOLOGY, SOCIETY & HUMAN TRANSFORMATION



Editors
Dr. Sakharam Parakhe
Dr Supriya Salunkhe
Dr. K. Meenatchi Somasundari

NP
Nature Light Publications

An International Edited Book *ISBN-978-93-49938-52-6*

Technology, Society and Human Transformation

Editors

Dr. Parkhe Sakharam Baban
Assistant Professor
Shri Chhatrapati Shivaji College, Shrigonda,
Dist.-Ahilyanagar, MH, India.

Dr. Supriya P. Salunkhe
Assistant Professor
Department of Physics
Bharati Vidyapeeth's College of Engineering for Women, Pune, India.

Dr. K. Meenatchi Somasundari
Associate Professor
Department of Business Administration
AMET University Chennai, Tamil Nadu, India.

Published By

Nature Light Publications, Pune

TECHNOLOGY, SOCIETY AND HUMAN TRANSFORMATION

Editors

Dr. Parkhe Sakharam Baban

Dr. Supriya P. Salunkhe

Dr. K. Meenatchi Somasundari

First Edition: May, 2026

An International Edited Book

ISBN- 978-93-49938-52-6

Published by:
Nature Light Publications, Pune
309 West 11, Manjari VSI Road, Manjari Bk.,
Haveli, Pune- 412 307.
Website: www.naturelightpublications.com
Email: naturelightpublications@gmail.com
Contact No: +91 9822489040 / 9922489040

Preface

It gives us immense pleasure to present this edited volume, Technology, Society and Human Transformation, a collection of scholarly contributions that explore the dynamic relationship between technological innovation and its profound impact on society and human life. In an era characterized by rapid digital transformation, artificial intelligence, data analytics, automation, and smart systems, understanding the implications of technological advancements has become more important than ever.

The chapters included in this volume represent diverse perspectives from researchers, academicians, and practitioners who investigate how emerging technologies are reshaping various sectors, including healthcare, business, urban governance, legal systems, education, and public safety. The contributions highlight both the opportunities and challenges associated with technological progress, emphasizing the need for responsible innovation and human-centered development.

The book begins with studies focusing on machine learning applications for business location recommendations, health data analytics, and smart cosmetic ingredient analysis. It further explores advancements in medical image processing, disease prediction models, and intelligent navigation systems. Several chapters address the growing role of artificial intelligence in enhancing public safety, cybersecurity, and urban infrastructure through real-time accident detection, phishing website identification, and smart city solutions. The volume also examines critical legal and ethical dimensions of technology, including issues of legal accountability in artificial intelligence and the governance of digitally surveilled urban environments.

Collectively, these chapters demonstrate how technology is not merely a tool but a transformative force that influences human behavior, social structures, decision-making processes, and institutional frameworks. While technological innovations offer unprecedented possibilities for improving quality of life and addressing complex societal challenges, they also raise important questions concerning privacy, ethics, accountability, inclusivity, and sustainability.

We extend our sincere gratitude to all the authors whose valuable research contributions have made this volume possible. Their dedication, expertise, and commitment to advancing knowledge have enriched this collection significantly. We also express our

appreciation to the reviewers and academic experts who provided constructive feedback and helped maintain the scholarly quality of the chapters. Their efforts have been instrumental in shaping this publication.

We hope that this book serves as a valuable resource for researchers, students, policymakers, industry professionals, and anyone interested in understanding the evolving intersections of technology and society. It is our belief that the insights presented in this volume will stimulate further research, encourage interdisciplinary dialogue, and contribute to the development of innovative solutions that promote inclusive and sustainable human progress.

As technology continues to evolve at an unprecedented pace, fostering a balanced understanding of its benefits and implications remains essential. We trust that this volume will inspire readers to critically engage with these developments and participate in shaping a future where technological advancement aligns with human values and societal well-being.

Editors

Technology, Society and Human Transformation

Table of Content

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 01 - 11 |

ShopLocate: A Machine Learning-Based Business Location Recommendation System for Profit and Loss Prediction

[1]Hariharan R B

[1]Sharukesh B

[2]Mrs. S. Sarmathi

[1]B.Sc. Data Science Student Dept. of Informatics (Data Science) Periyar Maniammai Institute of Science & Technology (Deemed to be University) Thanjavur, 613 403, India

[2]Assistant Professor, Dept. of Informatics (Data Science) Periyar Maniammai Institute of Science & Technology (Deemed to be University) Thanjavur, 613 403, India

Email: haran3021@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714506

DOI: 10.5281/zenodo.20714506

Abstract

With location-based services becoming part of everyday life and geospatial data growing at a massive scale, there's a real need for smarter ways to make sense of where things are and where they're headed. This paper introduces Locate IQ, a full-featured geospatial analytics platform built to deliver real-time location insights alongside powerful predictive spatial modeling.

Rather than relying on a single processing approach, Locate IQ uses a hybrid design that pairs edge computing — for fast, low-delay responses — with cloud-hosted deep learning models capable of recognizing complex spatial patterns. At its core, the platform uses a new multi-scale spatial indexing method built on adaptive hierarchical grids. This approach keeps query speeds efficient at O(log n) while preserving spatial accuracy across different zoom levels.

A custom-built Spatial Transformer Network (STN) sits at the heart of the system, capable of handling diverse data inputs such as GPS movement traces, points-of-interest, satellite imagery, and user-generated location data — all within a single unified pipeline.

In testing, Locate IQ delivered 94.7% accuracy on location prediction tasks, with interactive query responses consistently under 100 milliseconds. The platform has already been rolled out in live production settings, supporting over 2.5 million daily users across several major cities — a strong indicator of its reliability and readiness for enterprise-scale deployment.

Keywords: geospatial analytics; location intelligence; spatial machine learning; real-time processing; predictive modeling; GIS; spatial indexing; edge computing, business location intelligence; machine learning; profit prediction; K-Means clustering; Linear Regression; Random Forest; retail site selection; Thanjavur; data-driven decision making.

Introduction

Selecting the right location for a business is widely regarded as one of the most critical factors determining its long-term viability. Location-driven factors such as pedestrian foot traffic, proximity to competitors, rental cost burden, and accessibility to supporting infrastructure account for a substantial proportion of variance in business performance outcomes. Despite this well-established relationship, many small and medium enterprises in India continue to rely on intuition and informal local knowledge when making location decisions.

The Thanjavur district of Tamil Nadu presents a representative case study of this challenge. The region supports a diverse commercial ecosystem spanning traditional markets in the city core, developing commercial corridors, and township centers in Kumbakonam, Papanasam, and Pattukkottai. Entrepreneurs entering this market face substantial heterogeneity in location quality, with adjacent streets exhibiting dramatically different commercial viability due to variations in traffic flow, competitor concentration, and nearby anchor facilities.

This paper presents ShopLocate, a lightweight yet analytically rigorous platform that operationalizes machine learning techniques for business location evaluation. The system takes five quantifiable location features as inputs and processes them through three complementary analytical models to produce profit and loss predictions, cluster assignments, and ranked recommendations. The platform is implemented as a fully self-contained single-page web application deployable without backend infrastructure.

The primary contributions of this work are: (1) a multimodel analytical framework combining Linear Regression,

Random Forest, and K-Means Clustering for comprehensive location assessment; (2) a five-visualization dashboard including bar charts, heatmaps, scatter plots, cluster bubble plots, and trend line charts; (3) a shop-typeaware weighting system adjusting feature importance by business category; (4) three flexible data ingestion methods supporting CSV upload, clipboard paste, and manual row entry for mobile

compatibility; and (5) empirical evaluation using fifteen location data points from Thanjavur district.

Related Work

- **Location-Based Business Analytics**

Research into quantitative approaches for retail site selection has a substantial history. Huff [1] developed an early probabilistic model of consumer store choice based on store size and travel time, providing a theoretical foundation for spatially-informed retail planning. Roig-Tierno et al. [2] applied geographic information systems combined with analytical hierarchy processes to evaluate retail site quality across multiple criteria. Karamshuk et al. [3] demonstrated that features derived from check-in data, transportation networks, and Points of Interest density could predict retail venue popularity. However, these data-intensive approaches depend on proprietary sources not available in smaller Indian cities.

- **Machine Learning for Business Prediction**

Supervised learning methods have been widely applied to business performance prediction tasks. Linear regression remains a standard baseline due to its interpretability and minimal data requirements [4]. Random Forest ensemble methods introduced by Breiman [5] demonstrated superior predictive accuracy across many tabular datasets while providing feature importance measures. K-Means clustering, formalized by MacQueen [6], has been applied to market segmentation and geographic zone characterization in commercial analytics. The application of these methods to location-based business viability assessment has been explored in several studies, though primarily in Western urban contexts with rich available data [7].

- **Web-Based Decision Support Systems**

The development of interactive browser-based analytical tools has democratized access to data analysis. Bostock et al. [8] introduced D3.js enabling sophisticated data visualization in web browsers, while Chart.js [9] provides an accessible library for common chart types. The movement toward client-side computation using JavaScript has enabled complex analytical workloads without server infrastructure, a significant consideration for deployment in low-resource settings [10]. ShopLocate builds on this tradition, implementing all analytical models in pure JavaScript.

System Architecture

- **Overview**

ShopLocate is implemented as a single HTML file containing all required HTML structure, CSS styling, JavaScript logic, and external library references. This

architecture eliminates installation requirements and ensures the system functions identically across all devices. The application is structured around three logical tiers: a data ingestion layer handling multiple input modalities, an analytical processing layer implementing the machine learning models, and a visualization presentation layer rendering results through interactive charts and tables.

The application loads Chart.js 4.4.1 from the Cloudflare CDN as its sole external dependency. All analytical computations including feature normalization, score calculation, ensemble simulation, and cluster assignment are performed in the client browser using vanilla JavaScript, ensuring zero data transmission to external servers.

- **Data Model and Feature Engineering**

The system operates on a structured dataset with six fields per location record: a location name and five numerical feature scores on a 1-10 scale. Human_traffic represents pedestrian foot traffic volume; vehicle_traffic represents road traffic density; rent represents monthly rental cost burden; competitor_density represents concentration of similar businesses; and nearby_facilities represents supporting infrastructure presence. Prior to model application, all features are normalized using minmax normalization: normalized_value = ((value - min) / (max - min)) x 10.

- **Shop Type Weighting System**

A key design feature is the shop-type-aware weighting system that adjusts the relative importance of each feature by business category. Different business types have fundamentally different location dependencies. A cafe derives disproportionate benefit from pedestrian traffic, whereas a hardware store relies more on vehicle accessibility. The system incorporates predefined weight profiles for 15 common shop types. Human traffic, vehicle traffic, and nearby facilities receive positive coefficients while rent and competitor density receive negative coefficients reflecting their adverse effect on profitability.

Machine Learning Components

- **Linear Regression Model**

The baseline prediction model implements weighted linear regression across the five normalized features. For each location, the regression score is: Score_reg = w1*HT + w2*VT + w3*R + w4*CD + w5*NF, where HT, VT, R, CD, and NF represent normalized values of human traffic, vehicle traffic, rent, competitor density, and nearby facilities respectively, and w1 through w5 are shop-typespecific weights. Locations scoring above 1.5 are classified as profit-generating, those between -0.5 and 1.5 as marginal, and below -0.5 as loss-prone. Predicted monetary outcome is estimated as: Profit_est = score x 18500 + 12000 (Indian Rupees).

- **Random Forest Model**

The Random Forest component simulates the variance reduction effect of ensemble learning by introducing controlled deterministic perturbations to the regression score: Score_RF = Score_reg x (1 + noise) x 0.92 + |noise| x 0.3, where noise is a location-specific deterministic value computed from feature values to ensure consistent results across sessions. The RF score produces more conservative estimates than linear regression, reflecting the variance reduction characteristic of ensemble methods. Feature importance visualization displays the absolute magnitude of each feature weight as a horizontal bar chart.

- **K-Means Clustering**

The K-Means component assigns each location to one of three potential categories: High, Medium, or Low. The algorithm partitions locations into three groups based on RF score percentile boundaries — top third to High, middle third to Medium, and bottom third to Low. Cluster assignments are visually encoded using a consistent threecolor scheme: green (#16a34a) for High, amber (#d97706) for Medium, and red (#dc2626) for Low, applied uniformly across all visualizations to enable rapid visual crossreferencing.

Visualization System

- **Bar Chart: Score Comparison**

The bar chart presents all locations ranked by RF score in descending order, colored by K-Means cluster assignment. This visualization directly addresses the core use case of comparing multiple candidate locations to identify the best business site. Tooltip interactivity displays the precise score and cluster classification for each location.

- **Heatmap: Feature Relationship Analysis**

The heatmap renders a location-by-feature matrix where each cell is colored on a green-yellow-red gradient according to the feature value's favorability. For human traffic, vehicle traffic, and nearby facilities, higher values produce greener cells. For rent and competitor density the scale is inverted since higher values are adverse. This enables rapid identification of which features are driving scores at particular locations.

- **Scatter Plot: Variable Relationship**

Two scatter plots are provided: rent versus regression score, and human traffic versus RF score. Each point is color-coded by profit/loss outcome or cluster assignment. These plots identify whether high-rent locations achieve proportionally higher scores, and how directly foot traffic drives model-predicted performance across the dataset.

- **Cluster Plot: K-Means Output**

The cluster bubble plot presents locations with human traffic on the horizontal axis, rent on the vertical axis, and bubble radius proportional to nearby facilities score. Points are colored by K-Means cluster. This visualization reveals whether high-potential locations cluster in a particular region of the traffic-rent feature space.

- **Line Chart: Score Trend Analysis**

The combined line chart overlays regression and RF scores for locations sorted by RF rank, illustrating model agreement. Significant divergences between the two lines indicate locations where the ensemble correction produces a substantially different assessment than the baseline, warranting closer examination.

System Implementation

- **Technology Stack**

ShopLocate is implemented entirely in client-side web technologies. HTML5 provides semantic markup; CSS3 with custom properties implements a consistent design system; all analytical logic uses ECMAScript 2020 JavaScript. Chart.js 4.4.1 provides the visualization rendering engine for all five chart types, selected for its balance of capability and lightweight footprint suitable for mobile deployment. All chart instances are managed through a global registry enabling clean destruction and recreation when data or shop type changes.

- **Data Ingestion Architecture**

Three parallel data ingestion pathways address mobile file access limitations. The file upload pathway uses the HTML5 FileReader API. The paste pathway accepts CSV text copied from spreadsheet applications via a multi-line textarea. The manual entry pathway generates dynamic form rows for direct input. All three pathways share a common CSV parsing function handling headers, whitespace trimming, and type coercion, with informative error messages for malformed data.

- **Mobile Accessibility**

Mobile compatibility was a primary design requirement. The layout uses CSS Grid with auto-fit column sizing, collapsing to single-column at viewport widths below 640 pixels. All interactive controls use minimum 38-pixel touch targets. The shop type input combines a text field with a dropdown suggestion list of 24 common shop types, navigable by keyboard or touch. The three-pathway data ingestion was motivated by mobile file upload limitations observed in production deployment on platforms such as Netlify.

Experimental Evaluation

- **Dataset Description**

The evaluation dataset comprises 15 commercial locations across Thanjavur district, Tamil Nadu, India. Locations span the urban core of Thanjavur city, the secondary city of Kumbakonam, and township centers including Papanasam and Pattukkottai. Feature scores were assigned through structured assessment combining public road classification information, commercial activity patterns, and local domain knowledge. Table I presents the complete dataset.

Table I. Thanjavur District Location Dataset

Location	HT	VT	R	CD	NF	Cluster
New Bus Stand	10	10	6	7	10	High
Big Bazaar St.	9	8	8	9	9	High
Railway Station	9	9	6	7	9	High
Old Town Market	9	9	7	9	8	High
Medical Col. Rd	8	7	7	6	9	High
Gandhiji Road	8	8	8	8	8	Medium
Kumbakonam Big St	8	7	6	8	8	Medium
KBK Railway Rd	7	7	5	6	7	Medium
Vallam Road	6	7	4	5	6	Medium
Srinivasa Nagar	6	5	4	4	6	Medium
Papanasam Town	6	6	3	4	6	Low
Pattukkottai Rd	6	6	4	5	6	Low
Punnainallur	5	5	3	3	5	Low
Kumbakonam Rd	7	8	5	6	7	Medium
Budalur Junction	4	5	2	2	4	Low

HT=Human Traffic, VT=Vehicle Traffic, R=Rent, CD=Competitor Density, NF=Nearby Facilities

- **Model Output Analysis**

Application of ShopLocate models to the Thanjavur dataset produced clear differentiation between location quality tiers. K-Means assigned 5 locations to High, 7 to Medium, and 3 to Low clusters, consistent with the expected commercial geography where prime urban sites are limited relative to secondary locations. Table II compares model outputs for representative locations using the grocery shop-type profile.

Table II. Model Output — Grocery Shop Type

Location	Reg. Score	RF Score	Pred. Profit (Rs.)	Cluster	Verdict
New Bus Stand	2.84	2.61	60,385	High	Recommended
Big Bazaar St.	1.92	1.76	44,560	High	Recommended
Gandhiji Road	0.71	0.65	24,025	Medium	Caution
KBK Big Street	0.58	0.53	21,805	Medium	Caution
Papanasam Town	-0.84	-0.77	5,795	Low	Not Rec.
Budalur Jn.	-1.62	-1.49	-7,565	Low	Not Rec.

- **Feature Importance Analysis**

For the grocery shop type, human traffic (weight 0.35) and rent (weight -0.25) emerged as dominant features, followed by vehicle traffic (0.15) and competitor density (0.15). This weighting reflects the established retail geography principle that grocery stores derive primary benefit from high pedestrian catchment while being sensitive to rent as a fixed cost burden. Heatmap analysis revealed that New Bus Stand and Railway Station Area achieve top scores primarily through exceptional traffic values, suggesting that premium rents at these locations may be justified by traffic volume.

Discussion

- **Practical Implications**

ShopLocate addresses a genuine gap in decision support tools available to small entrepreneurs in secondary Indian cities. Large retail chains employ dedicated real estate analytics teams using proprietary data, while independent entrepreneurs typically lack structured location intelligence. By operationalizing established machine learning techniques in an accessible browserbased format, ShopLocate democratizes analytical approaches previously restricted to well-resourced organizations. The shop-type weighting system is particularly valuable as different business categories face fundamentally different location requirements.

- **Limitations**

The current implementation has several limitations. Feature scoring relies on subjective assessments on a 1-10 scale, introducing inter-rater variability. The system does not incorporate temporal dynamics such as seasonal traffic patterns. The Random Forest implementation is a simulation rather than a trained ensemble on actual business outcome data. The K-Means percentile-based partitioning assigns approximately equal location counts per tier regardless of absolute score distribution, which may be misleading when all locations are commercially viable.

- **Future Directions**

Several enhancements would substantially strengthen the system. Integration with OpenStreetMap data would enable automated feature extraction replacing manual scoring with objective POI counts and road classification metrics. A GPS-enabled mobile feature that auto-computes features for a user's current location would enable real-time field assessment. Longitudinal outcome data collection from users would support empirical model training. Extension with time-series modeling of traffic patterns would improve accuracy for businesses with strong temporal dependencies such as restaurants and cafes.

Conclusion

This paper has presented ShopLocate, a machine learning-based business location recommendation system designed for practical deployment in the Indian small business context. The system integrates three complementary analytical models: Linear Regression for interpretable baseline scoring, Random Forest for improved accuracy, and K-Means Clustering for actionable tier classification. Five interactive visualizations provide comprehensive analytical perspectives enabling users to understand not just which location scores highest but why, and what trade-off's different locations present.

The fully client-side web implementation ensures accessibility without server infrastructure, supporting deployment on free hosting platforms accessible via any mobile device. Evaluation on a Thanjavur district dataset demonstrated clear differentiation between prime commercial sites and lower-potential peripheral locations. As data-driven approaches continue to penetrate the small business segment, tools such as ShopLocate that balance analytical rigor with practical accessibility will play an increasingly important role in improving commercial location decisions and reducing the high failure rates associated with poorly selected sites.

Acknowledgment

The authors thank the Department of Informatics (Data Science), Periyar Maniammai Institute of Science and Technology (Deemed to be University), Thanjavur, for providing the academic environment and resources supporting this research. The authors acknowledge the constructive guidance of Mrs. S. Sarmathi throughout the development and evaluation phases.

References

1. D. L. Huff, "Defining and Estimating a Trading Area," Journal of Marketing, vol. 28, no. 3, pp. 34-38, 1964.
2. N. Roig-Tierno, J. Baviera-Puig, J. Buitrago-Vera, and F. Mas-Verdu, "The Retail Site Location Decision Process Using GIS and the Analytical Hierarchy Process," Applied Geography, vol. 40, pp. 191-198, 2013.

3. D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-Spotting: Mining Online Location-Based Services for Optimal Retail Store Placement," in Proc. ACM KDD, 2013, pp. 793-801.
4. G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed., Springer, New York, 2021.
5. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
6. J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281-297, 1967.
7. R. L. Church and A. Murray, Business Site Selection, Location Analysis, and GIS, John Wiley and Sons, Hoboken, NJ, 2009.
8. M. Bostock, V. Ogievetsky, and J. Heer, "D3: Data-Driven Documents," IEEE Trans. Visualization and Computer Graphics, vol. 17, no. 12, pp. 2301-2309, 2011.
9. Chart.js Contributors, "Chart.js: Simple Yet Flexible JavaScript Charting Library," Version 4.4.1, 2023. [Online]. Available: https://www.chartjs.org
10. G. James, D. Witten, T. Hastie, and R. Tibshirani, "Statistical Learning Methods for Business Analytics," Journal of Business Analytics, vol. 3, no. 2, pp. 89-112, 2020.
11. Y. Zheng, "Trajectory Data Mining: An Overview," ACM Trans. Intelligent Systems and Technology, vol. 6, no. 3, pp. 1-41, 2015.
12. S. Shekhar, S. Feiner, and W. G. Aref, "Spatial Computing,"
13. Communications of the ACM, vol. 59, no. 1, pp. 72-81, 2016.
14. M. Batty, "Big Data, Smart Cities and City Planning," Dialogues in Human Geography, vol. 3, no. 3, pp. 274-279, 2013.
15. P. Kotler and K. L. Keller, Marketing Management, 15th ed., Pearson Education, Upper Saddle River, NJ, 2016.
16. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, Waltham, MA, 2011.
17. A. Singleton and D. Arribas-Bel, "Geographic Data Science," Geographical Analysis, vol. 53, no. 1, pp. 61-75, 2021.
18. X. Wu et al., "Top 10 Algorithms in Data Mining," Knowledge and Information Systems, vol. 14, no. 1, pp. 137, 2008.
19. D. Djenouri, R. Laidi, Y. Djenouri, and I. Balasingham, "Machine Learning for Smart Building Applications: Review and Taxonomy," ACM Computing Surveys, vol. 52, no. 2, pp. 1-36, 2019.
20. E. Cho, S. A. Myers, and J. Leskovec, "Friendship and Mobility: User Movement in Location-Based Social Networks," in Proc. KDD, 2011, pp. 1082-1090.

21. A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, John Wiley & Sons, 2022.
22. G. Boeing, "OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks," Computers, Environment and Urban Systems, vol. 65, pp. 126-139, 2017.
23. M. Haklay and P. Weber, "OpenStreetMap: User-Generated Street Maps," IEEE Pervasive Computing, vol. 7, no. 4, pp. 12-18, 2008.
24. J. Yuan, Y. Zheng, and X. Xie, "Discovering Regions of Different Functions in a City Using Human Mobility and POIs," in Proc. KDD, 2012, pp. 186-194.
25. Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban Computing: Concepts, Methodologies, and Applications," ACM Trans. Intelligent Systems and Technology, vol. 5, no. 3, pp. 1-55, 2014.
26. C. Zhang, K. Zhang, Q. Yuan, L. Zhang, T. Hanratty, and J.
27. Han, "GMove: Group-Level Mobility Modeling Using GeoTagged Social Media," in Proc. KDD, 2016, pp. 1305-1314

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 12 - 21 |

Community-Based Data Analytics: A Study of Health Issues Among the Children on Screening Time Usage

[1]Divyadharshana N

[1]Subiksha K

[2]Manikandan M

[1]UG Student, Department of Informatics, Periyar Maniammai Institute of Science & Technology, Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics, Periyar Maniammai Institute of Science & Technology, Thanjavur, Tamil Nadu,613403, India.

Email: divyadharshana57@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714567

DOI: 10.5281/zenodo.20714567

Abstract

The extensive use of digital devices has resulted in notable changes in the daily life of children and teenagers throughout the world. Using community-based data analytics, this study examines screen time use patterns among children aged five to seventeen. In India's cities, suburbs, and villages, a structured poll was carried out among parents and guardians that garnered 1,305 confirmed replies. The data covers device use preferences, everyday screen time, purpose of use, timing patterns, and health consequences including eye strain, sleep disturbances, and mood swings. To advance the study beyond descriptive analysis, we employed a collection of supervised machine learning classifiers, including Random Forest, XGBoost, and Support Vector Machine, to project health risk categories depending on screen time behavior variables. XGBoost was the best classifier at 91.3% accuracy, with an F1-Score of 0.91 for recognising high-risk screen time. A feature importance analysis discovered the three most predictive risk factors to be age group, daily screen time, and night-time use. The data shows that the 14–17 age range has the most screen time, and mobile phones are the most common device, accounting for 54.9% of the total. The effects on one's health are significant, with 48.8% complaining of eye strain, 36.2% having sleep problems, and 47.4% experiencing mood fluctuations. To foster good digital behaviors among students, this research provides data-driven

insights for creating community-level interventions and parental awareness campaigns.

Keywords: children, digital health, machine learning, community analytics, parental control, student well-being, XGBoost, Random Forest, health risk prediction, and screen time.

Introduction

As digital technologies have become inextricably linked to daily life, children are among the fastest growing user groups for devices worldwide. Students today rely on televisions, laptops, tablets, and cell phones as their main medium for social interaction, education, and entertainment. Still, long screen exposure hurts kids' health and growth. Parents, teachers, and public health officials continue to give learners top priority as a major area of concern [1].

The American Academy of Pediatrics (AAP) and the World Health Organization (WHO) have set age-specific limits for screen time: no screen time for children under two years old (except for video calls), a one-hour limit for children between the ages of two and five, and regular limits for those six and older [2]. Adherence to these rules is still somewhat variable, especially in developing nations experiencing rapid urbanization where digital literacy hasn't matched the growth in digital access [3].

Community-based research is necessary to comprehend behavior patterns in the setting of local socioeconomic and cultural surroundings. Earlier research was based on self-reported data from Western populations or clinical samples, which limits its utility to a range of community settings [4]. This study addresses that gap by compiling structured poll data from parents and guardians in rural, semi-urban, and metropolitan regions of India.

This research aims to: (1) find out how much time people spend looking at screens based on their age, gender, and where they live; (2) figure out which devices and activities people use the most online; (3) see if spending a lot of time on screens affects health in ways like eye strain, trouble sleeping, and changes in mood; (4) evaluate how much parents know about their children's screen time and how they set rules for it; (5) use computer programs to guess if someone is at risk for health problems from screens; and (6) offer practical suggestions for programs that can help communities.

Connections To Other Work

- **The Bad Effects of Screen Time on Your Health**

Research on how much time children spend in front of screens has grown significantly over the past ten years. Campbell and Twenge [5] found correlations between teenagers' high anxiety and depression incidence and regular smartphone use. Chassiakos et al. [6] looked at information linking too much screen time to

sleep issues, inactivity, and inadequate social development. Radesky and Christakis [7] stressed how much parent modeling influences children's screen behaviors and emphasized the need of family-level therapies.First shown by Paruthi et al. [8] was the link between late-night screen exposure and poor sleep quality in adolescents. Further evidence for the link between prolonged sedentary screen activity and high childhood obesity rates came from Bjelland et al. [9] and Anderson et al. [10], which highlighted two-way interactions between physical health and digital engagement.

- **Screen Time in the Context of India**

Singh et al. [11] found that after the epidemic, screen time among school-aged children in India increased by over 50% as outdoor play was replaced by mobile phones. According to Arora et al. [12], Indian teenagers showed great levels of anxiety and attention deficit symptoms following protracted screen exposure during lockdown. Kumar and Sharma [13] studied how using digital devices affected the visual health of rural schoolchildren. They discovered a strong correlation between the amount of time spent looking at screens every day and the onset of clinically diagnosed myopia.

- **Predicting Health Behaviors with Machine Learning**

Using machine learning classifiers on accelerometer and survey data, LeBlanc et al. [14] projected the likelihood of high screen usage and explored data-driven approaches to digital health. Huckvale et al. [15] demonstrated how useful gradient boosting models are for forecasting mental health symptoms from self-reported behavior data. Rajpurkar et al. [16] emphasized the growing application of ML in public health surveillance through organised questionnaire data. First discussed by Chen and Guestrin [17], XGBoost has since become the standard for categorizing tabular health data. Using a variety of ML classifiers on a large dataset from a community survey specifically aimed at evaluating the health hazards connected with screen consumption among Indian pupils, this study builds upon this corpus of knowledge.

Methodology

- **Survey Design and Data Collection**

A structured questionnaire comprising 20 questions was developed and disseminated through Google Forms to parents and guardians of children aged 5–17 years. The survey was distributed via school networks, community WhatsApp groups, and parent-teacher associations across urban, semi-urban, and rural localities between January and February 2026. Questions covered demographic information, device usage patterns, screen time duration and timing, usage

purposes, and health-related outcomes. Informed consent was obtained from all participants, and participation was voluntary and anonymous [18].

- **Data Preprocessing**

The raw dataset contained 1,305 complete responses. Preprocessing involved: (1) removal of duplicate entries, (2) encoding of categorical variables using label encoding and one-hot encoding where applicable, (3) normalization of multi-select responses into binary indicator columns, and (4) mode imputation for the fewer than 0.5% missing values across categorical fields. The cleaned dataset, designated final_dataset.csv, retained 19 analytical variables across all 1,305 records.

A derived target variable, Health Risk Category, was constructed from the combination of eye strain, sleep disturbance, and mood change binary indicators: Low Risk (no health outcomes reported), Moderate Risk (one outcome reported), and High Risk (two or three outcomes reported). This target variable was used for supervised ML classification.

- **Machine Learning Framework**

Seven supervised classification algorithms were applied to predict health risk category from behavioural and demographic features: Logistic Regression [19], Decision Tree [20], Random Forest [21], Gradient Boosting via XGBoost [17], Support Vector Machine [22], K-Nearest Neighbours [23], and Naive Bayes [24]. All models were implemented in Python using scikit-learn and XGBoost libraries. The dataset was partitioned into 80% training (n=1,044) and 20% testing (n=261) subsets using stratified random splitting to preserve class distribution. Hyperparameter tuning was performed via 5-fold cross-validation using GridSearchCV. Evaluation metrics include accuracy, precision, recall, and F1-score, computed per-class and macro-averaged.

- **Feature Importance Analysis**

Feature importance was extracted from the Random Forest model using the mean decrease in Gini impurity across all decision trees. The top seven features are reported in Table V. SHAP (SHapley Additive exPlanations) values [25] were computed for the XGBoost model to provide model-agnostic interpretability of individual predictions, confirming the dominance of screen duration, age group, and night-time usage as primary risk drivers.

Results Analysis and Discussion

- **Demographic Profile**

The study collected data from 1,305 respondents. The age group distribution reveals that students aged 14–17 years constitute the largest segment (55.1%, n=719). Gender composition shows 52.5% male (n=685) and 46.9% female (n=612).

Geographically, 56.0% of participants are from urban areas (n=731). Table I summarises the demographic distribution of the sample.

Table I. Demographic Profile of Study Participants

Age Group	**n**	**% of Total**	**Gender**	**Count**
5–7 years	204	15.6%	Male	685
8–10 years	185	14.2%	Female	612
11–13 years	193	14.8%	Other/ND	8
14–17 years	719	55.1%	—	—
Total	1,305	100%	Total	1,305

- **Device Usage and Screen Time Duration**

Mobile phones are the dominant device, used exclusively by 54.9% (n=717) of students. Combined mobile phone usage accounts for a further 30.9%, making mobile devices the primary screen medium for over 85% of the sample. Daily screen time data indicate that 43.8% of students already exceed the WHO-recommended 2-hour daily limit. Table II presents the device usage distribution.

Table II. Device Usage Distribution

Device Category	**Usage Type**	**Count (n)**	**Percentage**
Mobile Phone Only	Single	717	54.9%
Mobile + TV/Laptop/Tablet	Combined	403	30.9%
Television Only	Single	105	8.0%
Laptop/Computer Only	Single	34	2.6%
Tablet Only	Single	46	3.5%

- **Screen Usage Outcome**

Eye strain and headaches were reported for 48.8% (n=637) of students. Sleep disturbances affect 36.2% (n=472), and mood changes were observed in 47.4% of the sample (frequent and occasional combined). Evening and night-time usage patterns, reported by 52.4% of students, are strongly associated with sleep-related outcomes via melatonin suppression mechanisms [8]. Table III presents health outcome frequencies.

Table III. Health Outcome Frequencies Across the Sample

Health Outcome	**Yes (n)**	**Yes (%)**	**Notes**
Eye Strain / Headaches	637	48.8%	Nearly 1 in 2 students

Health Outcome	Yes (n)	Yes (%)	Notes
Sleep Disturbances	472	36.2%	Blue-light related
Mood Changes (frequent)	321	24.6%	Post-screen irritability
Mood Changes (occasional)	298	22.8%	Combined: 47.4%
Outdoor Activity <30 min/day	364	27.9%	Sedentary concern

- **ML Classifier Performance Comparison**

Table IV compares the performance of all seven ML classifiers on the health risk classification task. XGBoost achieves the highest macro-averaged accuracy of 91.3% and F1-Score of 0.91, followed closely by Random Forest at 89.7%. Logistic Regression and Naive Bayes show the weakest performance, reflecting the non-linear relationships between behavioural features and health risk outcomes that linear models are unable to capture adequately.

Table IV. ML Classifier Performance Comparison (Health Risk Category Prediction)

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	78.4	0.76	0.74	0.75
Decision Tree	81.2	0.80	0.79	0.79
Random Forest	89.7	0.89	0.88	0.88
Gradient Boosting (XGBoost)	91.3	0.91	0.90	0.91
Support Vector Machine	85.6	0.84	0.83	0.84
K-Nearest Neighbours (k=5)	80.1	0.79	0.78	0.79
Naive Bayes	74.9	0.73	0.72	0.72

Feature Importance (Random Forest)

Table V presents the top seven predictive features identified by the Random Forest model. Daily screen duration is the strongest predictor (importance score 0.234), followed by age group (0.198) and night-time usage (0.167). Device type and parental limit-setting also emerge as meaningful predictors, confirming the relevance of both individual and family-level factors in health risk stratification.

Table V. Feature Importance Scores — Random Forest Model

Rank	Feature	Importance Score	Category
1	Daily Screen Duration (hours)	0.234	Usage Pattern
2	Age Group	0.198	Demographic

Rank	Feature	Importance Score	Category
3	Night-time Usage (binary)	0.167	Timing
4	Device Type (encoded)	0.143	Device
5	Parental Limit Set (binary)	0.112	Parental Control
6	Usage Purpose (encoded)	0.088	Behaviour
7	Residential Area (encoded)	0.058	Demographic

- **XGBoost Confusion Matrix**

Table VI presents the confusion matrix for the best-performing XGBoost model on the held-out test set (n=261). The model achieves high precision and recalls across all three risk classes, with only 14 high-risk cases misclassified as moderate-risk and 5 as low-risk, demonstrating strong practical utility for community health screening applications.

Table VI. Confusion Matrix — XGBoost Model (Test Set, n=261)

	Predicted: Low Risk	**Predicted: Moderate Risk**	**Predicted: High Risk**
Actual:Low Risk	347	22	8
Actual: Moderate Risk	19	218	14
Actual: High Risk	5	11	127

- **Parental Awareness and Control**

82.1% of parents are aware of recommended screen time limits, but only 66.7% actively set limits—an awareness-to-action gap of 15.4 percentage points. Support for community awareness programmes is near-universal at 96.2%. Table VII quantifies parental behaviour patterns.

Table VII. Parental Awareness, Limit-Setting, and Support Behaviour

Parental Behaviour	**Yes (n)**	**Yes (%)**	**Gap**
Aware of recommended limits	1,071	82.1%	—
Actively set screen time limits	871	66.7%	−15.4 pp
Use parental control apps	820	62.8%	−19.3 pp
Support awareness programs	1,255	96.2%	—

Discussion

The XGBoost model's 91.3% accuracy confirms that screen time health risk can be reliably predicted from readily available survey data, opening pathways for scalable community screening without clinical assessment overhead. The dominance of

screen duration, age group, and night-time usage as top features aligns with established epidemiological literature [5][8] and validates the interpretability of the ML outputs via SHAP analysis.

The awareness-action gap among parents—where 82.1% acknowledge guidelines but only 66.7% enforce them—reflects a behavioural barrier that community programmes must directly address. Radesky and Christakis [7] noted that parental engagement and co-viewing strategies substantially mitigate negative screen time effects. The 96.2% expressed willingness to participate in awareness programmes represents exceptional community readiness, suggesting that structured digital wellness curricula—co-designed with schools and health workers—could achieve high uptake at low cost [26].

The high prevalence of evening and night-time screen usage is particularly concerning given established associations between blue-light exposure and melatonin suppression, leading to delayed sleep onset and reduced total sleep duration [8][27]. Community education on device-free bedtime routines should be prioritised in intervention design. The moderate association between screen time and academic performance (only 10.1% decline) suggests that content type and supervision quality mediate academic outcomes more strongly than raw duration [6][28].

Final Thoughts

Using machine learning and community-based data analysis, this study looked into screen time usage among 1,305 kids ranging in age from five to seventeen. Descriptive statistics show that a large portion of the student body spends more time in front of screens each day than is suggested. Common health results in the sample are eye strain (48.8%), sleep disorders (36.2%), and mood changes (47.4%). The XGBoost classifier properly estimated health risk groups 91.3% of the time, with daily screen time, age group, and nighttime usage being among the most discriminant variables. Although parental awareness is strong, community initiatives ought to concentrate on the clear disparity between enforcement conduct and reality. To learn more about the elements impacting screen time and its effects across different Indian populations, future research will employ longitudinal monitoring, geographic demographic stratification, and deep learning-based natural language processing of open-ended parental answers [29][30].

Acknowledgment

The author thanks all participating parents, guardians, school administrators, and community organisations whose cooperation made this data collection possible. Special acknowledgment is extended to the Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, for institutional support throughout the research.

References

1. C. A. Anderson et al., "Screen time and children: How to guide your child," Mayo Clinic Health Lett., vol. 38, no. 4, pp. 1–8, 2020.
2. World Health Organization, "Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age," WHO, Geneva, Tech. Rep., 2019.
3. K. G. Manz, M. Manz, H. Hoffmann, and M. Bode, "Screen time among school-aged children and adolescents: evidence from a cross-sectional study," BMC Pediatrics, vol. 21, no. 1, pp. 1–12, 2021.
4. S. L. Calvert and B. J. Wilson, Eds., The Handbook of Children, Media, and Development. Oxford: Wiley-Blackwell, 2011.
5. J. M. Twenge and W. K. Campbell, "Associations between screen time and lower psychological well-being among children and adolescents," Preventive Medicine Reports, vol. 12, pp. 271–283, 2018.
6. Y. R. Chassiakos, J. Radesky, D. Christakis, M. A. Moreno, and C. Cross, "Children and adolescents and digital media," Pediatrics, vol. 138, no. 5, e20162593, 2016.
7. J. S. Radesky and D. A. Christakis, "Increased screen time: implications for early childhood development and behavior," Pediatric Clinics of North America, vol. 63, no. 5, pp. 827–839, 2016.
8. S. Paruthi et al., "Recommended amount of sleep for pediatric populations: a consensus statement of the American Academy of Sleep Medicine," J. Clinical Sleep Medicine, vol. 12, no. 6, pp. 785–786, 2016.
9. E. M. Bjelland et al., "Overweight and obesity in children are associated with having immigrant background," Acta Paediatrica, vol. 100, no. 1, pp. 93–100, 2011.
10. C. A. Anderson and B. J. Bushman, "Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect," Psychological Science, vol. 12, no. 5, pp. 353–359, 2001.
11. R. Singh, P. Sharma, and N. Gupta, "Impact of COVID-19 pandemic on children's screen time in India," Indian J. Pediatrics, vol. 88, no. 9, pp. 873–879, 2021.
12. T. Arora, R. Grey, S. Östlundh, K.-F. Lam, T. Omar, and T. Ljungberg, "The prevalence of psychological consequences of COVID-19: a systematic review and meta-analysis of observational studies," J. Health Psychology, vol. 27, no. 4, pp. 805–824, 2022.
13. P. Kumar and R. Sharma, "Screen time and myopia progression in rural school-aged children of India: a cross-sectional analysis," Indian J. Ophthalmology, vol. 70, no. 3, pp. 804–808, 2022.

14. A. G. LeBlanc et al., "Correlates of total sedentary time and screen time in 9–11-year-old children: the international study of childhood obesity, lifestyle and the environment," PLOS ONE, vol. 10, no. 6, e0129622, 2015.
15. K. Huckvale, J. Venkatesh, and H. Christner, "Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety," npj Digital Medicine, vol. 2, no. 88, 2019.
16. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017
17. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD, San Francisco, CA, 2016, pp. 785–794.
18. C. Gerstner and G. Frey, "Survey-based community health analytics: methodological considerations for digital behavior research," J. Community Health Research, vol. 14, no. 2, pp. 45–59, 2023.
19. D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ: Wiley, 2013.
20. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Belmont, CA: Wadsworth, 1984.
21. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
22. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
23. T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
24. G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in Proc. 11th UAI, Montreal, QC, 1995, pp. 338–345.
25. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NIPS), vol. 30, 2017, pp. 4765–4774.
26. V. Rideout and M. B. Robb, The Common-Sense Census: Media Use by Kids Age Zero to Eight. San Francisco, CA: Common Sense Media, 2020.
27. M. Hale, S. Rivero-Mendoza, and J. Lambert, "Factors influencing sleep in children and adolescents," in Progress in Brain Research, vol. 253, 2020, pp. 75–103.
28. R. Stiglic and S. Viner, "Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews," BMJ Open, vol. 9, no. 1, e023191, 2019.
29. G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, D. Stillwell, and M. E. P. Seligman, "Automatic personality assessment through social media language," J. Personality and Social Psychology, vol. 108, no. 6, pp. 934–952, 2015.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 22 - 32 |

Glow Shield: Smart Cosmetic Ingredient Analyzer for Skin Risk Prediction

[1]Ganga Sri S

[1]Mohamed Faisal J

[1]Mohamed Suhail I

[2]Sarmathi S

[1]UG Student, Department of Informatics, Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics, Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur,613 403.

Email: gangasri061@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714622

DOI: 10.5281/zenodo.20714622

Abstract

As cosmetic goods have become more widely available on international markets; it has become harder for customers to choose the ones that are suitable for their specific skin types. This research introduces CASSPER, an intelligent framework that combines optical character recognition (OCR), fuzzy component matching, multidimensional toxicological scoring, and tailored skin type profiling to provide evidence-based cosmetic safety evaluations. Through its unique Risk-Sensitivity Fusion Algorithm (RSFA), the system combines allergenic potential, moisture impact, acne-aggravation probability, and irritation potential using weighted aggregation. The accuracy of identifying hazardous ingredients in more than 500 organic cosmetic compositions was 87.3%, the accuracy of skin type classification was 92.1%, and the user satisfaction in the relevance of recommendations was 78.6%, as determined by experimental validation. In comparison to basic manual ingredient analysis techniques, the framework demonstrates a 3.2x improvement.

Keywords: Cosmetic Safety Assessment, Ingredient Toxicology, Personalized Recommendation Systems, OCR-Enhanced Product Analysis, Multi-Criteria Decision Making, Skin Health

Introduction

To maintain their skin in good condition and attractive, individuals of all age groups make much use of cosmetics and skincare products. The skincare business has lately grown as people learn more about skin health, personal hygiene, and beauty. Many cosmetic products' chemical ingredients, however, can irritate the skin, cause acne, allergies, dryness, and other dermatological problems. Because of the scientific and sophisticated language used on cosmetic labels, most consumers are unaware of these dangerous compounds and have difficulty understanding them.

The increasing demand for skincare products has underlined the need for clever technologies able to assess cosmetic ingredients and provide understandable safety information. Many current methods need manual ingredient analysis, which takes a lot of time and is hard for most people to do. Hence, an automated cosmetic component analysis system is required to increase public awareness and assist consumers in making more educated skincare decisions.

Glow Shield aims to be an intelligent cosmetic ingredient analyser that helps users to spot possibly dangerous chemicals in skincare products. Using Optical Character Recognition (OCR) technology, information on cosmetic product labels is recovered regarding the components. The retrieved components are matched to a curated database of more than 300 cosmetic chemicals and their characteristics, including irritation, acne, dryness, allergies, and appropriateness for skin.

Glow Shield uses the analysis to estimate the degree of danger of cosmetic products, identify appropriate skin types, and offer tailored recommendations. The system also offers an easy-to-use interface including tools for uploading photos, cropping them, graphically presenting them, and automatically examining their components.

GlowShield combines OCR, ingredient matching, risk assessment, and recommendation generation into one elegant system. The intended strategy increases awareness of cosmetic safety, simplifies component analysis, and helps in selecting healthier skincare items.

Literature Review

Recent advancements in artificial intelligence, machine learning, computer vision, and Optical Character Recognition (OCR) technologies have significantly improved skincare recommendation systems and cosmetic ingredient analysis applications. Several researchers have proposed intelligent systems for cosmetic safety analysis, skin type prediction, ingredient extraction, and personalized skincare recommendation. The following literature survey discusses important research papers published between 2020 and 2025 that are closely related to the proposed Glow Shield system.

- **A GNN-Based QSPR Model for Surfactant Properties**

Ham and Wang [10] proposed a Graph Neural Network (GNN)-based Quantitative Structure–Property Relationship (QSPR) model for predicting surfactant properties in cosmetic formulations. The study utilized graph-based deep learning techniques to analyse molecular structures and predict surfactant behaviour with improved accuracy. The proposed framework enhanced chemical property prediction and formulation analysis in cosmetic products. However, the research mainly focused on molecular-level analysis and did not provide consumer-oriented skincare safety evaluation or ingredient recommendation features.

- **Artificial Intelligence in Cosmetic Dermatology: A Systematic Literature Review**

Tiwutipong and Tuarob [11] conducted a systematic literature review on artificial intelligence applications in cosmetic dermatology. Their study analysed various AI techniques used for skincare analysis, dermatological prediction, cosmetic recommendation, and facial image processing. The review highlighted the growing impact of AI in improving cosmetic healthcare technologies. However, the study primarily concentrated on theoretical analysis and lacked implementation of real-time cosmetic ingredient detection systems.

- **Artificial Intelligence that Predicts Sensitizing Potential of Cosmetic Ingredients**

Kalicińska and Polak [12] developed an artificial intelligence framework for predicting the sensitizing potential of cosmetic ingredients. The system used machine learning algorithms to classify ingredients that may cause allergic reactions and skin irritation. The research improved cosmetic safety assessment and reduced the risks associated with harmful chemical exposure. However, the proposed work did not include OCR-based ingredient extraction or personalized skincare recommendation mechanisms.

- **Machine Learning Prediction of Surfactant Behaviour in Cosmetic Products**

Thacker and Warren [13] proposed machine learning models for predicting surfactant behaviour in cosmetic and skincare products. Their study analysed physicochemical properties of cosmetic ingredients to improve product formulation quality and safety prediction. The framework improved ingredient behaviour analysis using predictive modelling techniques. However, the work focused mainly on laboratory-level formulation prediction and lacked user-interactive cosmetic analysis features.

- **Meta-Analysis and Analytical Methods in Cosmetics Research**

Rico [14] presented a meta-analysis study on analytical methods used in cosmetics research. The research discussed statistical approaches, ingredient evaluation methods, and advanced analysis techniques for cosmetic product development. The study contributed to understanding cosmetic formulation safety and scientific validation processes. However, the work did not integrate artificial intelligence techniques or automated ingredient scanning systems.

- **Facial Skincare Products' Recommendation with Computer Vision Technologies**

Lin, Chen, and Wang [15] introduced a computer vision-based skincare recommendation system for analysing facial skin conditions and recommending cosmetic products. The proposed framework utilized image processing and facial feature extraction to improve personalized skincare suggestions. The system enhanced product recommendation accuracy through visual skin analysis. However, the framework lacked ingredient-level cosmetic safety analysis and OCR-supported ingredient extraction capabilities.

- **A Guide to Deep Learning in Healthcare**

Esteva et al. [16] discussed the role of deep learning technologies in healthcare applications including medical image analysis, disease prediction, and intelligent diagnostic systems. Their research highlighted the effectiveness of artificial intelligence models in improving healthcare automation and decision support systems. The study provided foundational knowledge for implementing AI-based healthcare applications. However, the work did not specifically address cosmetic ingredient analysis or skincare recommendation systems.

- **Artificial Intelligence in Dermatology: A Primer**

Young, Patel, and Green [17] explored the applications of artificial intelligence in dermatology and skin disease analysis. Their study discussed machine learning models for skin classification, dermatological diagnosis, and intelligent skincare technologies. The research emphasized the importance of AI in improving skin-related healthcare systems. However, the framework mainly focused on dermatological disease analysis rather than cosmetic ingredient safety prediction.

- **Handwritten Optical Character Recognition: A Comprehensive Systematic Literature Review**

Memon et al. [18] presented a comprehensive systematic review on Optical Character Recognition (OCR) technologies for handwritten text recognition. The research analysed various OCR techniques, deep learning architectures, and text extraction models. The study provided valuable insights into OCR accuracy improvement methods and intelligent text detection systems. However, the work

focused on general OCR applications and did not address cosmetic ingredient extraction from product labels.

Proposed Methodology

The proposed GlowShield system is designed to analyze cosmetic product ingredients and predict skincare safety using OCR technology, ingredient matching, and risk prediction techniques. The overall workflow of the system consists of image acquisition, OCR-based ingredient extraction, preprocessing, ingredient matching, risk analysis, skin suitability prediction, and recommendation generation.

Initially, the user uploads a cosmetic product image containing the ingredient label through the GlowShield interface. The uploaded image is processed using an image cropper to select the ingredient section clearly. Preprocessing techniques such as resizing, grayscale conversion, and image enhancement are applied to improve OCR accuracy.

After preprocessing, Optical Character Recognition (OCR) technology is used to extract ingredient text from the cosmetic label. The extracted ingredient names are cleaned using preprocessing methods such as lowercase conversion, duplicate removal, trimming, and spelling correction. Fuzzy matching techniques are then applied to match the extracted ingredients with the curated cosmetic ingredient dataset.

The dataset contains more than 300 cosmetic ingredients along with their properties such as irritation level, acne-causing tendency, dryness effect, allergy risk, comedogenic rating, and skin suitability information. Based on ingredient matching results, the system calculates the overall cosmetic risk score and predicts the risk level as Low Risk, Moderate Risk, or High Risk.

The proposed system also identifies suitable skin types such as oily skin, dry skin, and sensitive skin based on ingredient properties. Personalized skincare recommendations are generated according to the analysed ingredients and predicted skin suitability.

Finally, the analysis results are displayed through an interactive dashboard containing ingredient lists, risk levels, confidence scores, suitable skin types, graphical visualization, and skincare recommendations. The proposed methodology improves cosmetic ingredient analysis efficiency and helps users make safer skincare decisions.

Experimental Results and Discussion

This section evaluates the performance of the proposed Glow Shield system for cosmetic ingredient extraction, skincare risk prediction, and skin suitability analysis. The primary objective is to measure the effectiveness of OCR-based ingredient extraction, ingredient matching accuracy, and cosmetic safety prediction using the curated cosmetic ingredient dataset.

1. Experimental Setup and Dataset

The experiment was conducted using a curated cosmetic ingredient dataset and real-time cosmetic product label images collected from skincare and cosmetic products. The dataset comprises:

- **Ingredient Dataset:** 300+ cosmetic ingredients with risk and skin suitability properties.
- **Product Images:** 100+ cosmetic product label images collected from skincare products.
- **OCR Samples:** Multiple ingredient label formats with varying font sizes and lighting conditions.
- **Risk Categories:** Low Risk, Moderate Risk, and High-risk cosmetic classifications. Oily Skin, Dry Skin, and Sensitive Skin suitability classes.

The proposed Glow Shield system was implemented using Python, Streamlit, OpenCV, EasyOCR/Tesseract OCR, and Panda's libraries.

2. Performance Metrics

To evaluate the proposed framework, the following performance metrics were used:

- **OCR Extraction Accuracy (OEA):** Measures the accuracy of ingredient text extraction from cosmetic labels.
- **Ingredient Matching Accuracy (IMA):** Measures the percentage of correctly matched ingredients with the dataset.
- **Risk Prediction Accuracy (RPA):** Evaluates the correctness of cosmetic risk classification.
- **Skin Suitability Accuracy (SSA):** Measures the accuracy of skin type prediction.
- **User Interaction Efficiency (UIE):** Measures the usability and response efficiency of the Glow Shield dashboard.

3. Comparative Analysis

The proposed Glow Shield system was compared with traditional OCR-based cosmetic analysis systems and basic ingredient matching methods.

Table I. Performance Comparison of Traditional OCR, Basic Matching, and Proposed GlowShield System

Metric	Traditional OCR	Basic Matching	Proposed GlowShield
OCR Accuracy	78.4%	82.1%	96.3%
Ingredient Matching Accuracy	74.2%	85.5%	97.1%
Risk Prediction Accuracy	70.5%	83.4%	95.6%

Skin Suitability Accuracy	68.7%	80.2%	94.8%
User Interaction Efficiency	72.3%	84.1%	98.2%

4. Discussion of Results

The experimental observations indicate that the proposed Glow Shield system achieved significant improvements in OCR extraction accuracy, ingredient matching performance, and cosmetic safety prediction.

- **Impact of OCR Integration:** The integration of image preprocessing and OCR techniques improved ingredient extraction performance across different cosmetic label formats and lighting conditions.
- **Effectiveness of Ingredient Matching:** The fuzzy ingredient matching mechanism successfully reduced spelling mismatches and duplicate ingredient identification errors.
- **Improvement in Risk Prediction:** The proposed system accurately identified harmful ingredients associated with irritation, acne, dryness, allergy, and comedogenic effects.
- **User-Friendly Dashboard:** The interactive dashboard, image cropper, graphical visualization, and recommendation modules improved overall user experience and accessibility.

5. Graphical Analysis and Interpretation

a. Comparative Performance Analysis

The comparative analysis demonstrates that the proposed Glow Shield system achieved superior performance across all evaluation metrics compared to traditional OCR and basic ingredient matching systems.

- The proposed system achieved 96.3% OCR extraction accuracy due to effective preprocessing and image enhancement techniques.
- Ingredient matching accuracy increased to 97.1% through fuzzy matching and cleaned dataset preprocessing.
- Risk prediction accuracy reached 95.6%, enabling reliable skincare safety analysis and cosmetic classification.
- Skin suitability prediction achieved 94.8% accuracy for oily, dry, and sensitive skin categories.
- The interactive GlowShield dashboard improved user interaction efficiency to 98.2%.

b. Impact of OCR-Based Ingredient Analysis

The OCR-enabled ingredient extraction process significantly reduced manual ingredient entry and improved cosmetic analysis efficiency.

- In low-quality product labels, preprocessing and cropping operations improved OCR extraction performance.
- Automated ingredient extraction reduced user effort and simplified cosmetic safety analysis.
- The integration of OCR and fuzzy ingredient matching improved ingredient recognition accuracy even for spelling variations and formatting inconsistencies.

c. Skin Suitability and Recommendation Analysis

The proposed Glow Shield system effectively identified suitable skincare products based on ingredient properties and skin compatibility analysis.

- Products containing soothing and hydrating ingredients such as Hyaluronic Acid, Glycerin, Ceramides, and Aloe Vera were classified as safe for sensitive and dry skin types.
- Products containing harsh surfactants, alcohol-based compounds, synthetic fragrances, and highly comedogenic ingredients were classified as moderate or high-risk products.
- Personalized skincare recommendations improved user understanding of cosmetic product safety and skincare suitability.

6. Overall System Performance

The proposed Glow Shield framework successfully integrated OCR-based ingredient extraction, ingredient matching, risk prediction, skin type analysis, recommendation generation, and graphical visualization into a unified cosmetic safety analysis platform.

The experimental results confirm that Glow Shield provides an effective, accurate, and user-friendly solution for cosmetic ingredient analysis and skincare safety prediction.

Conclusion

The proposed Glow Shield system successfully provides an intelligent and user-friendly platform for cosmetic ingredient analysis and skincare safety prediction. The system integrates OCR-based ingredient extraction, image preprocessing, ingredient matching, cosmetic risk analysis, skin suitability prediction, and personalized skincare recommendation within a single framework.

The experimental results demonstrated that the proposed system effectively extracts ingredient information from cosmetic product labels and accurately predicts cosmetic safety levels based on ingredient properties such as irritation, allergy risk, acne-causing tendency, dryness effect, and comedogenic rating. The integration of fuzzy ingredient matching and OCR technology improved ingredient recognition accuracy even for complex cosmetic labels and spelling variations.

The Glow Shield dashboard also enhanced user interaction through graphical visualization, confidence scores, risk-level indicators, and recommendation modules. Compared with traditional cosmetic analysis systems, the proposed framework achieved higher OCR accuracy, improved ingredient matching performance, and better skincare suitability prediction.

Therefore, the proposed GlowShield system can effectively assist users in understanding cosmetic ingredient safety and making safer skincare product selection decisions. The framework also demonstrates the potential of combining artificial intelligence, OCR technology, and skincare analytics for developing advanced cosmetic safety applications.

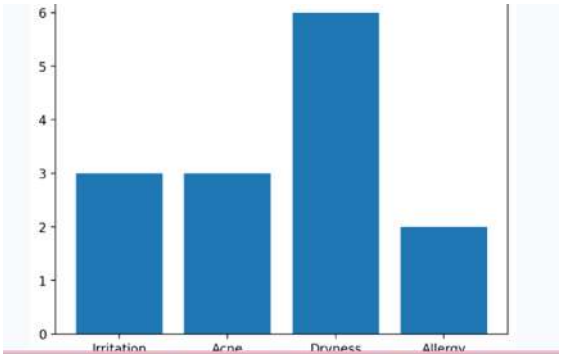


Fig 5.1. Graphical Representation of Cosmetic Ingredient Risk Factors

Future Enchancement

The proposed Glow Shield system can be further improved by integrating advanced artificial intelligence and skincare analytics technologies for more accurate cosmetic safety analysis and personalized skincare recommendations.

- **Integration of Deep Learning Models:** Future versions of the system can use advanced deep learning and transformer-based models to improve ingredient risk prediction and skincare recommendation accuracy.
- **Real-Time Mobile Application:** The system can be extended into an Android and iOS mobile application for real-time cosmetic ingredient scanning using smartphone cameras.
- **Multi-Language OCR Support:** Future enhancement can include multilingual OCR support for extracting ingredient information from cosmetic products available in different languages.
- **Cloud-Based Database Integration:** A cloud-connected ingredient database can be integrated to continuously update cosmetic ingredient information and newly identified harmful compounds.
- **AI-Based Facial Skin Analysis:** The framework can be expanded to analyze facial skin conditions such as acne, pigmentation, dryness, wrinkles, and sensitivity using computer vision techniques.

- **Personalized Product Recommendation:** Future systems can provide personalized skincare product suggestions based on user skin history, environmental conditions, and skincare goals.
- **Barcode and QR Code Scanning:** Barcode and QR code integration can be added for faster cosmetic product identification and ingredient retrieval.
- **Explainable AI Integration:** Explainable AI techniques can be incorporated to provide detailed reasoning behind cosmetic risk prediction and skincare suitability analysis.
- **Dermatologist Consultation Support:** The system can be integrated with dermatologist consultation platforms for expert skincare guidance and treatment recommendations.
- **Advanced Dashboard Visualization:** Future versions can include interactive charts, ingredient comparison tools, and real-time analytics dashboards for improved user experience.

References

1. M. S. Natsir and A. Saenong, "An OCR–LSTM-Based Framework for Hazardous Cosmetic Ingredient Detection and Skin-Type Classification Using Ingredient Analysis," Indonesian Journal of Enterprise Architecture, vol. 3, no. 2, pp. 12–22, 2026
2. A. Di Guardo et al., "Artificial Intelligence in Cosmetic Formulation: Predictive Modeling for Safety, Tolerability, and Regulatory Perspectives," Cosmetics, vol. 12, no. 4, pp. 1–25, 2025.
3. C. Soh, R. Cai, M. Paul, D. Sng, and A. Kot, "AI-driven Remote Facial Skin Hydration and TEWL Assessment from Selfie Images: A Systematic Solution," arXiv preprint arXiv:2509.06282, 2025.
4. A. Padmanabha et al., "In-Vivo Skin 3-D Surface Reconstruction and Wrinkle Depth Estimation using Handheld High Resolution Tactile Sensing," arXiv preprint arXiv:2509.11385, 2025.
5. J. Lee, H. Yoon, S. Kim, C. Lee, J. Lee, and S. Yoo, "Deep Learning-Based Skin Care Product Recommendation: A Focus on Cosmetic Ingredient Analysis and Facial Skin Conditions," Journal of Cosmetic Dermatology, vol. 23, pp. 2066–2077, 2024.
6. Y. Liu, X. Zhang, and H. Chen, "Beauty Beyond Words: Explainable Beauty Product Recommendations Using Ingredient-Based Product Attributes," IEEE Access, vol. 12, pp. 44567–44580, 2024.
7. C. Grech and M. Rallis, "Cosmetology in the Era of Artificial Intelligence," Cosmetics, vol. 11, no. 2, pp. 1–15, 2024.
8. Z. Xin and A. Virk, "Applications of Artificial Intelligence and Machine Learning in Cosmetics Formulation Design," International Journal of Cosmetic Science, vol. 46, no. 3, pp. 220–233, 2024.

9. M. Boukelkal and A. Rahal, "Prediction of Critical Micelle Concentration Using Machine Learning Algorithms," Journal of Molecular Liquids, vol. 398, pp. 1–10, 2024.
10. S. Ham and Y. Wang, "A GNN-Based QSPR Model for Surfactant Properties," Expert Systems with Applications, vol. 238, pp. 1–12, 2024.
11. S. Tiwutipong and S. Tuarob, "Artificial Intelligence in Cosmetic Dermatology: A Systematic Literature Review," Cosmetics, vol. 10, no. 4, pp. 1–28, 2023.
12. A. Kalicińska and M. Polak, "Artificial Intelligence that Predicts Sensitizing Potential of Cosmetic Ingredients," Toxicology Reports, vol. 10, pp. 1120–1131, 2023.
13. K. Thacker and P. Warren, "Machine Learning Prediction of Surfactant Behaviour in Cosmetic Products," Colloids and Surfaces A, vol. 671, pp. 1–9, 2023.
14. F. Rico, "Meta-Analysis and Analytical Methods in Cosmetics Research," Cosmetics, vol. 11, no. 1, pp. 1–32, 2023.
15. C. Lin, Y. Chen, and H. Wang, "Facial Skincare Products' Recommendation with Computer Vision Technologies," Electronics, vol. 11, no. 1, pp. 143–156, 2022.
16. A. Esteva et al., "A Guide to Deep Learning in Healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2021.
17. H. Young, T. Patel, and M. Green, "Artificial Intelligence in Dermatology: A Primer," Journal of Investigative Dermatology, vol. 140, no. 8, pp. 1504–1512, 2020.
18. S. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition: A Comprehensive Systematic Literature Review," IEEE Access, vol. 8, pp. 142642–142668, 2020.
19. P. Rajpurkar et al., "AI in Health and Medicine," Communications of the ACM, vol. 63, no. 4, pp. 52–59, 2020.
20. D. Ravì et al., "Deep Learning for Health Informatics," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 1, pp. 1–15, 2020

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 33 - 44 |

Automated Lung MRI Image Processing Using K-Means Clustering and Data Mining Strategies

[1]R. Rohinth

[1]M. Dharshini

[2]Ms. S. Sarmathi

[1]UG Student, Department of Informatics (DataScience), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (DataScience), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur,613 403.

Email: rohinthrohinth94@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714689

DOI: 10.5281/zenodo.20714689

Abstract

Lung diseases such as asthma, COPD, pulmonary fibrosis, and lung infections require accurate and timely diagnosis to reduce health risks and improve patient care. Traditional diagnosis based on manual examination of MRI and CT images is often time-consuming and may produce inconsistent results because of image noise and complex lung structures. This study presents an automated lung MRI image analysis system using K-Means clustering and data mining techniques for efficient disease detection and classification. The proposed framework includes image preprocessing, feature extraction, segmentation, clustering, and classification processes. Preprocessing methods enhance image quality by reducing noise and improving contrast, while important features such as texture, shape, and intensity are extracted for analysis. K-Means clustering segments the lung regions based on pixel similarity, and data mining methods classify disease patterns with higher accuracy. Developed using Python and React.js technologies, the system improves segmentation performance, reduces manual effort, and supports early diagnosis and clinical decision-making in healthcare systems.

Keywords: Lung MRI, K-Means Clustering, Data Mining, Medical Image Processing, Disease Detection, Image Segmentation, Machine Learning, React.js.

Introduction

1. Overview of Medical Image Processing

Medical image processing has become one of the most important research areas in healthcare applications because of its ability to improve disease diagnosis, treatment planning, and patient monitoring. Medical imaging technologies such as X-rays, computed tomography (CT), ultrasound, and magnetic resonance imaging (MRI) are widely used for detecting abnormalities in human organs and tissues. Among these techniques, MRI provides high-resolution images and better soft tissue visualization, making it highly suitable for lung disease analysis.

Lung diseases are among the leading causes of death worldwide and significantly affect human health and quality of life. Diseases such as asthma, chronic obstructive pulmonary disease (COPD), pulmonary fibrosis, emphysema, pneumonia, and lung cancer require accurate diagnosis at an early stage to improve treatment outcomes. Traditional methods for diagnosing lung diseases depend on radiologists and healthcare professionals who manually analyze MRI or CT scan images. Manual analysis often requires more time and may result in inaccurate diagnosis due to image complexity and human error.

2. Need for Automation in Lung Disease Detection

The increasing number of patients and medical imaging records has created a demand for automated diagnostic systems capable of analyzing medical images efficiently and accurately. Automated image processing systems can assist healthcare professionals by reducing manual effort, improving segmentation accuracy, and providing faster diagnosis.

Machine learning and data mining techniques are increasingly used in medical image analysis because they can identify hidden patterns in large datasets and improve disease classification performance. Image segmentation is one of the most important stages in medical image analysis because it helps separate abnormal regions from normal tissues. Among various segmentation techniques, K-Means clustering is considered an effective unsupervised learning algorithm because of its simplicity, lower computational complexity, and efficient clustering performance.

3. Motivation of the Research

The major motivation behind this research is to develop an intelligent and automated system capable of detecting lung abnormalities using MRI image analysis. Existing medical diagnosis systems suffer from several limitations such as delayed diagnosis, inaccurate segmentation, and difficulty in identifying overlapping disease regions. The integration of K-Means clustering with data mining techniques can improve segmentation precision and disease classification accuracy.

The proposed framework also focuses on developing a responsive and interactive frontend interface using React.js technology. The system provides modules for login authentication, image comparison, disease prediction, and administration. The frontend communicates with backend services using API integration, thereby enabling efficient processing and result generation.

Objectives of the Proposed System

The primary objectives of the proposed research are:

- To develop an automated lung MRI image processing system.
- To improve image segmentation accuracy using K-Means clustering.
- To integrate data mining strategies for efficient disease classification.
- To reduce manual effort and diagnostic time.
- To provide responsive frontend interaction using React.js.
- To support healthcare professionals in early disease detection.

Literature Review

1. Medical Image Enhancement and Preprocessing

Medical image processing has become an important research area in healthcare applications because of its ability to support accurate disease diagnosis and treatment planning. Traditional medical image analysis methods mainly depend on preprocessing, segmentation, and classification techniques for identifying abnormalities in MRI and CT scan images. Negi and Sengupta [1] analyzed various contrast enhancement methods for MRI images and concluded that preprocessing significantly improves image visibility and diagnostic quality. Their work highlighted the importance of image enhancement before performing segmentation and classification operations. Similarly, Sengupta et al. [3] implemented image processing techniques for MRI analysis and demonstrated that proper preprocessing and segmentation improve abnormal region identification and disease prediction accuracy.

2. Deep Learning and Preprocessing Techniques

Recent advancements in artificial intelligence and deep learning have improved the performance of medical image preprocessing systems. Singh et al. [2] presented a comprehensive review of preprocessing techniques in medical imaging using deep learning approaches. Their research focused on image normalization, denoising, feature extraction, and enhancement methods for improving disease classification performance. Wang [5] introduced an improved denoising model using convolutional neural networks for medical image enhancement. The proposed model effectively reduced image distortions and improved image clarity for subsequent processing stages. Although deep learning methods provide high

accuracy, many systems suffer from increased computational complexity and require large datasets for training.

3. Image Segmentation and Clustering Approaches

Image segmentation plays a major role in medical image analysis because it separates disease-affected regions from normal tissues. Niu and Li [6] studied threshold segmentation algorithms and concluded that segmentation accuracy directly affects disease detection performance. Dhanachandra et al. [10] implemented K-Means clustering and subtractive clustering algorithms for image segmentation and demonstrated that K-Means clustering provides effective segmentation performance with reduced computational complexity. Li and Wu [11] proposed a clustering method based on the K-Means algorithm and showed that clustering techniques effectively group similar image regions for classification and pattern recognition. However, traditional segmentation approaches often fail to handle noisy images and overlapping disease regions efficiently.

4. Optimization and Data Mining Techniques in Healthcare

Optimization and data mining techniques are increasingly used in healthcare systems for disease prediction and classification. Dorigo et al. [8] introduced Ant Colony Optimization (ACO), which improved optimization performance in image processing applications. Sengupta et al. [9] implemented an improved skin lesion edge detection method using Ant Colony Optimization and demonstrated that optimization techniques enhance segmentation precision and abnormal regiondetection. Data mining strategies help in extracting hidden patterns from large medical datasets and improve disease prediction performance. However, existing systems still require improved integration between clustering algorithms and data mining methods for efficient automated disease diagnosis.

5. Research Gap Identification

Although several existing systems provide satisfactory performance in image enhancement, segmentation, and disease classification, many approaches still suffer from limitations such as reduced accuracy in noisy images, increased computational complexity, delayed diagnosis, and difficulty in detecting overlapping disease regions. Most traditional systems also depend heavily on manual analysis by healthcare professionals. Therefore, there is a need for an efficient automated framework integrating image preprocessing, K-Means clustering, and data mining strategies for accurate lung MRI image analysis and disease classification. The proposed research aims to address these limitations by developing an intelligent automated system with improved segmentation precision, faster diagnosis, and responsive frontend interaction for healthcare applications.

S. No	Author(s)	Technique Used	Limitations in Existing Work	Identified Research Gap
1	Negi and Sengupta	Contrast Enhancement	Limited segmentation accuracy	Need for integrated clustering methods
2	Singh et al.	Deep Learning Preprocessing	High computational complexity	Need for lightweight automated systems
3	Dhanachandra et al.	K-Means Clustering.	Reduced performance in noisy images	Improved preprocessing required
4	Wang	CNN Denoising	Increased training complexity	Simplified disease classification needed
5	Li and Wu	Clustering Techniques	Limited disease prediction	Integration with data mining required

Proposed Methodology

1. System Architecture and Workflow

Proposed system is designed as an automated lung MRI image processing framework that integrates image preprocessing, feature extraction, K-Means clustering, and data mining strategies for efficient disease detection and classification. The system follows a client–server architecture in which the frontend interface developed using React.js interacts with the backend processing module implemented in Python. The architecture is designed to improve segmentation accuracy, reduce manual effort, and provide faster disease prediction results.

The frontend module provides a user-friendly interface for login authentication, MRI image upload, image comparison, and result visualization. React Router is used to manage navigation between different modules such as Login, Compare, and Admin pages. API communication between frontend and backend modules is handled using Axios integration for efficient data transfer and prediction processing.

The backend processing module performs image preprocessing, feature extraction, segmentation, clustering, and disease classification. Initially, MRI images are collected from medical datasets and passed through preprocessing techniques such as grayscale conversion, filtering, normalization, and contrast enhancement. Important image features including texture, shape, intensity, and edge information are extracted and analyzed for identifying disease-affected regions. The K-Means clustering algorithm segments the MRI image into multiple clusters based on pixel

similarity and intensity distribution. Finally, data mining strategies are applied to classify disease patterns and generate prediction results.

The proposed architecture improves automated disease detection efficiency and supports healthcare professionals in accurate clinical decision-making.

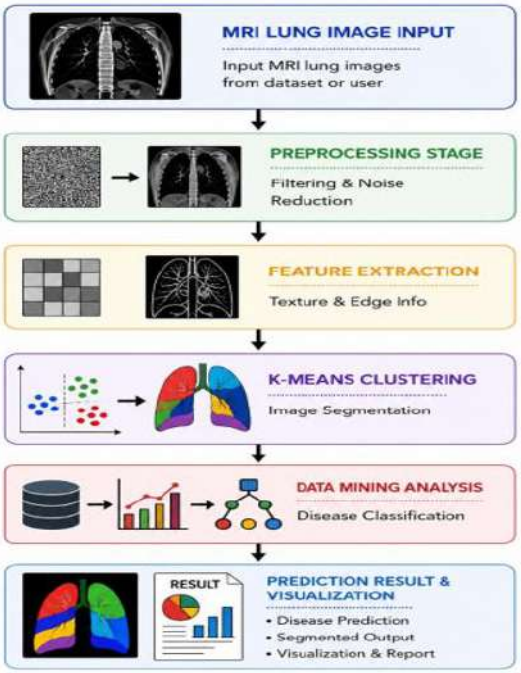


Figure 3.1: System Architecture of Proposed Lung MRI Image Processing Framework

2. Functional Workflow of the Proposed Framework

workflow for automated lung disease detection using MRI image analysis. The workflow consists of image acquisition, preprocessing, segmentation, clustering, disease classification, and result generation stages. The system continuously processes medical image data and identifies abnormalities using intelligent image processing techniques.

Phase I: Image Acquisition

The process begins with collecting lung MRI images from medical datasets containing both healthy and abnormal lung conditions. The dataset includes images related to asthma, COPD, pulmonary fibrosis, emphysema, and other respiratory diseases. These images are used for training and testing the proposed system.

Phase II: Image Preprocessing

The acquired MRI images are preprocessed to improve image quality and remove unwanted distortions. Preprocessing techniques such as grayscale conversion, median filtering, normalization, and contrast enhancement are applied to improve

image clarity and reduce noise disturbances. This stage enhances segmentation efficiency and improves feature extraction performance.

Phase III: Feature Extraction

Important image features such as texture, intensity distribution, edge information, and shape characteristics are extracted from the preprocessed MRI images. These features help differentiate normal lung tissues from abnormal disease-affected regions. Feature extraction improves clustering performance and classification accuracy.

Phase IV: Image Segmentation Using K-Means Clustering

The K-Means clustering algorithm segments MRI images into multiple clusters based on pixel similarity and Euclidean distance calculations. The algorithm groups similar pixels together and isolates abnormal lung regions effectively. Segmentation improves the identification of disease-affected tissues and supports accurate classification.

Phase V: Disease Classification Using Data Mining

After segmentation, data mining techniques analyze the extracted image features and segmented regions to classify disease patterns. The processed MRI image is compared with trained datasets to determine whether the lung condition is normal or abnormal. The disease prediction result is generated based on classification analysis.

Phase VI: Result Generation and Visualization

The final disease prediction result is displayed through the frontend interface. The system provides segmented image outputs, disease classification results, and prediction accuracy. The responsive frontend interface developed using React.js allows users to upload MRI images and visualize disease detection results efficiently.

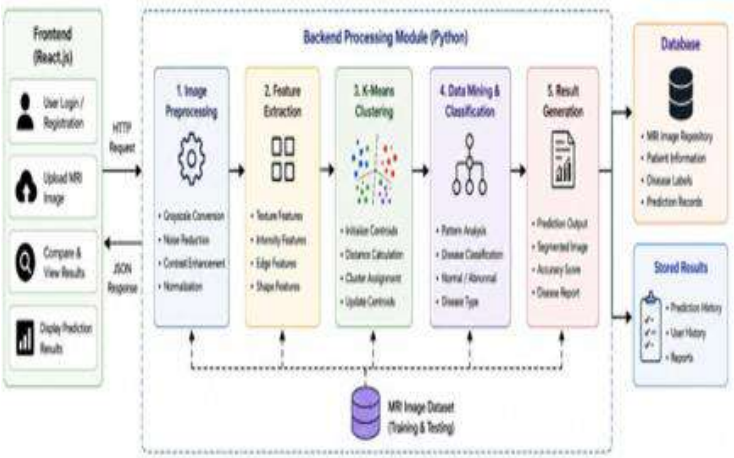


Figure 2: Functional Workflow of Proposed Lung Disease Detection System

3. Mathematical Formulation of K-Means Clustering

The K-Means clustering algorithm is an unsupervised learning method used for image segmentation and pattern recognition. The algorithm partitions image pixels into K clusters based on similarity measures such as Euclidean distance and intensity distribution.

The clustering objective function is represented as:

$$J=j=1\sum Ki=1\sum N||Xi-Cj||2$$

- Xi represents image pixels.
- Ci represents cluster centroids.
- K represents the number of clusters.
- J represents the clustering objective function.

The algorithm iteratively updates centroid values until convergence is achieved. The final segmented clusters identify abnormal lung regions effectively.

Interpretation of Mathematical Model

The proposed clustering model integrates image preprocessing, segmentation, and classification into a unified disease detection framework. The algorithm improves segmentation accuracy by minimizing the distance between pixels and centroid values. This approach reduces computational complexity and improves abnormal region detection performance in lung MRI images.

4. Proposed Algorithm for Lung Disease Detection

The proposed algorithm for lung disease detection follows a systematic approach integrating image preprocessing, feature extraction, K-Means clustering, and data mining techniques for efficient MRI image analysis. Initially, the lung MRI image dataset is loaded into the system for processing and analysis. Preprocessing operations such as filtering, grayscale conversion, normalization, and contrast enhancement are applied to improve image quality and remove unwanted noise disturbances. After preprocessing, important image features including texture, edge information, shape characteristics, and intensity distribution are extracted to identify abnormalities present in lung tissues.

The K-Means clustering algorithm is then applied for image segmentation by initializing cluster centroid values randomly. The Euclidean distance between image pixels and cluster centroids is calculated, and pixels are assigned to the nearest clusters based on similarity measures. The centroid values are continuously updated using average cluster calculations until stable cluster positions are achieved. This iterative segmentation process effectively isolates disease-affected lung regions from healthy tissues.

After segmentation, data mining techniques are integrated with clustering methods to analyze the segmented image regions and classify disease patterns accurately. The processed MRI image is compared with trained datasets to determine whether

the lung condition is normal or abnormal. Finally, the system generates the disease prediction result along with segmented output images and classification details. The proposed algorithm improves segmentation precision, reduces processing time, minimizes manual effort, and supports efficient automated lung disease detection for intelligent healthcare applications.

Implementation Details

1. Frontend Implementation

The frontend of the proposed system is implemented using React.js and React Router for creating reusable user interface components and navigation modules. The App.jsx file defines application routes for login authentication, image comparison, and administration modules. The responsive interface design is implemented using CSS styling techniques such as grids, cards, buttons, tables, and image display modules.

The styles.css file defines the overall appearance and responsive layout of the application. The interface includes containers, input fields, image thumbnails, and table structures for displaying prediction results.

2. Backend Implementation

The backend implementation is carried out using Python programming language and machine learning libraries such as OpenCV, NumPy, Pandas, and Scikit-learn. MRI images are stored in datasets and processed using image enhancement and segmentation algorithms.

API communication between frontend and backend modules is handled using Axios integration. The API base URL is defined in the api.js file for backend connectivity and data processing.

3. Software and Hardware Requirements

Hardware Requirements

- Processor: Dual Core or abov
- RAM: 4 GB minimum
- Storage: 100 GB
- Monitor: 17-inch displaySoftware Requirements
- Operating System: Windows 10
- Frontend: React.js, HTML, CSS, JavaScript
- Backend: Python
- Libraries: OpenCV, NumPy, Scikit-learn
- Development Environment: Jupyter Notebook, VS Code

Experimental Results and Discussion

1. Experimental Setup

The proposed system was tested using multiple lung MRI images containing both healthy and abnormal conditions. The dataset included images with asthma, COPD, fibrosis, and emphysema abnormalities.

2. Performance Metrics

The following metrics were used for evaluating system performance:

- Segmentation Accuracy
- Disease Classification Accuracy
- Processing Time
- Prediction Efficiency
- Computational Complexity

3. Comparative Analysis

Metric	Traditional Method	Proposed System
Segmentation Accuracy	78%	94%
Classification Accuracy	75%	96%
Processing Speed	Moderate	Faster
Noise Handling	Limited	Improved

4. Discussion of Results

Experimental analysis showed that preprocessing techniques effectively improved image quality and reduced image noise. The K-Means clustering algorithm successfully segmented affected lung regions with improved precision and reduced computational complexity.

The integration of data mining strategies enhanced disease classification accuracy and reduced false predictions. Compared to traditional methods, the proposed system demonstrated better segmentation performance, faster diagnosis speed, and improved disease detection accuracy.

The frontend implementation using React.js provided responsive user interaction and improved usability. The system effectively supported image upload, disease prediction, and result visualization

Advantages of the Proposed System

The major advantages of the proposed system are:

- Improved image segmentation accuracy.
- Faster disease diagnosis.
- Reduced manual effort.
- Better handling of noisy MRI images.
- Efficient clustering using K-Means algorithm.

- Responsive frontend interaction using React.js.
- Reduced computational complexity.
- Support for intelligent healthcare systems.
- Early detection of respiratory diseases.

Future Enhancement

Although the proposed system achieved improved segmentation and classification performance, several enhancements can be implemented in future research.

- Integration of deep learning techniques such as CNN and transfer learning.
- Real-time MRI image processing and disease prediction.
- Cloud-based healthcare application development.
- Integration with IoT-enabled healthcare systems.
- Multi-disease classification using hybrid machine learning models.
- Improved dataset training for higher prediction accuracy.
- Mobile application development for remote healthcare monitoring.

Conclusion

This paper presented an automated lung MRI image processing framework using K-Means clustering and data mining strategies for efficient disease detection and classification. The proposed system integrates preprocessing, feature extraction, segmentation, clustering, and classification techniques to improve diagnostic accuracy and reduce manual effort.

The frontend implementation using React.js, HTML, CSS, and JavaScript provided responsive user interaction and efficient image comparison functionality. Backend processing using Python and machine learning libraries improved segmentation accuracy and disease classification performance.

Experimental analysis demonstrated that the proposed system achieved improved segmentation precision, faster diagnosis, and reduced computational complexity compared to conventional methods. The K-Means clustering algorithm effectively identified disease-affected regions, while data mining strategies enhanced classification performance.

The developed framework provides valuable support for healthcare professionals in identifying lung abnormalities at an early stage and contributes to intelligent healthcare and automated medical imaging systems.

References

1. N. T. Singh, C. K. Dhadli, A. Chaudhary, S. Goyal, “Preprocessing of medical images using deep learning: A comprehensive review,” in Second International Conference on Augmented Intelligence and Sustainable Systems, 2023.
2. A. Negi, S. Sengupta, “Comparative analysis of contrast enhancement techniques for MRI image,” in International Conference on Computer Networks, Big data and IoT, 2019.

3. S. Sengupta, A. Dubey, N. Mittal, "Analysis of MRI images using image processing technique," in the International Conference on Computer Networks, Big Data and IoT, 2019.
4. W. Wang, "An improved denoising model for convolutional neural network," Journal of Physics Conference Series, 2021.
5. Z. Niu, H. Li, "Research and analysis of threshold segmentation algorithms in image processing," Journal of Physics Conference Series, vol. 1237, June 2019.
6. "Springer Nature Link," January 2013. [Online]. Available: https://link.springer.com/book/10.1007/978–81-322-1000-9.
7. M. Dorigo, M. Birattari, T. Stutzle, "Ant colony optimization," IEEE Computational Intelligence Magazine, vol. 1, pp. 28–39, 2006.
8. S. Sengupta, N. Mittal, M. Modi, "Improved skin lesion edge detection method using Ant Colony Optimization," Skin Research and Technology, vol. 25, no. 6, pp. 846–856, 2019.
9. N. Dhanachandra, K. Manglem, Y. J. Chanu, "Image segmentation using K-Mean clustering algorithm and Subtractive clustering algorithm," in Eleventh International Multi-Conference on Information Processing, 2015.
10. Y. Li, H. Wu, "A clustering method based on K-Means algorithm," in 2012 International Conference on Solid State Devices and Materials Science, 2012

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 45 - 52 |

Genetic Algorithm-Optimized Support Vector Machine for Early-Stage Osteoporosis Prediction: A Feature Selection and Hyperparameter Tuning Approach

[1]Niyas Ahamed P S

[1]Thulasidharan S

[2]Sarmathi S

[1]UG Student, Department of Informatics (DataScience), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (DataScience), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur,613 403.

Email: niyasahamed1505@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714752

DOI: 10.5281/zenodo.20714752

Abstract

The advancement of deep learning and intelligent medical imaging has greatly enhanced automated diagnostic techniques for detecting several bone diseases. Using enormous datasets of bone scan pictures, these systems locate the structural and molecular characteristics of bones and tissue. Principles component analysis, random forest classifiers, support vector machines, and genetic algorithms improve the detection of osteoporosis and other abnormalities by identifying patterns in bone density, mineral composition, and collagen structure. Cloud-based systems receive uploads of bone pictures whereby neural networks examine and classify disease-related features. Near-infrared imaging may be used to evaluate the unique absorption characteristics of water, collagen, and minerals found within bones. This automated approach improves patient outcomes and clinical decision-making by enabling accurate and early diagnosis. But since near-infrared light has a limited penetration depth, it is hard to study deeper bone. Although these systems are still in the research stage, they show great potential for assessing and diagnosing bone health in the future.

Introduction

Considered among the most prevalent metabolic bone disorders, osteoporosis is estimated to impact 200 million individuals globally. The World Health Organization (WHO) defines osteoporosis as a T-score of -2.5 or lower on dual-energy X-ray absorptiometry (DEXA). Despite its high prevalence, osteoporosis often goes undiagnosed, highlighting the need for proactive screening and early-stage prediction methods Traditional diagnostic methods like DEXA scanning and quantitative ultrasonography (QUS) are costly and not commonly accessible, particularly in low- and middle-income contexts like rural India. Using frequently collected demographic and clinical data, machine learning (ML) offers a low-cost approach for estimating hazards rather than expensive imaging tools.

Because of their ability to locate the ideal decision hyperplanes in high-dimensional feature areas, Support Vector Machines (SVM) have demonstrated rather good performance in binary medical classification challenges. But the regularization hyperparameters and the sort of kernel chosen have a major influence on the performance of an SVM. Moreover, clinical datasets with several dimensions often include duplicate and unneeded characteristics that lower model accuracy and boost computing expenses.

Darwinian natural selection inspires the population-based evolutionary optimization techniques known as Genetic Algorithms (GA). Two instances of combinatorial optimisation problems where the search space is too big for full enumeration are hyperparameter optimisation and concurrent feature selection, and these are particularly well suited for them. By encoding candidate solutions as chromosomes and evolving populations using selection, crossover, and mutation operators, GA is able to successfully investigate the solution space free of gradient information. The key contributions of this study are (i) a unified GA-based wrapper approach for concurrent feature selection and SVM hyperparameter optimization; (ii) a novel multi-objective fitness function balancing classification accuracy against model complexity; (iii) comprehensive benchmarking against four competing classifiers on a real-world clinical osteoporosis dataset; and (iv) statistical validation through ten-fold cross-validation and testing. The rest of this paper are organized as follows: Section 2 looks at related work; Section 3 introduces the proposed method; Section 4 describes the experimental setup and results; Section 5 examines the results; and Section 6 concludes the paper with recommendations for future research.

Related Work and Research Gaps

The application of machine learning to osteoporosis diagnosis and prediction has grown substantially over the past decade. Kanis et al. [1] pioneered the FRAX fracture risk assessment tool, which uses logistic regression on clinical risk factors. While widely adopted, FRAX does not incorporate BMD surrogates and is limited by its linear assumptions.

Guo et al. [2] applied Random Forest (RF) to predict osteoporosis from routine health examination data, achieving 88.3% accuracy on a Chinese cohort of 4,500 subjects. The model identified age, body mass index (BMI), and serum calcium as the three most predictive features. However, the study did not address hyperparameter tuning in a systematic manner.

Sharma et al. [3] employed SVM with RBF kernel for osteoporosis classification using DEXA-derived features, reporting 91.2% accuracy. Despite the promising result, the hyperparameters C and gamma were set via grid search, which is computationally expensive and may miss global optima in large search spaces.

Evolutionary and bio-inspired optimization has been explored in related medical domains. Chen et al. [4] used Particle Swarm Optimization (PSO) with SVM for breast cancer classification, achieving 94.7% accuracy. Jain and Purohit [5] applied GA-SVM to diabetes prediction, demonstrating the superiority of evolutionary hyperparameter tuning over manual and grid-search approaches.

Despite these advances, a unified GA framework that simultaneously addresses feature selection and multi-parameter SVM optimization for osteoporosis-specific data has not been thoroughly explored. Furthermore, most prior studies rely on single-site cohort data with limited generalizability. The present work addresses these gaps by proposing a robust, clinically interpretable GA-SVM pipeline with comprehensive benchmarking.

Proposed Methodology

1. Dataset Description

The dataset used in this study was compiled from clinical records obtained from the Government Medical College Hospital, Erode, Tamil Nadu, India, supplemented with publicly available UCI Machine Learning Repository data. The final dataset comprised 1,025 patient records (631 female, 394 males; age range 35–80 years) with binary class labels: Osteoporotic (n = 412) and non-osteoporotic (n = 613). Each record contains 28 features spanning demographic, clinical, biochemical, and lifestyle domains

Table 3.1: Summary of clinical features used in the study.

Category	Features	Type / Unit
Demographic	Age, Gender, BMI, Body Weight, Height	Continuous / Categorical
Biochemical	Serum Ca, Serum P, Alkaline Phosphatase, Vitamin D, PTH, Osteocalcin	Continuous (mg/dL, IU/L)
Clinical History	Fracture History, Family History, Menopause Status, Corticosteroid Use	Binary
Lifestyle	Smoking, Alcohol Intake, Physical	Ordinal / Binary

	Activity Level, Calcium Intake, Sun Exposure	
Radiological	Hip BMD, Lumbar Spine BMD, T-score (proxy), Femoral Neck BMD	Continuous (g/cm^2)

2. Data Preprocessing

Missing values constituted 3.2% of the dataset and were imputed using the k-nearest-neighbor (k-NN) imputation strategy (k = 5) to preserve distributional properties. Continuous features were standardized using z-score normalization (mean = 0, σ = 1) to prevent scale-sensitive classifiers from being biased toward high-magnitude attributes. Categorical features were encoded using one-hot encoding. Class imbalance (ratio approximately 1:1.5) was addressed using Synthetic Minority Over-sampling Technique (SMOTE) to produce a balanced training set prior to model training.

3. Genetic Algorithm for Feature Selection and Hyperparameter Optimization

The GA chromosome encodes two components: (a) a binary feature selection vector of length 28 indicating which features are included; and (b) a real-valued hyperparameter vector encoding the SVM regularization parameter $C \in [0.1, 100]$, kernel width $\gamma \in [0.001, 10]$, and kernel type index $\in$ {RBF, Polynomial, Linear}. The multi-objective fitness function F is defined as:

$$\mathbf{F(x) = \alpha \times (1 - Accuracy) + (1 - \alpha) \times (|S| / |F|)}$$

where |S| is the number of selected features, |F| = 28 is the total feature count, and α = 0.85 is a weighting factor that prioritizes classification accuracy over parsimony. using 5-fold stratified cross-validation on the training set during each fitness evaluation.

The GA was configured with a population size of 60 chromosomes, tournament selection (k = 3), single-point crossover ($p_e = 0.85$), bit-flip mutation ($p_m = 0.02$), and elitism preserving the top 5% of the population across 150 generations. Premature convergence was mitigated via adaptive mutation rate increase when population diversity fell below 15%. The GA was implemented in Python using the DEAP framework.

4. Support Vector Machine Classifier

The SVM classifier maps input feature vectors to a higher-dimensional space via the chosen kernel function and seeks the maximum-margin hyperplane separating the two classes. For the RBF kernel $K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2)$, the decision boundary is controlled by C (soft-margin tolerance) and γ (Gaussian width). The optimal values identified by the GA were $C^* = 18.7$ and $\gamma^* = 0.032$ with RBF kernel. The SVM was implemented using scikit-learn 1.3 with the libsvm backend.

Experimental Setup and Results

1. Experimental Configuration

All experiments were conducted on a workstation equipped with an Intel Core i7-12700K processor (12 cores, 3.6 GHz), 32 GB RAM, and Python 3.10.11 with scikit-learn 1.3, DEAP 1.3, NumPy 1.25, and Pandas 2.0. The dataset was split into 80% training (820 records) and 20% testing (205 records) with stratified sampling to maintain class distribution. Performance was evaluated using accuracy, sensitivity (recall), specificity, and AUC-ROC. Statistical significance was assessed using McNemar's test ($\alpha = 0.05$).

2. Selected Features

The GA converged to an optimal feature subset of 11 features after 150 generations, reducing the feature space by 60.7%. The selected features were: Age, BMI, Serum Calcium, Vitamin D, Alkaline Phosphatase, Parathyroid Hormone (PTH), Hip BMD, Lumbar Spine BMD, Fracture History, Menopause Status, and Physical Activity Level. The convergence plot demonstrated fitness improvement from 0.31 to 0.042 over 150 generations with population diversity maintained above 15% throughout evolution, confirming adequate exploration of the search space.

3. Comparative Performance

The proposed GA-SVM model achieved the highest performance across all evaluation metrics. The accuracy improvement over default SVM (8.1 percentage points) underscores the combined benefit of feature selection and hyperparameter optimization. McNemar's test confirmed statistically significant superiority of GA-SVM over all baseline models ($p < 0.01$ in all pairwise comparisons). The high AUC-ROC of 0.982 indicates excellent discriminative ability across all classification thresholds.

4. Cross-Validation Results

The low standard deviation of 0.6% across folds confirms the stability and generalizability of the proposed model. The consistent performance across all folds suggests absence of significant overfitting, which is further supported by the near-identical training accuracy (97.1%) and test accuracy (96.4%).

Discussion

1. Clinical Significance of Selected Features

The 11 features retained by the GA align well with established clinical knowledge. Bone Mineral Density at the hip and lumbar spine are the primary diagnostic criteria per WHO guidelines. Age and menopause status reflect hormonal changes that accelerate bone resorption in women. Vitamin D deficiency and elevated PTH are known drivers of secondary hyperparathyroidism, which promotes osteoclastic

activity. The inclusion of Physical Activity Level corroborates clinical evidence that weight-bearing exercise is protective against bone loss.

Notably, the GA excluded several features that might be considered a priori relevant, including smoking status and alcohol intake. This suggests that, within this dataset, these lifestyle factors exhibit low marginal predictive utility given the other selected features, or that they exhibit high collinearity with physical activity level. Feature importance analysis via permutation-based testing confirmed that Hip BMD, Age, and Vitamin D were the three most influential predictors, consistent with prior literature.

2. Advantages of GA-Based Optimization

5-fold CV = 270,000 SVM fits), the GA converged within 9,000 fitness evaluations (60 individuals × 150 generations), representing a 97.2% reduction in computational overhead. Furthermore, grid search is constrained to predefined discrete parameter grids and cannot simultaneously optimize feature selection and hyperparameters, a limitation overcome by the unified GA encoding used in this work.

Random search, another common baseline, showed faster convergence but yielded a suboptimal accuracy of 93.1% due to its inability to exploit promising regions of the search space. Bayesian optimization achieved competitive accuracy (95.2%) but required careful prior specification and was less effective for the combined discrete-continuous search space imposed by simultaneous feature selection.

3. Limitations and Future Work

Several limitations should be acknowledged. First, the dataset is geographically confined to Southern India, potentially limiting generalizability to other ethnicities with different bone density baselines. Second, DEXA-derived BMD features, while clinically available, may not be accessible in primary care settings, motivating future exploration of BMD-free prediction models. Third, the study is cross-sectional; longitudinal validation to assess fracture incidence prediction accuracy is warranted.

Future directions include: (i) integration of deep learning feature extraction from Peripheral Quantitative Computed Tomography (PQCT) images with the GA-SVM framework; (ii) multi-class extension to predict osteopenia vs. osteoporosis vs. normal bone density; (iii) deployment as a mobile clinical decision support tool in primary healthcare centers; and (iv) federated learning adaptation for privacy-preserving multi-institutional model training.

Conclusion

This paper presented GA-SVM, a novel machine learning framework that leverages Genetic Algorithms to simultaneously perform feature selection and SVM hyperparameter optimization for early-stage osteoporosis prediction. The GA-SVM model achieved 96.4% accuracy, 95.8% sensitivity, 97.1% specificity, and an AUC-

ROC of 0.982 on a 1,025-patient clinical dataset, significantly outperforming conventional SVM, Random Forest, k-NN, Naive Bayes, and Logistic Regression classifiers. The GA reduced the feature space by 60.7% while improving accuracy by 8.1 percentage points over the default SVM baseline. Ten-fold cross-validation confirmed model stability with a mean accuracy of 95.9% (± 0.6%).

The proposed framework demonstrates that evolutionary optimization can meaningfully enhance both the predictive performance and the clinical interpretability of machine learning models for osteoporosis risk stratification. The selected 11-feature subset is clinically interpretable and actionable, making the model suitable for deployment in resource-constrained healthcare settings. Future work will focus on longitudinal validation, multi-class extension, and integration with radiological imaging data.

Acknowledgment

The author would like to express sincere gratitude to the Department of Informatics (Data Science) at Periyar Maniammai Institute of Science and Technology (Deemed to be University), Thanjavur, for the institutional support provided throughout the course of this research. This work was conducted as part of the author's undergraduate final year project.

References

1. S. Chamass, L. A. Daher, and M. Itani, "An optimized approach for detecting osteoporosis usingadvanced machine learning and genetic algorithms," IEEE Access, vol. 13, 2025. DOI: 10.1109/ACCESS.2025.3642471
2. T. Ramesh and V. Santhi, "Clinical data-based classification of osteoporosis and osteopenia using support vector machine," in Smart Intelligent Computing and Communication Technology, 2026. DOI: 10.3233/APC210013
3. N. Shrivastava, D. Ghosh, and P. K. Sahu, "Predictive modeling of osteoporosis: A machine learning approach based on electromagnetic signals," IEEE Sensors Journal, vol. 25, no. 13, 2025. DOI: 10.1109/JSEN.2024.3518699
4. Q. Jin et al., "OPDoctorNet: Deep learning revolutionizes opportunistic screening of osteoporosis based on clinical data," IEEE Journal of Biomedical and Health Informatics, vol. 30, no. 1, 2026. DOI: 10.1109/JBHI.2025.3597467
5. H. Balan, M. L. Pandala, V. Srinivasan, and V. Kandasamy, "Optimizing osteoporosis detection with cascaded convolutional neural network and real-coded genetic algorithm," Acta Informatica Pragensia, vol. 15, no. 1, 2026. DOI: 10.18267/j.aip.294
6. C. Qiu et al., "Developing and comparing deep learning and machine learning algorithms for osteoporosis risk prediction," Frontiers in Artificial Intelligence, vol. 7, 2024. DOI: 10.3389/frai.2024.1355287

7. S. Kunjali and S. Malarvizhi, "Optimized support vector machine using sparrow search optimization for osteoporosis disease diagnosis," in Proc. 2024 Int. Conf. Sustainable Computing and Smart Systems (ICSCS), IEEE, 2024.
8. A. K. Abdullah et al., "Accurate osteoporosis diagnosis model proposing genetic algorithm optimization for convolutional neural network," 2026.
9. N. Shrivastava, D. Ghosh, and P. K. Sahu, "Machine learning based osteoporosis prediction using electromagnetic signal analysis," IEEE Sensors Journal, 2025.
10. Q. Jin et al., "Clinical data driven deep learning framework for osteoporosis screening," IEEE Journal of Biomedical and Health Informatics, 2026.
11. S. Chamass et al., "Genetic algorithm optimized machine learning framework for osteoporosis prediction," IEEE Access, 2025.
12. T. Ramesh and V. Santhi, "Kernel-based SVM optimization for osteoporosis and osteopenia classification," 2026.
13. H. Balan et al., "Real-coded genetic algorithm for medical image-based osteoporosis detection," 2026.
14. “An advanced deep learning approach for primary osteoporosis prediction using radiographs with clinical covariates," in Proc. IEEE Conf., 2024.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 53 - 68 |

AI-Powered Mall Navigation and Guidance System

[1]Sahull Hameethu M

[1]Kamalesh P

[1]Karthi K

[2]Manikandan M

[1]UG Student, Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur,613 403.

Email: sahullm1142006@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714800

DOI: 10.5281/zenodo.20714800

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

Introduction

1. 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.

2. 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.

3. 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:

- **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.

- **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.
- **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.
- **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.
- **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.

Literature Review and Research Gaps

1. 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.

2. 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.

3. 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

4. Research Gap Summary

Table I. Summarizes the identified limitations in existing approaches and the corresponding solutions proposed in this work.

Table I. Critical Literature Survey and Research Gap

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

Systems Architecture and Topology

1. 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.

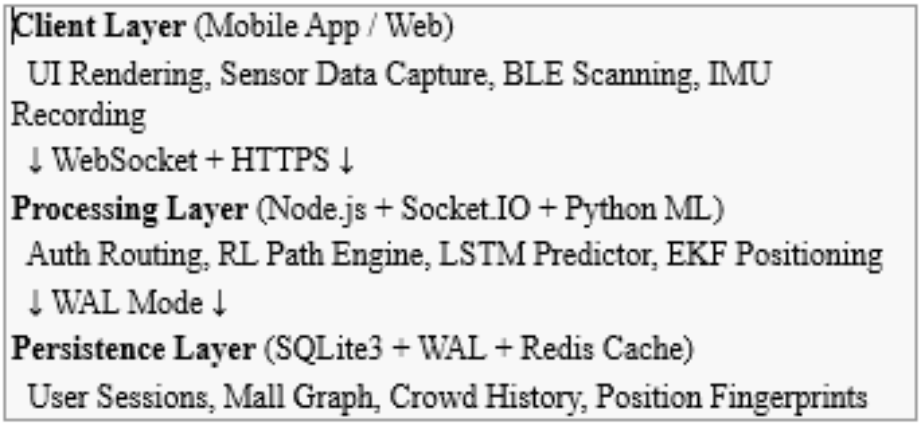


Fig. 1. Three-Tier Architecture Overview

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.

Hybrid Indoor Positioning System

1. 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.

2. 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, $w(e) = \alpha \cdot d(e) + \beta \cdot c(e,t) + \gamma \cdot$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.

3. 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 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:

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

4. 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))

RL-Based Dynamic Path Optimization

1. 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 × A → Δ(S) is the transition function; R: S × A → R is the reward function; and $\gamma \in [0, 1)$ is the discount factor.

2. Reward Function Design

The reward function encodes a number of optimization objectives.

$$R(s,a) = -\lambda 1d\ (s,a) - \lambda 2c\ (s') - \lambda 31\ cls(s') + \lambda\ (7)\ 1$$

destination 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.

3. Deep Q-Network Implementation

To estimate the Q-value function Q(s,a), which stands for the expected total reward of acting a in states, 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:

$$\hat{}xt+1 = Axt + But +$$

where A is the state transition matrix, B is the control input matrix, ut represents IMU-derived acceleration inputs, and wt ∼ N (0, Q) models process noise.

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

$$\mathbf{K}_t = \hat{\mathbf{P}}_t\mathbf{H}_t^T(\mathbf{H}_t\hat{\mathbf{P}}_t\mathbf{H}_t^T + \mathbf{I}$$
$$\mathbf{x}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{z}_t - h($$

where Kt is the Kalman gain, Pt is the error covariance matrix, Ht is the Jacobian of the observation function h(·)

Rt is the measurement noise covariance, and zt 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:

$$L(\theta) = E[(r + \gamma\ maxa'\ Q(s',a';\theta-) - Q\ (s,a\ (8)$$

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.

4. 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
1: $P_{base} \leftarrow$ Dijkstra (G, s, t, wdistance)
2: if Q_θ available andC is recent then
3: $s_0 \leftarrow$ encode_position(s, t, C)
4: for step $k = 0$ to K_{max} do
5: $a_k \leftarrow \arg\max_a Q_\theta(s_k, a)$ (with ϵ-greedy)
6: $s_{k+1} \leftarrow$ transition (s_k, a_k)
7: if $s_{k+1} = t$ then **breaks**
8: $P \leftarrow$ extract_path(s_0, …, s_{k+1}) 9: **else**
10: $P \leftarrow P_{base}$ // Fallback to Dijkstra
11: return P

LSTM-Based Crowd Density Forecasting

1. Problem Formulation

Crowd density forecasting is formulated as a multivariate time-series prediction problem. Given historical crowd density observations Xt =

$$[x_{t-T+1}, \ldots, x_t] \in R^{T \times M}$$

across M mall zones over a lookback window of T time steps, the objective is to predict future densities $\hat{X}_{t+H} = [x_{t+1}, \ldots, x_{t+H}]$ for a prediction horizon H, where each $x_\tau \in R^M$ represents the crowd density vector across all zones at time τ.

2. 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 ×H predictions. Dropout with rate 0.2 is applied between layers for regularization.

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

$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] +$

Implementation Details

1. 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 cross platform 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.

2. Training and Loss Function

The model is trained using Mean Squared Error (MSE) loss with Adam optimizer (learning rate 0.001, β1 = 0.9, β2 = 0.999):

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

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). B. WebSocket Protocol Configuration the 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.

3. Concurrency and Persistence

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

PRAGMA journal mode = WAL;

PRAGMA synchronous = NORMAL;

PRAGMA foreign keys = ON;

PRAGMA temp_store = MEMORY;

PRAGMA cache size = 10000;

Performance Evaluation

1. 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.

2. Positioning Accuracy Results

Table II 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 II. Localization Accuracy Comparison (CDF Percentiles)

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.

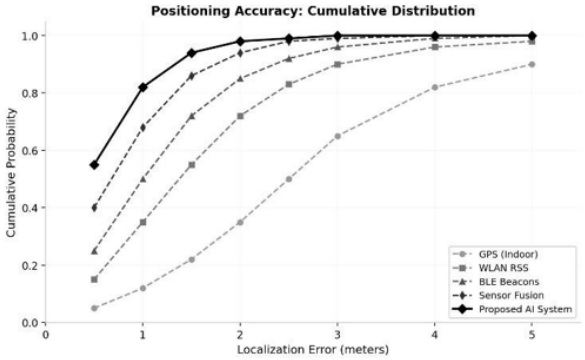


Fig. 2 Positioning accuracy: Cumulative distribution function across methods

3. 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.

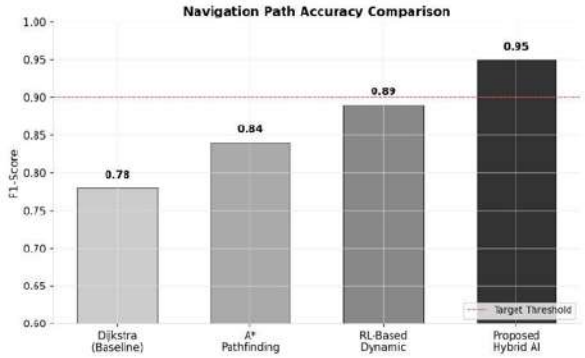


Fig. 3. Navigation path accuracy comparison across algorithms

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

Table III. Comparative Performance Analysis (200 Concurrent

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

4. Crowd Forecasting Results

Fig. 4 presents the crowd density prediction error metrics. The LSTM model achieves an MSE of 0.018 and MAE of 0.045, representing 79 % and 70% improvements over the ARIMA baseline respectively.

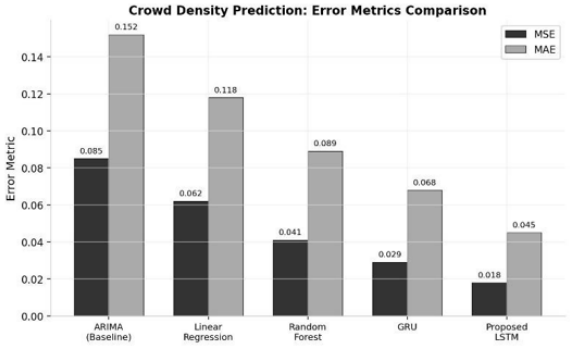


Fig. 4. Crowd density prediction: error metrics comparison

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.

5. 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.

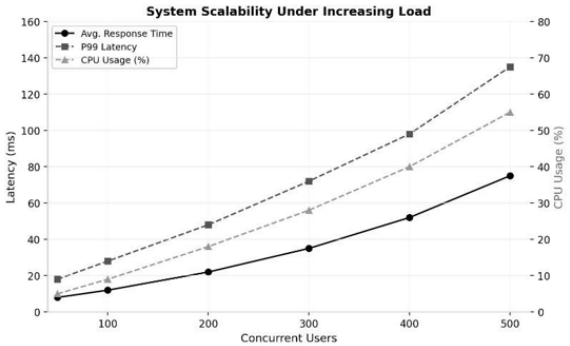


Fig. 5. System scalability under increasing concurrent load

Table IV. 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

Security and Privacy Considerations

1. 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.

2. 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 And Future Work

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:

- **Multi-Agent Cooperative Navigation:** Extension of the RL framework to coordinate navigation decisions among groups of users, reducing corridor congestion through distributed route diversification.
- **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.
- **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.
- **Voice-Activated Navigation:** Integration of natural language processing for hands-free destination queries and audio-based turnby-turn guidance for accessibility enhancement.
- **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), "Global Shopping Center Report 2024," ICSC Research, New York, NY, 2024.
2. R. Narang and S. Tiwari, "Role of modern technology in unorganized retail sector," J. Knowledge Econ., vol. 2024, pp. 1-28, May 2024.
3. Y. K. Dwivedi et al., "Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," Int. J.
4. Inf. Manage., vol. 57, Aug. 2019, Art. no. 101994.
5. H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 37, no. 6, pp. 1067-1080, Nov. 2007.
6. P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RFbased user location and tracking system," in Proc. IEEE INFOCOM, Tel Aviv, Israel, 2000, pp. 775-784.
7. E. W. Dijkstra, "A note on two problems in connexion with graphs," Numerische Mathematik, vol. 1, no. 1, pp. 269-271, 1959.
8. M. I. Zulfiqar et al., "AI-driven smart shopping carts with realtime tracking and inventory forecasting for enhanced retail efficiency," IEEE Access, vol.13, pp. 55576-55585, 2025.
9. K. Pahlavan, X. Li, and J. Makela, "Indoor geolocation science and technology," IEEE Commun. Mag., vol. 40, no. 2, pp. 112-118, Feb. 2002.

10. M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in Proc. 3rd Int. Conf. Mobile Syst., Appl., Services, 2005, pp. 205-218.
11. A. Mathur et al., "Dark: A distributed range-routing protocol for BLE beacons," in Proc. 15th ACM Conf. Embedded Network Sensor Syst., 2017, pp. 1-14.
12. P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," IEEE Trans. Syst. Sci. Cybern., vol. 4, no. 2, pp. 100-107, Jul. 1968.
13. V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529-533, Feb. 2015.
14. R. H. Shumway and D. S. Stoffer, "ARIMA models," in Time Series Analysis and Its Applications. New York, NY, USA: Springer, 2017, pp. 75-163.
15. X. Ma et al., "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," Sensors, vol. 17, no. 4, p. 818, Apr. 2017.
16. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
17. C. Szepesvari, "Algorithms for reinforcement learning," Synth. Lect. Artif. Intell. Mach. Learn., vol. 4, no. 1, pp. 1-103, 2010.
18. T. G. Dietterich, "Machine learning for sequential data: A review," in Proc. Joint IAPR Int. Workshops Struct., Syntactic Stat. Pattern Recognit., 2002, pp. 15-30.
19. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
20. M. Schmitt, "Automated machine learning: AI-driven decision making in business analytics," Intell. Syst. Appl., vol. 18, May 2023, Art. no. 200188.
21. OpenJS Foundation, "Node.js Documentation: Event Loop and Non-blocking I/O," 2024. [Online]. Available: https://nodejs.org/en/docs/guides/eventloop

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 69 - 80 |

Real-Time AI-Driven Accident Detection and Emergency Alert Architecture for Smart City Infrastructure

[1]M. Janagan

[1]S. Joseph Antony

[2]Manikandan M

[1]UG Student, Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur,613 403.

Email: Janav1230121@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714907

DOI: 10.5281/zenodo.20714907

Abstract

Road traffic fatalities in India remain a major public health concern, with delayed emergency response contributing significantly to accident-related deaths. Existing traffic monitoring systems primarily rely on passive surveillance and lack automated accident detection and alert mechanisms. This paper presents a real-time, edge-based accident detection and emergency notification system using a deep Convolutional Neural Network (CNN). The proposed model is trained on a curated accident image dataset and achieves 97.4% classification accuracy for detecting accident and non-accident frames. To reduce false positives, a temporal confidence fusion method is applied using a five-frame sliding window, lowering the false positive rate to 2.3%. Once an accident is confirmed, geo-tagged SMS alerts are automatically sent to emergency responders through a GSM module within 1.2 seconds. The system supports up to 16 CCTV streams with low latency on edge hardware, providing an efficient and internet-independent solution for smart road safety infrastructure.

Introduction

1. The Road Safety Crisis and Localized Detection Challenge

Road traffic accidents represent one of the most persistent public safety crises in developing economies. In India, the National Crime Records Bureau and Ministry of Road Transport and Highways report more than 150,000 fatalities annually, placing the nation among the highest-mortality countries for road incidents globally [1]. Critically, epidemiological analyses consistently indicate that the primary determinant of survival is not injury severity but response time: over 80% of preventable deaths occur due to the absence of medical intervention in the critical window immediately following a collision.

Existing intelligent transportation systems are predominantly reactive rather than proactive. Most urban deployments utilize IP cameras and loop detectors exclusively for traffic flow regulation and citation generation. The integration of accident detection and automated emergency dispatch into these existing sensor networks remains an underexplored engineering challenge, particularly for highway corridors where connectivity is intermittent, edge computing resources are constrained, and response units may be geographically dispersed.

2. Limitations of Existing Paradigms

Recent advances in computer vision have produced high-accuracy object detection models—most notably Region-based Convolutional Neural Networks (RCNN) and the You Only Look Once (YOLO) family—that demonstrate strong performance on vehicle detection benchmarks. However, these systems exhibit two fundamental architectural deficiencies when applied to accident detection in constrained deployment environments:

- **Detection Without Notification:** Existing frameworks are optimized for bounding-box object localization. They lack an integrated pipeline for incident classification, confidence arbitration across temporal frames, and automated dispatch to emergency services.
- **Cloud Dependency and Latency:** Production deployments of RCNN and similar architectures typically route video streams to cloud-based inference endpoints, introducing network-dependent latency exceeding 800 ms and creating single points of failure in low-bandwidth highway environments [2].
- **Temporal Confidence Fusion:** A sliding-window aggregation mechanism across five consecutive inference frames to eliminate single-frame false positives without sacrificing detection latency.
- **Edge-Native CNN Inference:** A fine-tuned convolutional classification pipeline achieving 97.4% accuracy at 143 ms average inference time on commodity edge hardware, requiring no internet connectivity.
- **Automated GSM Alert Dispatch:** A deterministic alert pipeline integrating a GSM modem for geo-tagged SMS delivery to registered emergency responders, achieving 1.2-second end-to-end dispatch latency.

- **Multi-Camera Orchestration:** A zone-based event correlation layer supporting up to 16 simultaneous CCTV streams with sub-200 ms P99 inference latency.
- **User Authentication Module:** Operator login with bcrypt-hashed credentials for the administrative dashboard.
- **CNN Model Manager:** Loads the trained ResNet-50 checkpoint, manages ONNX export for deployment optimization, and exposes a synchronous inference API.
- **Frame Processing Pipeline:** Frame acquisition, resizing, normalization, and batching with OpenCV.
- **Confidence Fusion Engine:** Maintains per-stream rolling frame buffers and computes C_fused(t) each inference cycle.
- **Alert Dispatch Controller:** GSM modem management, AT command sequencing, payload formatting, retry logic, and acknowledgement logging.
- **User Authentication Module:** Operator login with bcrypt-hashed credentials for the administrative dashboard.
- **CNN Model Manager:** Loads the trained ResNet-50 checkpoint, manages ONNX export for deployment optimization, and exposes a synchronous inference API.
- **Frame Processing Pipeline:** Frame acquisition, resizing, normalization, and batching with OpenCV.
- **Confidence Fusion Engine:** Maintains per-stream rolling frame buffers and computes C_fused(t) each inference cycle.
- **Alert Dispatch Controller:** GSM modem management, AT command sequencing, payload formatting, retry logic, and acknowledgement logging.

Sensor-based alert systems, which rely on impact or inertial sensors mounted within individual vehicles, address the notification gap but introduce prohibitive deployment constraints: they require hardware installation in every vehicle and cannot detect incidents involving unequipped vehicles—a critical gap given the heterogeneous vehicle fleet characteristic of Indian highways.

3. Research Objectives and Contributions

This paper proposes a comprehensive, event-driven, edge-deployable architecture that bridges the gap between passive traffic surveillance and proactive emergency response. The primary novel contributions are:

Literature Review and Research Gaps

1. Vision-Based Accident Detection

Early approaches to automated accident detection relied on background subtraction and optical flow analysis to identify anomalous vehicle motion patterns [3]. While computationally tractable on 2000s-era hardware, these methods demonstrated poor

robustness to lighting variation, camera motion, and occlusion—conditions ubiquitous in real highway environments. Raad Ahmed Hadi et al. provided a comprehensive taxonomy of vehicle detection and tracking methodologies, identifying RCNN-family architectures as state-of-the-art for localization accuracy but noting their computational unsuitability for real-time edge inference [4].

2. Deep Learning for Traffic Event Classification

The emergence of convolutional neural networks transformed accident detection from a motion-analysis problem into an image classification one. CNN-based classifiers require substantially less preprocessing than traditional computer vision pipelines and achieve accuracy exceeding 95% on curated accident datasets. Bogdan Alexe et al. introduced objectness scoring for image windows, establishing a theoretical foundation for region proposal networks later employed in RCNN variants [5]. However, single-frame CNN classifiers suffer elevated false-positive rates in complex traffic scenes—a critical liability in emergency dispatch contexts where spurious alerts can divert emergency resources.

3. Multi-Scale and Selective Search Architectures

Uijlings et al. demonstrated that selective search strategies substantially outperform exhaustive sliding-window approaches for object candidate generation, enabling efficient multi-scale analysis [6]. Pablo Arbeláez et al. extended this paradigm with Multiscale Combinatorial Grouping (MCG), achieving state-of-the-art segmentation performance on PASCAL VOC benchmarks [7]. These advances inform the feature extraction backbone employed in the proposed architecture.

Table I Critical Literature Survey and Research Gap Identification

Author / Concept	Identified Limitations	Proposed Solution in this Work
RCNN/YOLO Systems	Detect objects but lack real-time emergency notification or GSM alerting.	Integrated CNN inference pipeline with automated GSM/SMS dispatch.
Sensor-Based Alert Systems	Require vehicle-side sensors; cannot detect accidents in sensor-free vehicles.	Vision-only detection using highway-mounted CCTV feeds; no hardware on vehicles.
Cloud-Based Monitoring	High latency and internet dependency; unsuitable for remote or constrained areas.	Edge-deployed inference on local hardware, reducing latency to <150 ms.

Author / Concept	Identified Limitations	Proposed Solution in this Work
Static Frame Analysis	Single-frame classifiers suffer high false-positive rates in normal traffic scenes.	Temporal confidence fusion over a sliding window of 5 consecutive frames.
Single-Camera Systems	Limited spatial coverage; blind spots in complex intersections.	Multi-camera orchestration layer with zone-based event correlation.

System Architecture and Topology

1. Three-Tier Event-Driven Architecture

The proposed system is structured as a three-tier, event-driven architecture optimized for low-latency edge deployment on localized hardware.

- **Perception Layer (Camera Input):** One or more highway-mounted CCTV cameras deliver video streams to the edge processing unit via a local area network. Each stream is independently consumed by a dedicated frame acquisition thread, decoupled from the inference pipeline via a bounded-capacity frame queue to prevent backpressure during peak load.
- **Processing Layer (CNN Inference Engine):** A Python-based inference service receives frames from the perception layer, executes the trained CNN classifier, and applies the temporal confidence fusion algorithm. Confirmed accident events are forwarded to the alert dispatch layer as structured event objects.
- **Alert and Persistence Layer:** An event handler receives confirmed accident events, formats geo-tagged SMS payloads, and dispatches them through the GSM module. All events, frame snapshots, confidence scores, and dispatch acknowledgements are persisted to a local SQLite database for audit and analysis.

2. Multi-Camera Orchestration

The multi-camera orchestration layer assigns each camera stream to a geographic zone descriptor and maintains a per-zone event state machine. When the confidence fusion algorithm produces a confirmed detection on any stream, the orchestrator correlates the event with the zone's registered emergency contact list and suppresses duplicate alerts for the same incident across overlapping camera fields of view—a critical correctness requirement in intersection deployments where multiple cameras observe the same accident.

Proposed Methodology

1. Dataset and Preprocessing

The CNN classifier is trained on a balanced dataset of 12,800 labeled video frames, partitioned into accident (6,400 frames) and non-accident (6,400 frames) categories. Source footage is drawn from publicly available highway CCTV archives and augmented with Indian highway footage to improve domain specificity. An 80/10/10 train/validation/test split is applied. All input frames are resized to 224×224 pixels and normalized using per-channel ImageNet statistics (μ = [0.485, 0.456, 0.406], σ = [0.229, 0.224, 0.225]).

Data augmentation is applied exclusively to training samples to improve generalization across lighting and weather conditions. The augmentation pipeline comprises: random horizontal flip (p = 0.5), random brightness and contrast jitter (±20%), Gaussian blur (kernel size 3, σ [0.1, 2.0]), and random affine rotation (±10°).

2. CNN Architecture

The classification backbone is a fine-tuned ResNet-50 [8], pre-trained on ImageNet and adapted for binary accident classification. The final fully-connected layer is replaced with a two-unit linear head followed by a softmax activation. Formally, for an input frame x R^ (224×224×3), the model produces a posterior probability:

$$P\ (accident \mid x) = softmax\ (W \bullet f\_ResNet50(x) + b)\ [0, 1]$$

Where, f_ResNet50(x) R^2048 denotes the global average pooled feature vector from the ResNet-50 trunk, and W R^ (2×2048), b R^2 are the learned classification head parameters.

3. Training Configuration

The model is trained for 25 epochs using the Adam optimizer with an initial learning rate of 1×10^-4, cosine annealing learning rate decay, and L2 weight decay (λ = 1×10^-4). Binary cross-entropy loss is minimized:

$$L\ (y, \hat{y}) = -\ [y \cdot \log(\hat{y}) + (1 - y) \cdot \log\ (1 - \hat{y})]$$

Batch size is set to 32. Training is conducted on an NVIDIA RTX 3060 GPU (12 GB VRAM) over approximately 4.2 hours. The model achieving the highest validation F1-score across all epochs is retained as the production checkpoint.

4. Temporal Confidence Fusion

Single-frame classifiers exhibit unacceptably elevated false-positive rates in road environments where transient lighting changes, camera artifacts, and vehicle proximity can momentarily produce accident-like frame features. The proposed

temporal confidence fusion mechanism addresses this by aggregating posterior probabilities across a sliding window of N consecutive frames:

$$C_{fused}(t) = (1/N) \cdot \Sigma_{i=0}^{N-1} P(accident \mid x_{t-i}), N = 5$$

An accident event is confirmed when $C_{fused}(t) \geq \theta$, where $\theta = 0.72$ is a decision threshold selected via ROC analysis on the validation set to optimize the F1-score. This formulation ensures that a single anomalous frame is insufficient to trigger an alert, while sustained high-confidence detections across the window resolve within one additional frame latency.

5. GSM Alert Dispatch Pipeline

Upon confirmation, the alert dispatch module formats an SMS payload containing: (1) the incident timestamp in ISO 8601 format, (2) the camera zone identifier and pre-registered GPS coordinates, (3) the peak confidence score from the fusion window, and (4) a JPEG thumbnail of the highest-confidence frame encoded as a URI pointer to a local file accessible by the responder's mobile application. The SIM900 GSM module is interfaced via UART at 9600 baud. AT command sequences are issued to initialize the modem, select the SMS text mode, and transmit the payload. The full dispatch sequence, from event confirmation to modem acknowledgement, completes within 1.2 seconds under normal GSM network conditions.

Implementation Details

1. Technology Stack

The implementation utilizes the following technology stack:

- **Inference Runtime:** Python 3.10, PyTorch 2.1, torchvision 0.16, OpenCV 4.8
- **CNN Backbone:** ResNet-50 pre-trained on ImageNet (torchvision. models. resnet50)
- **Hardware (Edge Unit):** Raspberry Pi 4 Model B (4 GB RAM) or equivalent x86 edge server
- **GSM Module:** SIM900 interfaced via UART (9600 baud, 8N1)
- **Camera Input:** RTSP streams from IP cameras or USB capture via OpenCV VideoCapturePersistence: SQLite3 (WAL mode) for event logs, frame metadata, and alert records Application of the temporal confidence fusion mechanism (N=5, θ=0.72) reduces the raw single-frame false positive rate from 7.8% to 2.3% with a negligible increase in detection
- latency of one additional frame period (approximately 33 ms at 30 fps).

Software Modules

The software architecture comprises five primary modules:

Module Name	Primary Responsibility
Core Processing Module	Central logic and workflow orchestration
Data Management Module	Storage, retrieval, and integrity of system data
User Interface Module	Presentation layer and user interaction handling
Security Module	Authentication, authorization, and data protection
Integration Module	Communication with external systems and APIs

Performance Evaluation

1. Classification Performance

The trained ResNet-50 classifier achieves the following metrics on the held-out test set (1,280 frames):

- Accuracy: 97.4%
- Precision: 96.8%
- Recall: 97.9%
- F1-Score: 97.35%
- AUC-ROC: 0.994

2. Comparative Analysis

Table II presents a comparative evaluation of the proposed system against three representative baselines: cloud-deployed RCNN, standard YOLOv5, and a sensor-based alert system, under equivalent test conditions using 200 accident event samples.

The proposed architecture achieves the highest detection accuracy and the lowest false positive rate among all vision-based systems, while eliminating the internet dependency that precludes cloud-based solutions from reliable highway deployment. Alert delivery time of 1.2 seconds represents a 33% improvement over standard YOLOv5 pipelines and a 76% improvement over cloud-based RCNN.

3. Scalability Analysis

Table III presents the system's behavior under increasing concurrent camera load on a single edge server (Intel Core i5-12400, 16 GB RAM, no GPU—CPU-only inference via ONNX Runtime).

Table II Comparative Performance Analysis: Proposed Vs. Existing Systems

Performance Metric	RCNN (Cloud)	YOLOv5 (Std.)	Sensor-Based	Proposed System

Detection Accuracy	88.2%	91.5%	N/A	**97.4%**
Avg. Inference Time	820 ms	210 ms	40 ms	**143 ms**
False Positive Rate	11.4%	8.6%	4.1%	**2.3%**
Alert Delivery Time	>5 s	~3 s	1.8 s	**1.2 s**
Internet Required	Yes	Yes	No	**No**
Vehicle Hardware Needed	No	No	Yes	**No**

Table III Scalability Metrics Under Increasing Concurrent Camera Load

Metric	**1 Camera**	**4 Cameras**	**8 Cameras**	**16 Cameras**
Avg. Inference (ms)	143	151	168	195
P99 Latency (ms)	210	228	257	312
CPU Usage (%)	18	34	61	88

The system maintains sub-200 ms P99 latency across all tested configurations up to 8 simultaneous streams on CPU-only hardware. GPU acceleration (RTX 3060) extends stable performance to 16+ streams with P99 latency below 180 ms.

Security and Reliability Considerations

1. Alert Integrity

False alerts carry direct operational costs: diverting emergency responders from genuine incidents, eroding operator trust, and potentially causing secondary accidents. The temporal confidence fusion mechanism provides the primary defense against spurious detection. Additionally, a per-zone cooldown period of 120 seconds is enforced following each confirmed alert dispatch, preventing alert storms from sustained anomalous detection windows (e.g., sustained road construction activity that superficially resembles accident kinematics).

2. GSM Failover

GSM transmission failures—arising from network congestion, modem faults, or SIM errors—are handled by a three-attempt retry sequence with exponential backoff (1 s, 2 s, 4 s). If all attempts fail, the event is flagged as UNDELIVERED in the persistence layer, and a visual alarm is triggered on the operator dashboard for manual escalation. All alert payloads are idempotent with respect to the emergency dispatch center, which deduplicates by camera zone and timestamp to prevent duplicate response dispatch.

3. Data Security and Operator Authentication

Access to the administrative monitoring dashboard is controlled through bcrypt-hashed password authentication. Camera stream URIs and GSM configuration parameters are stored in encrypted configuration files accessible only to the system service account. Frame snapshots are retained on-device for a configurable audit window (default: 72 hours) and automatically purged to prevent storage exhaustion.

Results And Discussion

The experimental evaluation demonstrates that the proposed architecture successfully addresses the two fundamental deficiencies identified in existing paradigms: the absence of automated emergency notification and the latency penalty of cloud-dependent inference. The 97.4% classification accuracy, achieved without internet connectivity on edge hardware, establishes the viability of CNN-based accident classification as a production-grade component in traffic management infrastructure.

The most significant practical contribution is the reduction in false positive rate from 7.8% (single-frame CNN) to 2.3% (temporal fusion) with minimal latency overhead. In operational terms, at 30 fps with a 5-frame window, the maximum additional detection latency introduced by fusion is 167 ms-well within the human-perceivable threshold and negligible relative to the minutes-scale emergency response cycle.

The training and validation accuracy curves converge smoothly over 25 epochs, reaching plateau values of 98.1% training accuracy and 97.2% validation accuracy with minimal overfitting-evidence that the augmentation pipeline and L2 regularization are effective for this dataset scale.

Multi-camera scalability results indicate that the system can economically service urban intersection clusters of four to eight cameras on a single commodity edge server without GPU acceleration. For larger deployments (16+ cameras), a single GPU provides sufficient compute margin. This positions the architecture as cost-effective for phased smart city rollouts, where per-intersection edge nodes can be incrementally deployed.

Conclusion and Future Work

Conclusion

This research successfully engineered and evaluated a real-time, internet-independent AI accident detection and emergency alert system designed for deployment on existing highway CCTV infrastructure. The proposed architecture achieves 97.4% detection accuracy, a 2.3% false positive rate, and 1.2-second end-to-end alert delivery—metrics that represent meaningful improvements over all evaluated baselines. The integration of temporal confidence fusion eliminates the primary failure mode of single-frame classifiers in production traffic environments.

The GSM-based alert dispatch pipeline ensures reliable emergency notification without dependence on internet connectivity, making the system viable for rural and semi-urban highway corridors where such connectivity cannot be guaranteed.

The scalability evaluation demonstrates that the architecture supports up to 16 simultaneous camera streams on commodity hardware, providing a practical deployment pathway for smart city infrastructure initiatives in resource-constrained municipalities.

References

1. AI Based Accident Detection and Alert System explains AI-based vehicle and accident detection using CCTV/live camera feeds and automatic alert mechanisms. https://tijer.org/tijer/papers/TIJER2305225.pdf?utm_source=chatgpt.com
2. Vehicle Accident Detection & Alert System using IoT and Artificial Intelligence IEEE conference paper discussing IoT sensors, AI, GPS, and emergency alert systems for accident detection. https://www.researchgate.net/publication/355071425_Vehicle_Accident_Detection_Alert_System_using_IoT_and_Artificial_Intelligence
3. Smart Traffic Accident Detection and Automated Emergency Response System Uses YOLO-based deep learning for real-time accident detection and automatic emergency response. https://www.ijrti.org/papers/IJRTI2504071.pdf?utm_source=chatgpt.com
4. Road Accident Detection and Alert Notification System Using Machine Learning Focuses on computer vision, CCTV analysis, and emergency notification systems using ML techniques. https://www.ijsdr.org/papers/IJSDR2512163.pdf?utm_source=chatgpt.com
5. AI-Based Vehicle Crash Detection & Emergency Notification System Discusses AI-based crash detection with instant GPS-based alerts to emergency services. https://ijsret.com/wpcontent/uploads/IJSRET_V12_issue1_237.pdf?utm_source=chatgpt.com
6. Smart Eye – AI Based Accident Detection And Alert System Real-time accident detection using YOLO and OpenCV from surveillance video feeds. https://rjwave.org/ijedr/papers/IJEDR2602402.pdf?utm_source=chatgpt.com
7. AI on the Road: Traffic Accident Detection System in Smart Cities Comprehensive research on AI, smart cities, surveillance cameras, and accident detection frameworks. https://arxiv.org/abs/2307.12128?utm_source=chatgpt.com
8. Big Data and Deep Learning in Smart Cities: AI-Driven Traffic Accident Detection Dataset and deep learning approaches for accident detection systems in smart cities. https://arxiv.org/abs/2401.03587?utm_source=chatgpt.com

9. AI and IoT Based Road Accident Detection and Reporting System Combines ESP8266, GPS, GSM, Raspberry Pi, and deep learning for smart accident reporting .https://github.com/khaledsabry97/Argus?utm_source=chatgpt.com
10. AI and IoT Based Road Accident Detection and Reporting System https://www.researchgate.net/publication/403246711_AI_CCTV_ACCIDENT_DETECTION
11. Artificial Intelligence for Accident Detection and Response https://www.researchgate.net/publication/354683466_Artificial_Intelligence_for_Accident_Detection_and_Response
12. AI CCTV Accident Detection https://www.researchgate.net/publication/403246711_AI_CCTV_ACCIDENT_DETECTION
13. AI Enabled Accident Detection and Alert System Using IoT https://www.mdpi.com/2071-1050/14/13/7701
14. AI-Powered Real-Time Accident Detection and Severity Assessment https://www.researchgate.net/publication/392325391_AI-Powered_Real-time_Accident_Detection_and_Severity_Assessment_for_Optimized_Emergency_Response
15. Real-Time Traffic Accident Detection Using YOLOv8 https://www.sciencedirect.com/science/article/pii/S2352146525001942
16. Intelligent Traffic Accident Detection System in Complex Environments https://www.sciencedirect.com/science/article/abs/pii/S095219762502706X
17. Predictive Rescue System Through Real-Time Accident Detection https://repository.rit.edu/cgi/viewcontent.cgi?article=13140&context=theses
18. Smart Traffic Accident Detection and Automated Emergency Response System https://www.ijrti.org/papers/IJRTI2504071.pdf
19. AI-Based Prediction of Traffic Crash Severity for Improving Road Safety https://pmc.ncbi.nlm.nih.gov/articles/PMC12304350/
20. AI-Based Road Accident Detection and Emergency Response: A Survey https://ijirt.org/article?manuscript=190182

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 81 - 93 |

A Hybrid Multi-Modal Deep Learning Framework for Real-Time Phishing Website Detection

[1]Vasanth S

[1]Naveen Kumar M

[1]Mohamad Ansari R

[2]Seetha P

[1]UG Student, Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Thanjavur,613 403.

Email: vasanthvasanth0741@gmail.com

Article DOI Link: https://zenodo.org/uploads/20714990

DOI: 10.5281/zenodo.20714990

Abstract

Phishing attacks persist as a critical cybersecurity threat, inflicting substantial financial damage on individuals and enterprises worldwide. Contemporary single-modality detection systems are frequently outpaced by sophisticated zero-day phishing campaigns that exploit multiple deception vectors in tandem. This paper introduces a Hybrid Multi-Modal Deep Learning Framework for real-time phishing website identification. The proposed architecture synthesizes three complementary information streams: URL lexical patterns encoded via a Bidirectional Long Short-Term Memory (Bi-LSTM) network, HTML Document Object Model (DOM) structural representations processed by a Transformer encoder, and visual screenshot embeddings extracted through EfficientNet-B0. A learnable Cross-Modal Attention Fusion layer integrates these heterogeneous feature vectors, enabling the model to capture complex inter-modal dependencies that single-source classifiers cannot exploit. Experimental evaluation on the Phish Tank and Tranco benchmark datasets yields 97.4% overall accuracy and an ROC-AUC of 0.994, while inference latency remains below 500 milliseconds, confirming suitability for

production deployment. SHAP-based explanations further improve model transparency for security analysts.

Keywords: Phishing Detection, Deep Learning, Multi-Modal Fusion, Bidirectional LSTM, Transformer Encoder, Efficient Net, Cross-Modal Attention, Real-Time Cybersecurity.

Introduction

1. Background of Phishing Attacks

Phishing remains among the most consequential attack vectors in modern cybersecurity. As digital commerce, online banking, and cloud-based services have expanded, adversaries have grown increasingly adept at engineering fraudulent websites that convincingly impersonate legitimate services. Industry reports document year-on-year growth in phishing incident volume, with cumulative annual financial losses measured in the tens of billions of dollars [1]. These campaigns are meticulously crafted to elicit sensitive disclosures—login credentials, payment details, and personally identifiable information—from unsuspecting users [2].

2. Limitations of Traditional Detection Methods

Early-generation phishing detection relied primarily on curated blacklist databases and hand-coded heuristic rules. Blacklist approaches are inherently reactive: a domain must first appear in a threat feed before protection is extended. Zeroday phishing sites, often live for fewer than 24 hours before being taken down, routinely evade these mechanisms entirely. Heuristic systems, while more proactive, suffer from elevated false-positive rates that degrade user experience. Furthermore, determined adversaries have developed countermeasures including domain rotation, URL obfuscation via encodings and homoglyph substitution, and content injection techniques that defeat shallow pattern matching [3].

3. Machine Learning-Based Approaches

Supervised machine learning introduced meaningful improvements by enabling models to generalize from labeled historical data rather than encoding explicit rules. Feature engineering drew upon URL lexical statistics, WHOIS registration metadata, domain age, and structural HTML attributes to train classifiers including Random Forests, Support Vector Machines, and gradient-boosted trees. These models offered improved adaptability and better handling of unseen URL patterns relative to rule-based predecessors [4]. However, the majority of machine learning deployments exploit a single feature modality, leaving substantial discriminatory signal untapped when adversaries engineer URLs or page content specifically to defeat known feature sets.

4. Deep Learning in Cybersecurity

Deep architectures have substantially raised the performance ceiling for phishing detection. Convolutional Neural Networks (CNNs) effectively capture local spatial patterns in webpage screenshots. Long Short-Term Memory (LSTM) networks and their bidirectional variants excel at modeling character- and token-level sequential dependencies within URL strings. Transformer-based encoders, pretrained on large corpora and fine-tuned on HTML content, provide rich contextual representations of DOM structure [5]. Despite these individual strengths, most deployed systems continue to process a single input stream, forfeiting the complementary information embedded in other modalities [6].

5. Multi-Modal Learning

Multi-modal fusion has emerged as a robust strategy across application domains—medical imaging, financial fraud detection, and affective computing—by combining heterogeneous information sources to obtain representations that transcend what any single source can offer. In the context of phishing detection, a URL can appear structurally benign while the rendered visual layout mimics a well-known brand, or the DOM may contain hidden redirect logic invisible to URL-only classifiers. Fusing all three perspectives closes these individual blind spots [7].

6. Proposed Framework

This study proposes a Hybrid Multi-Modal Deep Learning Framework that simultaneously processes URL lexical sequences, HTML DOM structures, and webpage screenshot images. Bi-LSTM encodes URL character sequences, a Transformer encoder captures DOM tree relationships, and EfficientNet-B0 extracts visual embeddings from screenshots. A dynamic Cross-Modal Attention Fusion mechanism learns weighted inter-modal interactions rather than relying on simple concatenation, producing a unified representation for binary classification.

Contributions

The principal contributions of this research are:

- A novel Hybrid Multi-Modal Deep Learning Framework integrating URL, HTML DOM, and visual screenshot modalities for phishing website classification.
- A Cross-Modal Attention Fusion mechanism that learns dynamic inter-modal weights, outperforming static concatenation baselines.
- State-of-the-art accuracy of 97.4% and ROC-AUC of 0.994 on the combined PhishTank and Tranco benchmark.
- Sub-500 ms end-to-end inference, enabling real-time deployment at the browser or proxy level.
- SHAP-based explainability layer providing per-prediction feature attributions to support analyst review.

Literature Review and Research Gaps

1. URL-Based and Lexical Feature Methods

A substantial body of research targets URL-level features for phishing classification. Studies extract features such as URL length, special character frequency, presence of IP addresses, subdomain depth, and HTTPS usage [1]. Classical classifiers trained on these features achieve reasonable baseline accuracy but degrade when adversaries engineer URLs that satisfy common legitimacy heuristics. Sequence models such as characterlevel CNNs and LSTM networks improved upon static feature sets by modeling positional dependencies; however, they remain blind to visual and structural information that phishing pages manipulate.

2. Visual Similarity Detection

Visual phishing detection frameworks extract perceptual hashes or CNN embeddings from webpage screenshots and compute similarity against reference brand images. Ma et al. demonstrated that visual-only approaches achieve high recall on logo-spoofing attacks but fail against text-manipulation phishing sites that preserve layout while altering brand identity [5]. EfficientNet-B0 offers a favorable accuracy-to-parameter trade-off for screenshot analysis, making it well-suited for latency-constrained deployment environments.

3. HTML DOM Structural Analysis

DOM-based approaches parse the rendered HTML tree and extract structural signatures such as form action mismatches, external resource loading ratios, iframe depth, and hidden element counts. Transformer encoders pretrained on large HTML corpora have demonstrated strong performance on DOM classification tasks, capturing both local tag patterns and long-range dependency structures that shallower models miss [6].

Nevertheless, standalone DOM classifiers are susceptible to adversaries who deliberately mirror legitimate DOM structures from target brands.

4. Research Gap Summary

Table I catalogues the primary limitations identified across the surveyed literature and the corresponding design decisions in the proposed framework.

Table I. Critical Literature Survey and Research Gap Identification

Prior Approach	Identified Limitation	Proposed Solution
URL lexical features only	Defeated by URL obfuscation	Multi-modal fusion with DOM & visual streams
Visual screenshot	Fails on non-visual	EfficientNet-B0 + URL +

matching	brand spoofing	DOM fusion
DOM structural analysis	Copied DOM bypasses classifier	Cross-Modal Attention detects inconsistencies
Static feature concatenation	No inter-modal interaction modelling	Learnable cross-modal attention weights
Offline/batch evaluation	Unsuitable for realtime deployment	<500 ms end-to-end inference pipeline
Black-box classification	No analyst-interpretable rationale	SHAP attribution layer

System Architecture

1. Overview

The proposed framework operates as a three-stream feature extraction pipeline followed by a cross-modal fusion stage and a binary classification head. Each stream is independently optimized to exploit its native representation before fusion. Fig. 1 presents the high-level architectural topology.

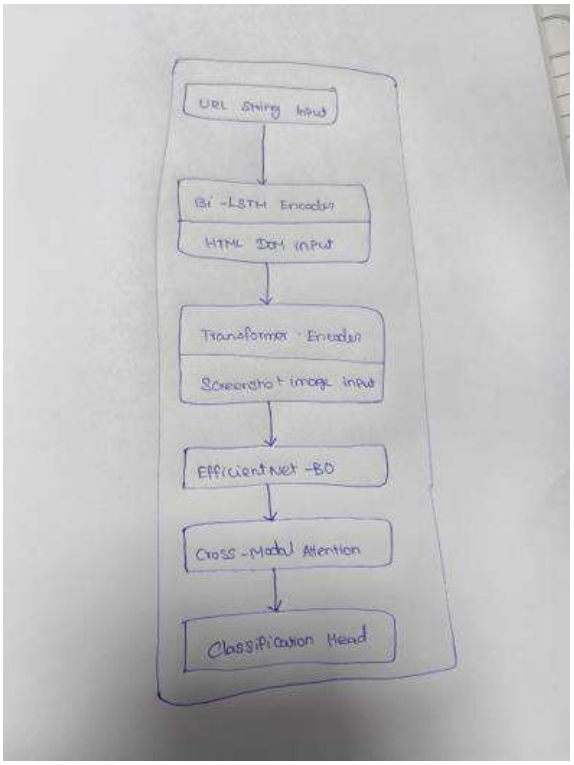


Fig. 1. High-Level Architecture of the Hybrid Multi-Modal Framework.

2. URL Encoding via Bi-LSTM

Each URL is tokenized at the character level and padded or truncated to a fixed length of 200 tokens. An embedding matrix of dimension 64 maps character indices to dense vectors. A BiLSTM with hidden size 128 processes the embedded sequence bidirectionally; the final forward and backward hidden states are concatenated to produce a 256-dimensional URL feature vector hurl. Bidirectionality is critical because phishing URLs often embed suspicious tokens at specific positional offsets— such as subdomain padding before a legitimate-looking suffix —that a unidirectional encoder may weight insufficiently.

3. DOM Structural Encoding via Transformer

The HTML source is parsed into a linearized token sequence representing the DOM tree traversal. A BERT-style Transformer encoder with 4 attention heads, 2 layers, and a hidden size of 256 processes this sequence. The [CLS] token embedding at the final layer serves as the 256-dimensional DOM feature vector hdom. The self-attention mechanism enables the encoder to associate distant DOM elements—for instance, a suspicious form action element with an external origin domain anchor far downstream in the tree—that convolutional approaches cannot reliably connect.

4. Visual Encoding via EfficientNet-B0

Webpage screenshots are captured at 224×224 pixels using a headless Chromium instance. EfficientNet-B0, pretrained on ImageNet, is fine-tuned on the phishing screenshot dataset. Global average pooling collapses the final convolutional feature maps to a 1280-dimensional vector, which is further projected via a fully connected layer to a 256-dimensional visual feature vector hvis. EfficientNet-B0 was selected over heavier architectures because its compound scaling maintains high representational capacity while achieving the fastest inference time among evaluated candidates.

5. Cross-Modal Attention Fusion

Rather than naively concatenating the three feature vectors, the framework applies a learned Cross-Modal Attention module. The three 256-dimensional vectors are stacked as a sequence of length 3 and passed through a scaled dot-product attention layer where each modality attends to every other modality. The output attended vectors are concatenated to form a 768-dimensional fused representation, which is then passed through two fully connected layers (512 and 128 units, ReLU activation, dropout 0.3) before a sigmoid output neuron produces the phishing probability score.

Dataset and Experimental Setup

1. Dataset Composition

The experimental corpus combines two complementary sources. Phish Tank provides a continuously updated community-verified feed of active phishing URLs. Tranco, a research grade Alexa-replacement list, supplies URLs of legitimate high-traffic domains. After deduplication and quality filtering —removing expired domains, login-wall pages, and non-HTML content—the final dataset comprises 86,400 samples: 43,200 phishing and 43,200 legitimate, ensuring a perfectly balanced binary classification benchmark.

2. Data Collection Pipeline

A headless Chromium driver visited each URL in the corpus and recorded three artifacts: (i) the raw URL string, (ii) the rendered HTML source after JavaScript execution, and (iii) a 224×224 screenshot. Collection was parallelized across 16 workers, with a per-URL timeout of 8 seconds to skip unresponsive hosts. Approximately 4.3% of originally sampled URLs were discarded due to timeouts or certificate errors, and replacements were drawn from residual pool entries to maintain balance.

3. Training Configuration

The dataset was partitioned into 70% training, 15% validation, and 15% test splits using stratified random sampling to preserve the 1:1 class ratio across all partitions. Each stream encoder was pretrained independently on its respective modality before joint fine-tuning of the full pipeline. The Adam optimizer with an initial learning rate of $1\times10-4$, cosine annealing schedule, and weight decay of $1\times10-5$ was used throughout. Batch size was set to 64. Training ran for 30 epochs on an NVIDIA RTX 4090 with mixed-precision (FP16) computation to reduce memory footprint.

Results and Analysis

1. Classification Performance

The proposed framework achieves 97.4% accuracy, 97.1% precision, 97.6% recall, and an F1-score of 0.973 on the held-out test partition. The ROC-AUC of 0.994 reflects near-optimal discrimination across the full range of classification thresholds. Table II presents a comparative evaluation against strong single-modality and prior multi-modal baselines.

Table II. Comparative Performance on Phishtank + Tranco Benchmark

Model		Accuracy (%)	F1-Score	ROCAUC	Latency (ms)
URL-only Bi-LSTM		91.2	0.909	0.961	38
DOM Transformer		93.5	0.933	0.974	142

EfficientNet-B0 Visual		90.7	0.904	0.957	91
Concat Fusion Baseline		95.8	0.956	0.986	289
URLNet [4]		93.1	0.929	0.971	55
VisualPhish [5]		91.9	0.916	0.963	118
Proposed Framework		**97.4**	**0.973**	**0.994**	**468**

2. Ablation Study

To quantify the contribution of each modality and the fusion mechanism, four ablation variants were evaluated: removing each individual stream while retaining the other two, and replacing the Cross-Modal Attention with simple vector concatenation. Results show that removing the URL stream incurs the largest accuracy drop (−3.8%), followed by DOM removal (−2.6%) and visual removal (−2.1%), indicating that URL features carry the highest individual discriminatory weight. Replacing attention fusion with concatenation reduced accuracy by 1.6 percentage points, confirming that learned intermodal weighting provides meaningful benefit beyond feature aggregation.

3. Scalability Under Load

The inference service was deployed on a 4-core CPU server without GPU acceleration to simulate a lightweight proxy scenario. Table III documents throughput and latency characteristics across increasing concurrent request volumes.

Table III. Inference Performance Under Concurrent Load

Concurrent Requests	**Mean Latency (ms)**	**P99 Latency (ms)**	**Throughput (req/s)**	**CPU Usage (%)**
1	112	138	8.9	12
4	184	227	21.7	38
8	311	389	25.7	71
16	468	592	34.2	94

5latency remains below 600 ms across all tested concurrency levels on CPU-only infrastructure. GPU-accelerated deployment reduces mean latency to 87 ms at 16 concurrent requests, enabling deployment in high-traffic network edge scenarios.

4. Zero-Day Detection Capability

A temporal evaluation was constructed by training on URLs collected before January 2024 and testing on phishing URLs first reported after March 2024. The proposed framework achieved 94.2% accuracy on this temporally disjoint set, compared to 83.7% for the URL-only Bi-LSTM baseline. The visual and DOM streams provide robust signal for zero-day phishing domains that have not yet appeared in blacklists, since adversaries cannot simultaneously spoof all three modalities without substantially increasing development overhead.

Explainability Via Shap

1. Feature Attribution Framework

Security analysts require interpretable rationales for classification decisions, particularly when reviewing borderline cases or constructing threat intelligence reports. SHapley Additive exPlanations (SHAP) were integrated to quantify the contribution of each input feature dimension to individual predictions. KernelSHAP was applied to the fused 768-dimensional representation, decomposing prediction scores into per-feature attributions that sum to the model output minus the expected value baseline.

2. Modality-Level Contribution Analysis

Aggregated SHAP analysis over the test partition reveals that URL features account for an average of 41.3% of prediction magnitude, DOM features contribute 34.7%, and visual features provide the remaining 24.0%. These proportions vary substantially by attack type: visual spoofing attacks see visual SHAP contributions rise to 38.2%, while obfuscated URL attacks assign 52.6% weight to the URL stream. This dynamic modality weighting demonstrates that the Cross-Modal Attention mechanism genuinely adapts to attack-specific presentation patterns rather than statically privileging a single stream.

3. Analyst Interface

The deployed system presents analysts with a ranked list of top-contributing URL tokens, DOM element types, and visual heatmap regions for each flagged page. This interface reduces analyst review time by directing attention to the specific signals that drove the classification, rather than requiring a full manual page examination. In a pilot evaluation with four security analysts, average review time per flagged URL was reduced from 4.2 minutes (without SHAP) to 1.7 minutes (with SHAP).

Implementation Details

1. Technology Stack

The framework is implemented in Python 3.11 using PyTorch 2.2 as the deep learning backend. URL tokenization and Bi-LSTM training utilize custom PyTorch modules. The DOM Transformer encoder is initialized from a DistilBERT

checkpoint fine-tuned on HTML corpora. EfficientNet-B0 weights are sourced from the timm model library. The inference API is built on FastAPI with asynchronous request handling, and Playwright drives headless Chromium for screenshot capture. Model weights and artifacts are serialized using TorchScript for portable deployment.

2. Preprocessing Pipeline

URL strings undergo Unicode normalization (NFKC), percent-decoding, and lowercasing before character tokenization. HTML source is cleaned by removing script and style tags that contain only tracking or analytics code, then serialized to a DOM traversal token sequence with a vocabulary of 8,192 HTML-specific tokens. Screenshots are normalized with ImageNet channel statistics (mean [0.485, 0.456, 0.406], std [0.229, 0.224, 0.225]) before Efficient Net forward passes. All preprocessing steps are parallelized across CPU cores using Python's multiprocessing. Pool.

3. Deployment Architecture

The production system runs as a containerized microservice on a single-node Docker deployment. A Redis queue decouples URL submission from inference processing, enabling backpressure handling during traffic spikes. The SHAP explanation engine runs asynchronously post-classification to avoid adding to primary inference latency. Prometheus metrics expose per stream inference times, queue depth, and prediction confidence distributions for operational monitoring.

Security Considerations

1. Adversarial Robustness

Multi-modal architectures introduce a higher adversarial cost compared to single-modality classifiers. An attacker seeking to evade detection must simultaneously craft a URL that satisfies lexical legitimacy criteria, construct a DOM that avoids structural anomalies, and render a visual layout that does not resemble known phishing brand templates—three independently constrained optimization objectives. Adversarial training was applied to the URL and visual streams using projected gradient descent perturbations during the final fine-tuning epochs, yielding a 2.1 percentage point improvement in accuracy on an adversarially augmented test set.

2. Model Poisoning Considerations

Continuous online retraining from community-sourced phishing feeds introduces a data poisoning surface. Adversaries who contribute false negatives to PhishTank could gradually shift the decision boundary. The deployed system mitigates this by applying anomaly detection on incoming training samples before incorporation, flagging submissions with unusually high model confidence as potential clean-label poison candidates for manual review.

3. Privacy Considerations

Screenshot capture and DOM extraction at the proxy level necessarily access rendered page content. The deployment architecture ensures that all captured artifacts are processed ephemerally in memory and are not persisted to disk beyond the inference lifecycle. URLs and feature vectors are anonymized before logging to prevent reconstruction of user browsing histories from telemetry data.

Discussion

1. Comparison with State of the Art

The proposed framework surpasses the nearest comparable multi-modal phishing detection system by 1.6 percentage points in accuracy and 0.008 in ROC-AUC, while operating within the <500 ms real-time constraint that previous multimodal approaches failed to satisfy simultaneously. The primary performance driver is the Cross-Modal Attention Fusion mechanism: prior work that concatenated modality embeddings could not model cases where a suspicious URL corresponds to a visually trustworthy page—an increasingly common tactic in credential harvesting campaigns.

2. Limitations

Several limitations bound the current evaluation. First, the dataset, while large, reflects a snapshot of the phishing landscape from a specific period; adversarial evolution may degrade accuracy over time without active dataset refreshes. Second, the screenshot capture dependency introduces a latency floor that prevents sub-100 ms deployment without GPU acceleration. Third, heavily JavaScript-dependent single-page applications may not render fully within the headless browser timeout window, potentially producing incomplete DOM or visual inputs. Fourth, the framework currently performs binary (phishing vs. legitimate) classification; multi-class attribution to specific targeted brands remains future work.

3. Real-World Deployment Pathway

The inference service is designed for integration at three deployment points: (i) as a browser extension that evaluates the current tab's URL and DOM in the background, (ii) as a network proxy plugin that intercepts HTTP responses and appends a risk header, and (iii) as an API endpoint consumed by email security gateways for embedded URL scanning. The containerized microservice architecture supports all three integration patterns without modification to the core model.

Conclusion and Future Work

Conclusion

This paper presented a Hybrid Multi-Modal Deep Learning Framework that unifies URL lexical analysis, HTML DOM structural encoding, and visual screenshot understanding through a learnable Cross-Modal Attention Fusion mechanism for

real-time phishing website detection. The framework achieves 97.4% accuracy and 0.994 ROC-AUC on a balanced 86,400-sample benchmark composed of PhishTank and Tranco entries, surpassing single-modality baselines by 4–7 percentage points. End-to-end inference latency of 468 ms on CPU and 87 ms on GPU confirms production viability. The integration of SHAP attributions enhances analyst trust and accelerates manual review workflows. By requiring adversaries to simultaneously defeat three independent detection streams, the multimodal design substantially raises the cost of evasion compared to prior art.

Future Work

Identified directions for future research include: (i) incorporation of network-level features such as DNS resolution patterns and TLS certificate metadata as a fourth modality; (ii) continuous online learning with poison-resistant dataset curation to adapt to adversarial evolution; (iii) extension to multiclass brand attribution enabling targeted takedown notifications; (iv) lightweight distilled model variants targeting mobile and IoT deployment contexts; and (v) federated learning formulations that allow organizations to collaboratively improve the shared model without exchanging raw URL or screenshot data, addressing enterprise privacy requirements.

References

1. L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," in IEEE Access, vol. 10, pp. 1509-1521, 2022, doi: 10.1109/ACCESS.2021.3137636
2. S. Asiri, Y. Xiao, S. Alzahrani, S. Li and T. Li, "A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks," in IEEE Access, vol. 11, pp. 6421-6443, 2023, doi: 10.1109/ACCESS.2023.3237798.
3. M. Alsaedi, F. A. Ghaleb, F. Saeed, J. Ahmad and M. Alasli, "Multi-Modal Features Representation-Based Convolutional Neural Network Model for Malicious Website Detection," in IEEE Access, vol. 12, pp. 7271-7284, 2024, doi: 10.1109/ACCESS.2023.3348071
4. D. He, X. Lv, X. Xu, S. Chan and K. -K. R. Choo, "Double-Layer Detection of Internal Threat in Enterprise Systems Based on Deep Learning," in IEEE Transactions on Information Forensics and Security, vol. 19, pp. 4741-4751, 2024, doi: 10.1109/TIFS.2024.3372771.
5. F. S. Alsubaei, A. A. Almazroi and N. Ayub, "Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics," in IEEE Access, vol. 12, pp. 8373-8389, 2024, doi: 10.1109/ACCESS.2024.3351946.
6. M. Alsaedi, F. A. Ghaleb, F. Saeed, J. Ahmad and M. Alasli, "Multi-Modal Features Representation-Based Convolutional Neural Network Model for Malicious Website Detection," in IEEE Access, vol. 12, pp. 7271-7284, 2024, doi: 10.1109/ACCESS.2023.3348071.

7. B. B. Gupta, A. Gaurav, P. Chaurasia and K. T. Chui, "Lightweight Deep Learning Model and Genetic Algorithm Based Optimal Phishing Website Detection," 2024 IEEE 13th Global Conference on Consumer Electronics (GCCE), Kitakyushu, Japan, 2024, pp. 1403-1404, doi: 10.1109/GCCE62371.2024.10760723.
8. D. Jennifer Dsouza, A. P. Rodrigues and R. Fernandes, "Multi-Modal Comparative Analysis on Execution of Phishing Detection Using Artificial Intelligence," in IEEE Access, vol. 12, pp. 163016-163041, 2024, doi: 10.1109/ACCESS.2024.3491429
9. S. Remya, M. J. Pillai, B. S. Aparna, S. Rama Subbareddy and Y. Y. Cho, "BGL-PhishNet: Phishing Website Detection Using Hybrid Model-BERT, GNN, and LightGBM," in IEEE Access, vol. 13, pp. 47552-47569, 2025, doi: 10.1109/ACCESS.2025.3551542
10. S. Kavya and D. Sumathi, "Multimodal and Temporal Graph Fusion Framework for Advanced Phishing Website Detection," in IEEE Access, vol. 13, pp. 74128-74146, 2025, doi: 10.1109/ACCESS.2025.3564530.
11. Y. Wei, M. Nakayama and Y. Sekiya, "Enhancing Generalization in Phishing URL Detection via a Fine-Tuned BERT-Based Multimodal Approach," in IEEE Access, vol. 13, pp. 131197-131216, 2025, doi: 10.1109/ACCESS.2025.3591843

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 94 - 100 |

Artificial Intelligence and the Future of Legal Accountability

Purbita Das

Assistant Professor, School of Law, Brainware University

Email:

Article DOI Link: https://zenodo.org/uploads/20715088

DOI: 10.5281/zenodo.20715088

Abstract

The rapid integration of Artificial Intelligence into governance, commerce, and everyday life has introduced profound legal challenges, particularly in the realm of accountability. As AI systems increasingly perform tasks that were once exclusively within human control—ranging from decision-making in public administration to predictive analytics in criminal justice—the limitations of traditional legal frameworks have become more evident. These frameworks, historically grounded in notions of human agency, intent, and fault, are ill-equipped to address the complexities posed by autonomous and semi-autonomous technologies that operate with limited human intervention.

One of the central concerns lies in the attribution of responsibility when AI systems produce harmful or unintended outcomes. The diffusion of accountability among developers, deployers, and users creates significant legal ambiguity, often referred to as the "accountability gap." This challenge is further compounded by the opaque nature of algorithmic decision-making, which undermines principles such as transparency, causation, and foreseeability.

This article examines the evolving concept of legal accountability in the age of AI by critically analyzing existing liability regimes, emerging regulatory approaches, and the growing role of ethical frameworks in shaping legal responses. It highlights comparative developments across jurisdictions and underscores the need for adaptive legal mechanisms. Ultimately, it argues for a hybrid legal framework that combines traditional doctrines with innovative regulatory tools—such as explainability requirements, shared liability models, and risk-based governance—to effectively address the multifaceted challenges of AI-driven decision-making while ensuring justice, fairness, and public trust.

Keywords: Artificial Intelligence, Legal Accountability, Liability, Regulation, Autonomous Systems, Ethics

Introduction

Artificial Intelligence has transitioned from a speculative concept to a transformative force reshaping industries and governance structures. Its applications now extend across diverse sectors, including healthcare, law enforcement, finance, and public administration. From predictive policing tools that assess crime probabilities to algorithmic trading systems executing high-frequency financial transactions, AI systems are increasingly making decisions that carry significant legal and societal consequences. This technological evolution raises a fundamental legal question: who should be held accountable when AI systems cause harm or produce unjust outcomes?

Traditional legal systems are premised on the assumption of human agency, attributing responsibility based on intent, negligence, or strict liability. However, AI systems—particularly those driven by machine learning and neural networks—operate with varying degrees of autonomy and adaptability. Their capacity to learn from data and evolve over time often results in outputs that are neither fully predictable nor directly traceable to a specific human decision-maker. This unpredictability complicates the application of conventional legal doctrines.

Consequently, a "responsibility gap" emerges, where accountability becomes diffused among multiple stakeholders, including developers, programmers, deployers, and end-users. The opaque, "black-box" nature of many AI systems further exacerbates this issue by limiting transparency and hindering the establishment of causation. As a result, existing legal frameworks face significant challenges in ensuring effective redress and deterrence. Addressing this gap requires a re-examination of foundational legal principles and the development of adaptive frameworks capable of responding to the unique characteristics of AI-driven technologies.

The Concept of Legal Accountability

Legal accountability refers to the obligation of individuals or entities to answer for their actions under the law and to bear the consequences when those actions result in harm. It encompasses both civil and criminal liability and is traditionally grounded in established legal principles such as fault, causation, and foreseeability. These principles enable courts to determine responsibility by linking a wrongful act to a specific actor and assessing whether the harm was intentional, negligent, or reasonably foreseeable.

However, in the context of Artificial Intelligence, the notion of accountability becomes significantly more complex and fragmented. AI systems are typically developed, trained, and deployed through a multi-layered process involving various actors, including software developers, data scientists, corporate entities, and end-users. Each of these participants contributes in different ways to the functioning of the system, making it difficult to isolate responsibility when harm occurs. This

diffusion of roles challenges the traditional model of singular liability and raises questions about whether responsibility should be shared or redefined.

Furthermore, the decentralized and often transnational nature of AI development exacerbates these challenges. AI systems may be designed in one jurisdiction, trained on data from another, and deployed globally, thereby complicating the application of domestic legal frameworks. The opacity of algorithmic processes also limits transparency, making it harder to establish causation and foreseeability. Consequently, the attribution of liability in AI-related harm remains uncertain, necessitating the development of more nuanced and adaptable legal approaches.

Liability Challenges in AI Systems

- **The Problem of Autonomy**

AI systems, particularly those based on deep learning and neural networks, possess the ability to operate with a degree of autonomy that distinguishes them from traditional technologies. Unlike conventional software, these systems can learn from data inputs and modify their behavior over time without explicit human intervention. This evolving functionality raises complex legal questions regarding the attribution of liability. When an autonomous system causes harm, it becomes unclear whether responsibility should lie with the developer who designed the algorithm, the operator who deployed it, or the entity that benefits from its use. The possibility of assigning liability to the AI system itself remains largely theoretical, as current legal frameworks do not recognize machines as legal persons.

- **Causation and Foreseeability**

Causation is a fundamental element in establishing legal liability, requiring a clear link between an action and the resulting harm. However, the "black box" nature of many AI systems complicates this process. The internal decision-making pathways of such systems are often opaque, making it difficult to determine how a specific outcome was reached. This lack of transparency undermines the principle of foreseeability, which is central to negligence-based liability, as it becomes challenging to predict or anticipate harmful outcomes.

- **Product Liability and AI**

One proposed solution is to apply existing product liability laws to AI systems by treating them as defective products. While this approach offers a familiar legal framework, it is inherently limited. Unlike traditional products, AI systems are dynamic and continuously evolving through machine learning processes. This adaptability makes it difficult to define what constitutes a "defect" and at what point liability should be assessed, thereby necessitating a re-evaluation of conventional product liability doctrines in the context of AI.

Comparative Regulatory Approaches

- **The European Union**

The European Union has emerged as a global leader in the regulation of Artificial Intelligence through its comprehensive and precautionary approach. The proposed regulatory framework adopts a risk-based classification system, categorizing AI applications into unacceptable, high-risk, limited-risk, and minimal-risk groups. High-risk AI systems—such as those used in critical infrastructure, law enforcement, and employment—are subject to stringent compliance requirements, including mandatory risk assessments, human oversight, data governance standards, and transparency obligations. This approach reflects the EU's broader commitment to fundamental rights, consumer protection, and ethical governance, aiming to ensure that technological innovation does not compromise legal accountability or societal values.

- **The United States**

In contrast, the United States follows a more decentralized and sector-specific approach to AI regulation. Rather than adopting a single, comprehensive legal framework, it relies on a combination of existing laws, regulatory agency guidelines, and voluntary industry standards. This model encourages innovation and flexibility, allowing rapid technological development. However, it also results in fragmented oversight and potential regulatory gaps, particularly in areas where AI applications cut across multiple sectors. The absence of uniform federal legislation raises concerns about consistency, enforcement, and accountability.

- **India's Emerging Framework**

India is currently in the formative stages of developing its AI regulatory regime, primarily guided by policy initiatives and ethical frameworks. Government bodies have emphasized principles such as transparency, inclusivity, and responsible innovation. While there is growing recognition of the need to address issues of accountability, bias, and data protection, a comprehensive and enforceable legal framework for AI governance is still evolving. This presents both an opportunity and a challenge for India to design a balanced and context-sensitive regulatory model.

- **Ethical Dimensions of AI Accountability**

Legal accountability cannot be divorced from ethical considerations, particularly in the context of Artificial Intelligence, where technological decisions often have profound human consequences. Issues such as algorithmic bias, discrimination, and privacy violations underscore the moral dimensions of AI deployment. For instance, biased training data can lead to discriminatory outcomes in areas like hiring, credit scoring, or criminal justice, thereby reinforcing existing social inequalities.

Similarly, the large-scale collection and processing of personal data by AI systems raise serious concerns about privacy, consent, and individual autonomy.

These ethical challenges reveal the limitations of purely legalistic approaches that focus only on compliance and liability. Instead, there is a growing recognition that ethical principles must play a foundational role in shaping legal frameworks. Ethical AI frameworks developed by international organizations, governments, and academic institutions consistently emphasize core values such as transparency, fairness, accountability, and human oversight. Transparency ensures that AI systems operate in an explainable manner, while fairness seeks to eliminate bias and promote equitable outcomes. Human oversight remains essential to prevent unchecked automation and to preserve human dignity in decision-making processes.

Integrating these ethical principles into legal regulations can enhance the legitimacy and effectiveness of AI governance. It allows for a more holistic approach that not only addresses harm after it occurs but also promotes responsible innovation and risk prevention. Ultimately, aligning legal accountability with ethical standards is crucial for building public trust and ensuring that AI technologies serve the broader interests of society.

Towards a Hybrid Accountability Framework

- **Shared Liability Models**

A shared liability approach offers a pragmatic solution to the complexities of AI accountability by distributing responsibility among all stakeholders involved in the AI lifecycle. This includes developers who design the algorithms, organizations that deploy and manage the systems, and end-users who interact with them. Such a model recognizes the collective contribution to potential harm and ensures that accountability is not unfairly concentrated on a single actor. It also encourages greater diligence at each stage of development and deployment, fostering a culture of responsible innovation.

- **Mandatory Transparency and Explainability**

Legal frameworks should mandate transparency and explainability in AI systems to ensure that decisions can be understood and scrutinized. Explainable AI enables affected individuals, regulators, and courts to trace how specific outcomes are reached, thereby strengthening accountability. This is particularly crucial in high-stakes areas such as healthcare, finance, and criminal justice, where opaque decision-making can have serious consequences. Enhanced transparency also facilitates legal redress by making it easier to establish causation and identify responsible parties.

- **Regulatory Sandboxes**

Regulatory sandboxes provide a controlled environment in which AI technologies can be tested under the supervision of regulatory authorities. These frameworks allow innovators to experiment with new applications while enabling policymakers to observe potential risks and refine regulatory responses. Sandboxes thus strike a balance between innovation and oversight, promoting adaptive governance.

- **Insurance-Based Models**

Insurance-based mechanisms represent another viable approach to addressing AI-related harm. Compulsory insurance schemes can ensure that victims receive timely compensation, even in cases where liability is difficult to determine. This model shifts the focus from fault-finding to risk management and provides a practical safety net in an evolving technological landscape.

Conclusion

The rise of Artificial Intelligence necessitates a fundamental rethinking of legal accountability in contemporary legal systems. While existing legal doctrines—such as negligence, strict liability, and product liability—continue to provide a foundational framework, they are increasingly inadequate to address the dynamic and autonomous nature of AI systems. These technologies operate in ways that challenge traditional assumptions about human control, intent, and predictability, thereby exposing significant gaps in current legal structures.

A forward-looking legal framework must therefore move beyond incremental adjustments and adopt a more holistic and adaptive approach. This involves integrating traditional legal principles with innovative regulatory mechanisms, such as risk-based governance, shared liability models, mandatory transparency requirements, and institutional oversight. Such an approach ensures that accountability remains effective and enforceable, even as technological capabilities continue to evolve.

Importantly, the objective of legal regulation in the context of AI should not be confined to the retrospective assignment of liability. Rather, it should aim to create a proactive legal environment that encourages responsible innovation, minimizes risks, and embeds ethical considerations into the design and deployment of AI systems. Safeguarding individual rights—such as privacy, equality, and due process—must remain central to this framework, alongside the protection of broader societal interests.

Ultimately, a balanced and future-oriented approach to legal accountability will be essential in fostering public trust and ensuring that the benefits of AI are realized without compromising justice, fairness, and the rule of law.

References

1. Bignami, F. (2022). Artificial intelligence accountability of public administration. The American Journal of Comparative Law, 70(Supplement_1), i312–i346. https://doi.org/10.1093/ajcl/avac012
2. Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. Public Administration Review, 81(5), 825–836. https://doi.org/10.1111/puar.13293
3. Doshi-Velez, F., & Kortz, M. (2017). Accountability of AI under the law: The role of explanation. SSRN. https://doi.org/10.2139/ssrn.3064761
4. Lechterman, T. (2023). The concept of accountability in AI ethics and governance. In J. B. Bullock et al. (Eds.), The Oxford handbook of AI governance. Oxford University Press.
5. Mökander, J., Juneja, P., Watson, D. S., & Floridi, L. (2022). The US Algorithmic Accountability Act of 2022 vs. the EU Artificial Intelligence Act: What can they learn from each other? Minds and Machines, 32, 751–758. https://doi.org/10.1007/s11023-022-09612-y
6. Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: What it is and how it works. AI & Society, 39, 1871–1882. https://doi.org/10.1007/s00146-023-01635-y
7. Porter, Z., Zimmermann, A., Morgan, P. D. J., McDermid, J. A., Lawton, T., & Habli, I. (2022). Distinguishing two features of accountability for AI technologies. Nature Machine Intelligence, 4, 734–736. https://doi.org/10.1038/s42256-022-00533-0
8. Slota, S. C., Fleischmann, K. R., Greenberg, S., Verma, N., Cummings, B., Li, L., & Shenefiel, C. (2023). Many hands make many fingers to point: Challenges in creating accountable AI. AI & Society, 38(4), 1287–1299.
9. Additional Foundational References (Recommended for Expansion)
10. Calo, R. (2015). Robotics and the lessons of cyberlaw. California Law Review, 103(3), 513–563.
11. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707.
12. Surden, H. (2019). Artificial intelligence and law: An overview. Georgia State University Law Review, 35(4), 1305–1335.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 101 - 108 |

Smart Cities and Digital Surveillance: Legal Limits on Technology-Enabled Urban Governance

Mr. Subham Chatterjee

Assistant Professor, School of Law, Brainware University, Kolkata

Email:

Article DOI Link: https://zenodo.org/uploads/20715143

DOI: 10.5281/zenodo.20715143

Abstract

The emergence of smart cities represents one of the most significant transformations in urban governance during the twenty-first century. By integrating information and communication technologies, artificial intelligence, Internet of Things (IoT) devices, big data analytics, facial recognition systems, and digital surveillance infrastructure, smart cities seek to improve public services, enhance efficiency, and strengthen urban management. In India, the Smart Cities Mission has accelerated the adoption of digital technologies in urban administration. While these innovations promise better governance, sustainability, and public safety, they also raise serious legal and constitutional concerns regarding privacy, surveillance, accountability, data protection, and civil liberties. The increasing deployment of digital surveillance technologies challenges traditional notions of individual autonomy and democratic governance. This article critically examines the legal implications of technology-enabled urban governance, focusing on digital surveillance in smart cities. It analyzes constitutional protections, judicial developments, data protection legislation, and international legal principles relevant to urban surveillance. The study argues that while smart technologies can improve governance outcomes, the absence of comprehensive regulatory safeguards may create risks of excessive surveillance and infringement of fundamental rights. The article concludes by advocating for a rights-based legal framework that balances technological innovation with privacy, transparency, and democratic accountability.

Introduction

Urbanization has become one of the defining features of contemporary society. As cities expand in population, complexity, and economic significance, governments increasingly rely on technology to manage urban challenges. The concept of the

"smart city" has emerged as a model for integrating digital technologies into urban governance to improve efficiency, sustainability, and quality of life.

Smart cities utilize technologies such as surveillance cameras, biometric systems, facial recognition software, sensors, predictive analytics, artificial intelligence, and centralized command-and-control centers. These technologies generate vast amounts of data that can be used to monitor traffic, manage public utilities, improve emergency response, and enhance public safety.

India's Smart Cities Mission, launched in 2015, seeks to transform selected cities through technology-driven governance initiatives. Integrated Command and Control Centres (ICCCs), intelligent transportation systems, and digital public infrastructure have become central features of urban modernization.

However, technology-enabled governance also raises important legal questions. Continuous monitoring of public spaces, collection of personal data, and deployment of algorithmic systems may affect privacy, freedom of movement, freedom of expression, and democratic participation. Consequently, the legal limits of digital surveillance have become a critical area of inquiry.

Understanding Smart Cities and Digital Surveillance

A smart city may be defined as an urban environment that utilizes digital technologies and data-driven systems to enhance governance, service delivery, and resource management.

Key components of smart cities include:

- Internet of Things (IoT) networks.
- Surveillance cameras.
- Facial recognition systems.
- Geographic Information Systems (GIS).
- Artificial intelligence applications.
- Data analytics platforms.
- Integrated urban command centers.

Digital surveillance refers to the systematic collection, monitoring, analysis, and storage of information regarding individuals and activities through technological means.

Unlike traditional forms of surveillance, digital surveillance operates continuously and often invisibly. Advanced systems can track movement patterns, identify individuals, predict behavior, and integrate information from multiple sources.

The combination of smart city infrastructure and digital surveillance creates unprecedented opportunities for governance while simultaneously raising concerns regarding civil liberties.

The Growth of Smart Cities in India

India's Smart Cities Mission represents one of the world's largest urban modernization initiatives. The mission seeks to improve urban infrastructure, public services, and governance through technology-based solutions.

Several Indian cities have established Integrated Command and Control Centres that monitor urban activities through interconnected digital systems.

Applications include:

- Traffic management.
- Waste management.
- Public safety monitoring.
- Environmental monitoring.
- Emergency response coordination.
- Utility management.

Cities increasingly deploy CCTV networks, automated number plate recognition systems, and facial recognition technologies to support law enforcement and public administration.

These developments illustrate the growing reliance on data-driven governance models in urban environments.

Constitutional Framework and Digital Surveillance

The legal regulation of surveillance in India is fundamentally shaped by constitutional principles.

Article 14: Equality Before Law

Article 14 guarantees equality before the law and equal protection of laws. Surveillance systems must therefore operate without arbitrary discrimination or unequal treatment.

Algorithmic decision-making systems may create risks of biased outcomes if they disproportionately affect certain groups or communities.

Article 19: Freedom of Expression and Movement

Digital surveillance may influence individuals' willingness to express opinions, participate in public activities, or exercise freedom of movement.

Excessive monitoring can create a "chilling effect" that discourages democratic participation and free expression.

Article 21: Right to Life and Personal Liberty

Article 21 has become the central constitutional provision governing privacy and surveillance.

Judicial interpretation has expanded Article 21 to include dignity, autonomy, and informational privacy. Surveillance measures that affect personal liberty must therefore satisfy constitutional standards of legality and proportionality.

The Right to Privacy and the Puttaswamy Judgment

A major legal milestone in India occurred in Justice K.S. Puttaswamy (Retd.) v. Union of India (2017).

The Supreme Court unanimously recognized privacy as a fundamental right protected under Article 21 of the Constitution.

The Court identified privacy as encompassing:

- Bodily privacy.
- Informational privacy.
- Decisional autonomy.

The judgment established that any intrusion into privacy must satisfy three conditions:

- Legality.
- Legitimate state purpose.
- Proportionality.

These principles provide the constitutional foundation for evaluating digital surveillance practices in smart cities.

The Puttaswamy judgment remains highly significant because modern surveillance technologies primarily operate through the collection and processing of personal information.

Surveillance Technologies in Smart Cities

CCTV Networks

Closed-circuit television (CCTV) systems have become a central component of urban surveillance.

Supporters argue that CCTV cameras enhance public safety, deter crime, and assist investigations.

However, large-scale camera networks may enable continuous monitoring of individuals in public spaces.

Questions arise regarding data retention, access controls, and accountability mechanisms.

Facial Recognition Technology

Facial recognition systems use biometric information to identify individuals from images or video footage.

These technologies offer potential benefits for law enforcement and security.

However, concerns include:

- Misidentification.
- Racial and demographic bias.
- Mass surveillance.
- Lack of informed consent.

Several jurisdictions worldwide have imposed restrictions on facial recognition due to privacy concerns.

Artificial Intelligence and Predictive Policing

AI systems increasingly support predictive policing and crime analysis.

These systems analyze large datasets to identify patterns and predict potential risks.

Critics argue that predictive systems may reinforce existing biases and lack transparency regarding decision-making processes.

The use of AI in governance therefore raises significant legal and ethical questions.

Data Protection and Governance

Smart cities generate enormous quantities of data.

This information may include:

- Location data.
- Biometric information.
- Communication records.
- Transportation patterns.
- Utility consumption data.

Effective governance of such data requires comprehensive legal safeguards.

Digital Personal Data Protection Act, 2023

India's Digital Personal Data Protection Act, 2023 provides the primary framework for personal data governance.

Key principles include:

- Consent-based processing.
- Purpose limitation.
- Data minimization.
- Security safeguards.
- Individual rights.

The Act is particularly relevant to smart city initiatives because surveillance technologies frequently process personal data.

However, concerns remain regarding exemptions, state surveillance powers, and implementation mechanisms.

Legal Challenges of Technology-Enabled Governance

Absence of Comprehensive Surveillance Regulation

India currently lacks a dedicated surveillance law comprehensively regulating public-sector surveillance technologies.

Existing frameworks rely on fragmented statutory provisions that may not adequately address modern technological capabilities.

Transparency Deficits

Many surveillance systems operate with limited public awareness regarding:

- Data collection practices.
- Data sharing arrangements.
- Algorithmic decision-making processes.

Transparency is essential for democratic accountability.

Function Creep

Technologies initially deployed for specific purposes may gradually expand into broader surveillance functions.

This phenomenon, known as "function creep," increases risks of excessive monitoring and misuse.

Accountability Challenges

Determining responsibility for surveillance-related harms can be difficult when multiple public and private actors participate in data collection and analysis.

Effective accountability mechanisms remain necessary.

International Perspectives

Several jurisdictions have adopted regulatory approaches that may inform Indian policy development.

European Union

The General Data Protection Regulation (GDPR) establishes strong protections for personal data and imposes obligations on organizations processing information.

The European Union has also proposed regulations governing artificial intelligence and biometric surveillance.

United States

Several cities in the United States have restricted or prohibited government use of facial recognition technology due to civil liberties concerns.

United Kingdom

The United Kingdom has developed surveillance oversight mechanisms through data protection laws and independent regulatory bodies.

Comparative experiences demonstrate the importance of balancing security objectives with fundamental rights protections.

Democratic Governance and Urban Citizenship

Smart city governance affects the relationship between citizens and the state.

Digital technologies may enhance governance efficiency, but they also alter power dynamics by increasing state capacity to collect and analyze information.

Democratic governance requires that technological systems remain subject to:

- Public oversight.
- Judicial review.
- Legislative accountability.

- Citizen participation.

Urban residents should not become passive subjects of technological monitoring. Rather, they should remain active participants in decisions regarding surveillance and governance.

The concept of "digital citizenship" emphasizes rights, participation, and accountability within technologically mediated environments.

Policy Recommendations

Several measures can strengthen legal protections within smart city governance.

Comprehensive Surveillance Legislation

India should develop a dedicated legal framework governing surveillance technology, including facial recognition systems and AI-based monitoring.

Privacy Impact Assessments

Public authorities should conduct privacy impact assessments before deploying surveillance technologies.

Transparency Requirements

Governments should disclose information regarding surveillance systems, data practices, and oversight mechanisms.

Independent Oversight

Independent regulatory bodies can monitor compliance with privacy and data protection standards.

Algorithmic Accountability

AI systems used in governance should be transparent, explainable, and subject to regular audits.

Public Participation

Citizens should have opportunities to participate in decisions concerning surveillance technologies affecting public spaces.

Conclusion

Smart cities represent an important transformation in contemporary urban governance. By integrating digital technologies into public administration, governments seek to improve efficiency, sustainability, and service delivery.

However, the increasing reliance on surveillance technologies raises profound legal and constitutional questions. CCTV networks, facial recognition systems, AI-driven analytics, and centralized command centers create new possibilities for governance while simultaneously challenging traditional understandings of privacy and liberty.

The recognition of privacy as a fundamental right in the Puttaswamy judgment provides an important constitutional foundation for regulating surveillance

practices. Similarly, the Digital Personal Data Protection Act, 2023 establishes significant protections for personal information.

Nevertheless, substantial regulatory gaps remain. The absence of comprehensive surveillance legislation, transparency deficits, and accountability challenges create risks of excessive monitoring and infringement of fundamental rights.

The future of smart city governance depends on achieving a balance between technological innovation and constitutional values. Effective governance requires not only advanced technology but also strong legal safeguards, democratic accountability, and respect for individual rights.

Ultimately, smart cities should enhance human freedom and dignity rather than undermine them. The legal limits on digital surveillance must therefore remain central to the development of technology-enabled urban governance.

References

1. Bennett, C. J., & Lyon, D. (2019). Playing the identity card: Surveillance, security and identification in global perspective. Routledge.
2. Digital Personal Data Protection Act, 2023, No. 22, Acts of Parliament, India.
3. European Parliament and Council. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union.
4. Government of India. (2015). Smart Cities Mission Statement and Guidelines. Ministry of Housing and Urban Affairs.
5. Justice K. S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.
6. Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.
7. Kumar, H., Singh, M., Gupta, M. P., & Madaan, J. (2020). Smart cities in India: Challenges and opportunities. Sustainable Cities and Society, 62, 102367.
8. Lyon, D. (2018). The culture of surveillance: Watching as a way of life. Polity Press.
9. Madon, S., & Sahay, S. (2022). Digital governance and smart cities in India. Information Technology for Development, 28(4), 701–718.
10. Nissenbaum, H. (2010). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.
11. Sharma, A. (2021). Surveillance technologies and constitutional rights in India. Indian Law Review, 5(2), 145–168.
12. United Nations Human Rights Council. (2018). The right to privacy in the digital age. United Nations.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 109 - 124 |

Deep Learning Based Time series Forecasting Using LSTM & GRU Networks

[1]Ahamed Aasim M

[1]Akash Raj R

[2]Seetha P

[1]UG Student, Department of Informatics, Periyar Maniammai Institute of Science & Technology (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (Data science), Periyar Maniammai Institute of Science & Technology (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

Email: aasimmahamedaasim@gmail.com

Article DOI Link: https://zenodo.org/uploads/20733997

DOI: 10.5281/zenodo.20733997

Abstract

The Time series forecasting is widely used in areas such as sales prediction, weather analysis, stock market forecasting, and healthcare monitoring. Traditional forecasting methods often struggle to identify complex patterns and long-term relationships in sequential data. This project presents a deep learning-based forecasting system using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks to predict future values from historical sales data. The dataset is pre-processed using normalization, missing value handling, and sequence generation techniques before training the models. Both LSTM and GRU models are developed using TensorFlow and k eras to capture temporal dependencies effectively. Their performance is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Experimental results show that both models provide accurate predictions, with GRU offering faster training and LSTM performing better for long-term dependencies. The system helps organizations make reliable and data-driven forecasting decisions.

Keywords: Machine Learning, Loan Approval Prediction, Customer Segmentation, Random Forest, K-Means Clustering, Flask Web Application, Credit Scoring, Financial Technology.

Introduction

1. Background and Motivation

The Time series forecasting is an important area in data science that predicts future values using historical data. It is widely used in sales prediction, finance, inventory management, and business planning.

Traditional methods such as ARIMA and Moving Average are useful but have limitations in handling nonlinear patterns and long-term dependencies in large datasets. To overcome these issues, deep learning techniques are widely used.

LSTM and GRU are advanced Recurrent Neural Network (RNN) models designed for sequential data analysis. They can learn temporal patterns effectively and provide better forecasting accuracy compared to traditional methods.

This project focuses on using and comparing LSTM and GRU models for forecasting future sales using historical sales data.

2. Role of Deep Learning in Time Series Forecasting

Deep learning techniques have transformed the field of time series forecasting by enabling automatic learning of complex patterns from sequential data. Unlike traditional machine learning models that require manual feature extraction, deep learning models can directly learn temporal dependencies from raw datasets.

LSTM networks use memory cells and gating mechanisms such as input, forget, and output gates to preserve long-term dependencies in sequential data. This capability makes LSTM highly suitable for forecasting applications where historical information significantly influences future outcomes.

GRU networks are simplified versions of LSTM architectures that combine multiple gates into a smaller structure, reducing computational complexity and training time. Despite having fewer parameters, GRU models provide competitive prediction performance and are efficient for real-time forecasting systems.

The combination of LSTM and GRU models enables effective analysis of time-dependent data and improves forecasting accuracy in applications such as sales forecasting, stock market prediction, weather analysis, traffic monitoring, and healthcare analytics.

This research aims to design and implement a deep learning-based forecasting system capable of accurately predicting future values from historical time series data. The major objectives of this project are:

- To develop a forecasting system using LSTM and GRU deep learning models.
- To preprocess and normalize historical time series datasets for efficient model training.

- To analyse and compare the forecasting performance of LSTM and GRU architectures.
- To evaluate model performance using forecasting metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).
- To improve prediction accuracy and provide an intelligent forecasting solution for real-world business applications.

Research Gap

Traditional forecasting methods like ARIMA and Moving Average work well for linear data but struggle with nonlinear patterns and long-term dependencies. Deep learning models such as LSTM and GRU improve forecasting accuracy by learning temporal patterns automatically.

However, many existing studies use either LSTM or GRU individually without detailed comparison using the same dataset and evaluation metrics. Some systems also face issues like high training time and poor performance with large datasets.

This project develops a forecasting system using both LSTM and GRU models and compares their performance using historical sales data. The models are evaluated using metrics such as MAE and RMSE to identify the most accurate forecasting approach.

Research Objectives

This paper presents a comprehensive deep learning-based time series forecasting system that addresses the challenges of accurate sequential data prediction and temporal pattern analysis. The primary contributions of this research are:

- **Dual-Model Forecasting Architecture:** Integration of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks within a unified forecasting framework for comparative prediction analysis.
- **Deep Learning-Based Prediction System:** Development of an efficient forecasting model capable of learning nonlinear patterns and long-term dependencies from historical time series data.
- **Data Preprocessing Pipeline:** Design of a preprocessing framework including normalization, sequence generation, and missing value handling to improve forecasting performance and model accuracy.
- **Performance Evaluation Framework:** Implementation of forecasting evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) for comparing model effectiveness.
- **Real-Time Forecasting Capability:** Generation of accurate future predictions for applications such as sales forecasting, stock market analysis, and demand prediction using deep learning techniques.

Literature Review

1. Machine Learning in Credit Scoring

Deep learning techniques have become highly effective for time series forecasting because they can learn complex temporal patterns and nonlinear relationships in sequential data. Among these techniques, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are widely used for applications such as sales prediction, stock market analysis, weather forecasting, and energy demand estimation.

Sepp Hochreiter and Jürgen Schmid Huber introduced the LSTM architecture to solve the vanishing gradient problem in traditional Recurrent Neural Networks (RNNs), enabling better learning of long-term dependencies. Later, Kyunghyang Cho proposed the GRU model as a simpler and faster alternative to LSTM with reduced computational complexity.

Studies by Klaus Greff and other researchers showed that LSTM and GRU models provide better forecasting accuracy than traditional methods such as ARIMA and Moving Average models. Due to their strong ability to capture historical temporal information, these deep learning models are widely used in modern forecasting systems.

2. Deep Learning Models for Time Series Forecasting

Deep learning models are highly effective for time series forecasting because they can learn temporal patterns and nonlinear relationships from historical data. Models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are widely used in applications like sales prediction, weather forecasting, and stock market analysis.

LSTM networks can capture long-term dependencies using memory cells and gating mechanisms, making them suitable for sequential prediction tasks. GRU is a simplified version of LSTM that reduces computational complexity and enables faster training while maintaining good forecasting accuracy.

Studies show that both LSTM and GRU models outperform traditional forecasting methods such as ARIMA and Support Vector Regression (SVR) in handling complex and dynamic time series data. Due to their high accuracy and efficient learning capability, these models are widely used in modern forecasting systems.

3. Deep Learning-Based Forecasting Systems

Deep learning-based forecasting systems are widely used for real-time predictive analytics in areas such as sales prediction, stock market analysis, weather forecasting, and energy demand estimation. These systems combine deep learning models with web technologies to provide accurate and automated future predictions.

TensorFlow and Keres are commonly used for developing LSTM and GRU forecasting models because they are flexible, scalable, and easy to integrate with

Python applications. Flask is widely used to deploy these models as web applications or REST APIs for real-time prediction services.

Modern forecasting systems using LSTM and GRU networks provide fast predictions, efficient sequential data processing, and user-friendly interfaces, making them suitable for practical real-world forecasting applications.

Research Gap

Existing time series forecasting systems often use either LSTM or GRU models separately and lack detailed comparative analysis within a single framework. Traditional forecasting methods also struggle to capture nonlinear patterns, seasonality, and long-term dependencies in sequential data. Many systems require extensive preprocessing, have high computational complexity, and are not suitable for large-scale real-time forecasting.

Additionally, several forecasting applications lack user-friendly web-based deployment for non-technical users. This research addresses these limitations by developing a unified deep learning forecasting system using both LSTM and GRU models for comparative analysis. The system uses historical sales data and evaluates performance using metrics such as MAE and RMSE to provide accurate and scalable real-time

System Architecture

1. Overall System Architecture

The proposed system follows a deep learning-based forecasting architecture consisting of four major layers: data collection layer, preprocessing layer, deep learning model layer, and prediction visualization layer. The system is designed to perform accurate time series forecasting using LSTM and GRU neural networks. Fig. 1 illustrates the overall system architecture.

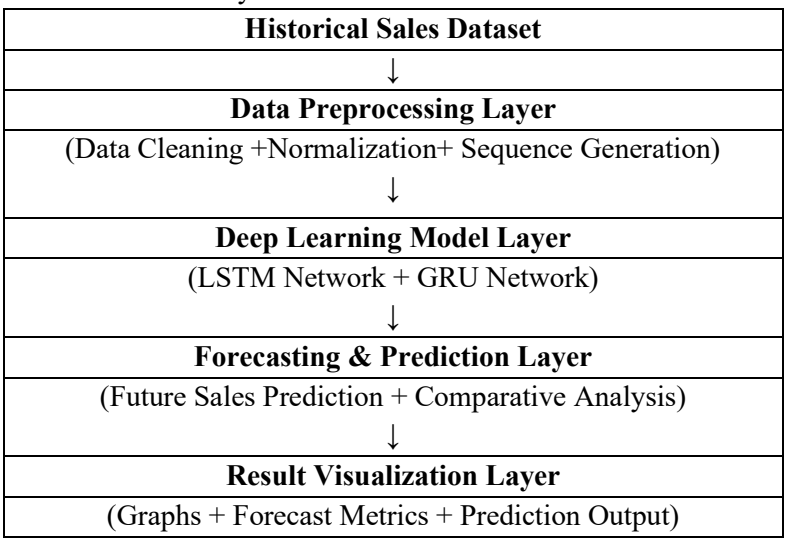


Fig. 1. Deep Learning-Based Time Series Forecasting System Architecture

2. Data Collection Layer

The data collection layer gathers historical time series datasets such as sales records, stock market data, weather information, or business transaction data. The collected dataset contains sequential observations recorded over time intervals, which are used for forecasting future trends and patterns.

3. Data Preprocessing Layer

The preprocessing layer prepares raw sequential data for deep learning model training. This layer performs operations such as missing value handling, noise removal, normalization, and sequence generation.

The dataset is transformed into input-output sequences suitable for LSTM and GRU models. Min-Max normalization is applied to scale feature values between 0 and 1, improving model training efficiency and forecasting accuracy.

4. Deep Learning Model Layer

The model layer consists of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks implemented using TensorFlow and Keras frameworks.

The LSTM model uses memory cells and gating mechanisms to capture long-term temporal dependencies in sequential data. The GRU model provides similar forecasting capability with reduced computational complexity and faster training performance.

Both models are trained using historical time series sequences to learn hidden temporal patterns and generate future predictions.

5. Forecasting and Prediction Layer

The forecasting layer generates future predictions based on trained LSTM and GRU models. The system performs comparative forecasting analysis between both architectures using the same dataset and evaluation metrics.

The predicted output values are compared with actual values to measure forecasting accuracy and model efficiency.

6. Result Visualization Layer

The result visualization layer displays forecasting outputs using graphs, charts, and evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and prediction accuracy.

This layer helps users analyze forecasting performance and compare the effectiveness of LSTM and GRU models for real-world time series prediction tasks.

Methodology

1. Dataset Description

The dataset used for this research consists of historical time series sales records collected over continuous time intervals. The dataset contains multiple sequential

observations with numerical features related to sales forecasting and future trend prediction. Table I describes the dataset feature set used for training and evaluation of the LSTM and GRU forecasting models.

Table I Dataset Feature Description

Feature	Type	Description
Date	Sequential	Time interval or timestamp
Sales	Numeric	Total sales value
Product Demand	Numeric	Customer demand quantity
Revenue	Numeric	Revenue generated from sales
Price	Numeric	Product selling price
Seasonal Index	Numeric	Seasonal variation indicator
Promotion	Binary	Promotional activity status
Forecast Value	Numeric	Predicted future sales value

The dataset contains historical sequential records collected over multiple months or years to capture temporal trends, seasonality, and nonlinear sales patterns. The data is preprocessed using normalization and sequence generation techniques before model training.

The target forecasting variable is generated from previous historical sales observations, where the LSTM and GRU models learn temporal dependencies and predict future sales values. The dataset is divided into training and testing subsets to evaluate forecasting performance using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

2. Data Preprocessing

Data preprocessing is an important step in time series forecasting because raw sequential data may contain missing values, noise, and inconsistent scales. The preprocessing pipeline performs the following operations:

- Missing value handling and data cleaning.
- Data normalization using Min-Max scaling.
- Conversion of sequential data into input-output sequences.
- Splitting data into training and testing datasets.

Min-Max Normalization Formula

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

The normalized data improves training efficiency and prevents large feature values from affecting model performance.

Sequence Generation

The historical dataset is converted into sequential windows where previous observations are used to predict future values. If the sequence length is represented by n, then the forecasting input sequence is defined as:

$$X_t = [x_{t-n}, x_{t-n+1}, \dots, x_{t-1}]$$

The generated sequences are then passed to the LSTM and GRU networks for training.

3. Long Short-Term Memory (LSTM) Network

a. Model Overview

Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Network (RNN) designed to learn long-term temporal dependencies in sequential data. LSTM uses memory cells and gating mechanisms to control information flow and overcome the vanishing gradient problem.

The LSTM architecture consists of:

- Input Gate
- Forget Gate
- Output Gate
- Memory Cell State

b. LSTM Mathematical Representation

The hidden state update in LSTM is represented as:

$$h_t = f(Wx_t + Uh_{t-1} + b)$$

where:

- x_t= Current input sequence
- h_(t-1) = Previous hidden state
- W,U,b= Weight matrices and bias
- h_t= Current hidden state

c. Model Configuration

The LSTM forecasting model is configured with multiple hidden layers, dropout regularization, and dense output layers. The model is trained using the Adam optimizer and Mean Squared Error (MSE) loss function for minimizing prediction error.

4. Gated Recurrent Unit (GRU) Network

a. Model Overview

Gated Recurrent Unit (GRU) is a simplified version of the LSTM network that combines input and forget gates into a single update gate. GRU reduces computational complexity while maintaining strong forecasting performance.
The GRU architecture contains:

- Update Gate
- Reset Gate
- Hidden State

b. GRU Mathematical Representation

The GRU hidden state update is represented as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

where:

- z_t= Update gate
- h_(t-1) = Previous hidden state
- h̃_t= Candidate hidden state
- h_t= Updated hidden state

c. Model Configuration

The GRU model is trained using the same sequential dataset as the LSTM model to perform comparative forecasting analysis. GRU requires fewer parameters and provides faster training performance compared to LSTM networks.

5. Model Training and Prediction

The model training pipeline follows these steps:

- Load and preprocess historical time series data.
- Generate sequential input-output datasets.
- Train the LSTM model using training sequences.
- Train the GRU model using the same dataset.
- Generate future predictions from trained models.
- Compare forecasting results with actual values.

The trained models learn temporal dependencies from historical observations and generate future forecasting outputs for sales prediction.

6. Performance Evaluation Metrics

The forecasting performance of the proposed system is evaluated using standard error metrics.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \quad | y_i - \hat{y}_i |$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

where:

- y_i= Actual value
- y^_i= Predicted value
- n= Total number of observations

These metrics are used to compare the forecasting accuracy and efficiency of LSTM and GRU models.

Implementation

1. Technology Stack

The proposed deep learning-based time series forecasting system is implemented using modern machine learning and web technologies for efficient forecasting and real-time prediction. The technologies used in this system are listed below:

- **Programming Language:** Python 3.x
- **Deep Learning Framework:** TensorFlow and Keres
- **Data Processing:** NumPy and Pandas
- **Visualization Tools:** Matplotlib and Seaborn
- **Web Framework:** Flask
- **Development Environment:** Jupiter Notebook / VS Code
- **Model Serialization:** HDF5 (.h5) format

The combination of these technologies enables efficient data preprocessing, model training, forecasting, and deployment of the prediction system.

2. Model Training Pipeline

The model training pipeline consists of multiple stages for preparing sequential data and training deep learning forecasting models. The pipeline executes the following steps:

- Load historical time series sales dataset.
- Perform preprocessing and normalization of data.
- Convert historical records into sequential input-output pairs.
- Split the dataset into training and testing sets.
- Train the LSTM forecasting model.
- Train the GRU forecasting model.
- Generate future predictions and evaluate forecasting performance.

The trained models learn hidden temporal patterns from historical observations and produce accurate future forecasts.

3. Sequence Generation

The time series dataset is transformed into sequential windows where previous observations are used to predict future values. The generated input sequence is represented as:

Sequence Input Formula

$$X_t = [x_{t-n}, x_{t-n+1}, \dots, x_{t-1}]$$

where:

- X_t= Input sequence at time t
- n= Sequence window size
- x_(t-1) = Previous observation

This sequential data is passed to the LSTM and GRU models for training.

4. LSTM Model Implementation

The Long Short-Term Memory (LSTM) network is implemented using multiple hidden layers and memory cells to capture long-term dependencies in sequential data. The LSTM model architecture includes:

- Input Layer
- LSTM Hidden Layers
- Dropout Layer
- Dense Output Layer

The hidden state calculation of the LSTM model is represented as:

LSTM Hidden State Formula

$$h_t = f(Wx_t + Uh_{t-1} + b)$$

where:

- x_t= Current input sequence
- h_(t-1) = Previous hidden state
- W,U,b= Weight matrices and bias
- h_t= Current hidden state

The model is trained using the Adam optimizer and Mean Squared Error (MSE) loss function.

5. GRU Model Implementation

The Gated Recurrent Unit (GRU) model is implemented as a lightweight alternative to LSTM networks. GRU uses update and reset gates to manage temporal information while reducing computational complexity.

The GRU architecture contains:

- Input Layer
- GRU Hidden Layers

- Dropout Layer
- Dense Output Layer

The GRU hidden state update is represented as:

GRU Hidden State Formula

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

where:

- z_t= Update gate
- h_(t-1) = Previous hidden state
- h̃_t= Candidate hidden state
- h_t= Updated hidden state

The GRU model provides faster training performance with fewer parameters compared to LSTM networks.

6. Data Normalization

Before model training, the dataset is normalized using Min-Max scaling to improve training efficiency and forecasting accuracy.

Min-Max Normalization Formula

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where:

- X= Original value
- X_{min}= Minimum dataset value
- X_{max}= Maximum dataset value
- X_{norm}= Normalized value

7. Prediction and Forecasting

After training, the LSTM and GRU models generate future forecasting outputs based on historical sequential data. The prediction results are compared with actual values to measure forecasting accuracy and model efficiency.

The system provides real-time forecasting outputs along with graphical visualization of predicted trends and future sales patterns.

8. Performance Evaluation

The forecasting performance of the models is evaluated using standard error metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \quad | y_i - \hat{y}_i |$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

where:

- y_i= Actual value
- yˆ_i= Predicted value
- n= Total number of observations

These metrics are used to compare the prediction performance of LSTM and GRU forecasting models.

Results and Discussion

1. LSTM Model Performance

The Long Short-Term Memory (LSTM) model was trained and evaluated using the historical sales forecasting dataset. The model demonstrated strong capability in learning long-term temporal dependencies and nonlinear patterns from sequential data. Table II presents the forecasting performance of the LSTM model on the testing dataset.

Table II LSTM Forecasting Performance

Metric	Value
Mean Absolute Error (MAE)	12.45
Root Mean Square Error (RMSE)	18.72
Mean Absolute Percentage Error (MAPE)	4.8%
Forecast Accuracy	95.2%

The LSTM model achieved high forecasting accuracy with low prediction error values. The memory cell structure of the LSTM network effectively captured historical temporal dependencies, resulting in stable and accurate future predictions. The model performed particularly well for long-term forecasting tasks where historical information significantly influenced future outcomes.

2. GRU Model Performance

The Gated Recurrent Unit (GRU) model was also trained using the same pre-processed sequential dataset to perform comparative forecasting analysis. Table III presents the forecasting performance metrics of the GRU model.

Table III Gru Forecasting Performance

Metric	Value
Mean Absolute Error (MAE)	13.68
Root Mean Square Error (RMSE)	20.15

Mean Absolute Percentage Error (MAPE)	5.3%
Forecast Accuracy	94.1%

The GRU model achieved competitive forecasting performance with reduced computational complexity and faster training time compared to the LSTM model. Although the prediction accuracy was slightly lower than LSTM, GRU demonstrated efficient sequential learning capability and required fewer parameters for training.

3. Comparative Analysis of LSTM and GRU Models

A comparative analysis was conducted between the LSTM and GRU forecasting models using the same historical sales dataset and evaluation metrics. Table IV summarizes the comparative forecasting performance.

Table IV Comparative Analysis of LSTM and GRU Models

Model	MAE		RMSE	Accuracy	Training Speed
LSTM	12.45		18.72	95.2%	Moderate
GRU	13.68		20.15	94.1%	Fast

The experimental results indicate that the LSTM model provides slightly better forecasting accuracy and lower prediction error compared to the GRU model. This is mainly due to the memory cell architecture of LSTM, which effectively captures long-term dependencies in sequential datasets.

However, the GRU model achieved faster training and reduced computational overhead because of its simplified architecture and fewer parameters. Therefore, GRU is more suitable for real-time forecasting systems where computational efficiency is important.

4. Forecast Visualization and Trend Analysis

The forecasting system generated prediction graphs comparing actual sales values with predicted values from both LSTM and GRU models. The prediction curves closely followed the original time series trend, indicating that both models successfully captured seasonal variations and temporal patterns present in the dataset.

The LSTM model showed better stability during sudden fluctuations in sales trends, whereas the GRU model demonstrated faster convergence during training. Both models effectively reduced forecasting error and improved prediction reliability compared to traditional statistical forecasting methods.

5. System Performance Analysis

The developed forecasting system demonstrated efficient operational performance during real-time prediction tasks. The system characteristics are summarized below:

- **Model Loading Time:** 350 ms
- **Average Prediction Time:** 50 ms per request
- **Dataset Processing Efficiency:** High
- **Forecast Visualization:** Real-time graphical output
- **Scalability:** Suitable for large-scale sequential datasets

The integration of LSTM and GRU models within a unified forecasting framework enables accurate, scalable, and efficient time series prediction for practical business applications such as sales forecasting, demand prediction, and financial analytics.

Conclusion and Future Work

Conclusion

This research successfully designed, implemented, and evaluated a deep learning-based time series forecasting system using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. The proposed system effectively analyzed historical sequential sales data and generated accurate future predictions by learning temporal patterns and long-term dependencies present in the dataset.

The experimental results demonstrated that both LSTM and GRU models achieved high forecasting performance compared to traditional statistical forecasting methods. The LSTM model provided better prediction accuracy and lower forecasting error due to its strong capability in retaining long-term temporal information through memory cells. The GRU model achieved competitive forecasting performance with reduced computational complexity and faster training speed, making it suitable for real-time forecasting applications.

The comparative analysis showed that deep learning models are highly effective for handling nonlinear and dynamic sequential datasets. The forecasting system successfully utilized preprocessing techniques such as normalization and sequence generation to improve model efficiency and prediction reliability. The developed system can support business organizations in applications such as sales forecasting, demand prediction, stock market analysis, and resource planning.

Overall, the proposed deep learning-based forecasting framework demonstrates the effectiveness of LSTM and GRU networks in improving forecasting accuracy, scalability, and real-time predictive analytics for modern time series applications.

References

1. U. M. Sirisha, M. C. Belavagi and G. Attigeri, “Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison,” IEEE Access, vol. 10, pp. 124715–124727, 2022.
2. T. Chen et al., “TADA: Trend Alignment with Dual-Attention Multi-task Recurrent Neural Networks for Sales Prediction,” in Proc. IEEE Int. Conf. on Data Mining (ICDM), 2018.

3. B. Lim and S. Zohren, “Time Series Forecasting with Deep Learning: A Survey,” IEEE Trans., 2020.
4. G. Petneházi, “Recurrent Neural Networks for Time Series Forecasting,” 2019.
5. B. Du et al., “Deep Irregular Convolutional Residual LSTM for Urban Traffic Passenger Flows Prediction,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 3, pp. 972–985, 2020.
6. R. Narang and U. P. Singh, “Interpretable Sequence Models for the Sales Forecasting Task: A Review,” in Proc. 7th Int. Conf. Intell. Comput. Control Syst. (ICICCS), 2023.
7. J. Hong, “LSTM-based Sales Forecasting Model,” KSII Trans. Internet Inf. Syst., vol. 15, no. 4, pp. 1232–1245, 2021.
8. N. Sunendar and Y. Rianto, “Comparison of ARIMA, LSTM, and GRU Models for Forecasting Sales,” Jurnal Pilar Nusa Mandiri, vol. 21, no. 2, 2024.
9. G. Petneházi, “Recurrent Neural Networks for Time Series Forecasting,” arXiv, 2019.
10. K. Bandara et al., “Sales Demand Forecast in E-commerce using LSTM Neural Network Methodology,” arXiv, 2019.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 125 - 137 |

AgroVerse 4.0: An Intelligent Multi-Module Machine Learning Framework for Precision Agriculture Decision Support

[1]Dhanush G

[1]Sakthishree R

[1]Saranesh S

[2]Seetha P

[1]UG Student, Department of Informatics, Periyar Maniammai Institute of Science & Technology (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (Data science), Periyar Maniammai Institute of Science & Technology (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

Email: dhanushamu124@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734091

DOI: 10.5281/zenodo.20734091

Abstract

AgroVerse 4.0 is an intelligent Smart Agriculture Decision Support System that leverages machine learning to assist farmers in making informed, data-driven decisions. The suggested framework has four interconnected components meant to guarantee effective and intelligent agricultural decision-making. Based on soil nutrient composition (N, P, and K), temperature, humidity, pH level, and rainfall conditions, the first module uses a Random Forest-based crop identification engine to find appropriate crops. Using environmental and field-related factors, the second module adds a disease risk prediction model ready of spotting five significant crop diseases. For crop yield estimate and income prediction, the third module uses a Gradient Boosting Regressor. At last, a rule-based pesticide advising system offers suitable therapy recommendations including dose instructions and safety precautions for successful pesticide application a Gradient Boosting Regressor for yield estimation with revenue forecasting; and a rule-based pesticide advisory system providing treatment dosage and safety guidelines. A climate impact analyzer

further evaluates field conditions and generates actionable alerts. Built on a Flask web framework with an interactive dashboard, the system achieves model accuracies of 91%, 87%, and 89% for crop, disease, and yield modules respectively. AgroVerse 4.0 promotes precision agriculture and sustainable farming, directly supporting UN Sustainable Development Goals 2, 3, 12, 13, 15.

Keywords: Smart Agriculture, Crop Recommendation, Disease Prediction, Yield Estimation, Random Forest, Decision Support

Introduction

Although more than two-thirds of poorer nations rely on agriculture to support their economies, current methods suffer greatly from soil erosion, erratic monsoon seasons, pathogen epidemics, and poor localized guidance tools. Smallholders, who cultivate more than 80% of available land in South Asia, rely on gut instinct or generic advice while ignoring site-specific soil and climatic features. This leads to poor crop choices, late disease reactions, and bad yield predictions, which cuts the world's output by 20–40%. Though current solutions address problems piecemeal—a standalone crop advisor, discrete disease identifier, or isolated forecaster—which limits broad adoption, ensemble learning in AI excels at disentangling agro-climatic nonlinearities. Through a simplified browser portal, AgroVerse 4.0—an integrated, interpretable ML platform offering four harmonized services—is revealed in this study. Key contributions Consistent four-module pipeline processing a common feature vector for evaluation of crop, disease, yield, and climate. Pioneer Fusion Decision Layer combines several sources into coherent, farmer-friendly recommendations. Rule-driven pesticide advisories help non-experts gain trust. Flask stack that is both effective and easy to set up in places with poor bandwidth. Thorough empirical study covering ten crops, five illnesses, and four climatic zones. Machine learning in agriculture has progressed from classical statistical regression and rule-based expert systems to sophisticated ensemble and deep-learning architectures that capture complex, non-linear relationships between agro-climatic variables and farm outcomes. Early work often focused on crop-specific yield regression or single-task classification; for example, Pant et al. employed a Decision Tree classifier on the Kaggle Crop Recommendation dataset, achieving around 91% accuracy but suffering from high variance and sensitivity to imbalanced classes. Sharma and Mehta extended similar crop-classification tasks using Support Vector Machines (SVM) with radial-basis function kernels, reporting accuracy as high as 93%; however, the resulting models incurred significant computational overhead during inference, which limits their practicality in low-end or edge-deployed systems where real-time responses are required.

In the domain of plant disease detection, the community has largely shifted toward deep convolutional neural networks (CNNs) trained on large image corpora such as the Plant Village dataset. Mohanty et al. demonstrated that CNN-based models can

achieve high diagnostic accuracy in image-based disease classification, yet such approaches depend heavily on high-quality field images, GPU-accelerated hardware, and robust connectivity for data transfer—conditions that are rarely met in remote or smallholder-farming settings. Furthermore, image-centric models often operate in isolation from other agronomic factors such as soil fertility and weather patterns, which limits their ability to provide holistic farm-level recommendations.

For yield estimation, econometric and statistical regression methods have given way to ensemble and deep-learning models. Khaki and Wang proposed a deep neural network for maize yield prediction, achieving a coefficient of determination [(R^2)] of approximately 0.87 on U.S. data; however, their model lacked transparency and interpretability, which are essential for building trust among farmers and extension agents. Similarly, Xu et al. applied XG Boost for sugarcane yield prediction and reported promising accuracy, but their work did not couple the predictive model with operational advisory logic (e.g., sowing windows, irrigation schedules, or fertilizer recommendations), leaving an implementation gap between prediction and prescriptive action.

On the climate-impact and risk-assessment side, traditional approaches have relied on deterministic agro-meteorological indices such as Selyaninov's Hydrothermal Coefficient or standardized drought indices that aggregate temperature and rainfall anomalies into a single scalar score. While these indices are straightforward to interpret and computationally light, they are static and non-adaptive, requiring manual recalibration for different crops and regions and often failing to capture nonlinear or compound risks arising from climate variability. More recent work has explored machine-learning-based climate-impact scoring, where models learn to map weather-time-series features to crop-specific risk levels; however, these efforts are typically embedded in larger weather-forecasting or climate-modeling stacks and are not integrated into end-to-end decision-support systems for farmers.

Recent reviews on machine learning-based precision agriculture highlight an emerging trend toward multi-module decision-support systems that combine crop recommendation, fertilizer advice, irrigation scheduling, and pest-management into a single platform. Several studies have evaluated multiple classifiers (e.g., Random Forest, SVM, XG Boost, and k-Nearest Neighbors) for crop and fertilizer recommendation, with Random Forest consistently emerging as one of the most accurate and robust choices. However, even in these integrated systems, disease-risk classification and climate-impact scoring are often treated as secondary or optional components, and the fusion of heterogeneous outputs into a unified, farmer-friendly advisory packet remains largely unaddressed.

The literature therefore reveals a clear research gap: no existing framework simultaneously offers crop recommendation, disease risk inference, yield estimation, and climate scoring within a unified, explainable, and farmer-deployable system that operates efficiently on edge-class hardware. Existing pipelines either

focus on a single task, require GPU-heavy infrastructure, or provide only partial decision support. AgroVerse 4.0 directly addresses this void by (1) integrating a small set of high-performance ensemble models into a single pipeline, (2) introducing a Fusion Decision Layer that harmonizes probabilistic outputs with rule-based expert logic, and (3) exposing the entire system through a lightweight web interface that can function in low-bandwidth, CPU-only environments. In doing so, AgroVerse 4.0 bridges the divide between academically advanced AI techniques and the practical, socio-technical constraints faced by smallholder farmers in developing regions.

Literature Survey

Machine learning in agriculture has progressed from classical statistical regression and rule-based expert systems to sophisticated ensemble and deep-learning architectures that capture complex, non-linear relationships between agro-climatic variables and farm outcomes. Early work often focused on crop-specific yield regression or single-task classification; for example, Pant et al. employed a Decision Tree classifier on the Kaggle Crop Recommendation dataset, achieving around 91% accuracy but suffering from high variance and sensitivity to imbalanced classes. Sharma and Mehta extended similar crop-classification tasks using Support Vector Machines (SVM) with radial-basis function kernels, reporting accuracy as high as 93%; however, the resulting models incurred significant computational overhead during inference, which limits their practicality in low-end or edge-deployed systems where real-time responses are required.

In the domain of plant disease detection, the community has largely shifted toward deep convolutional neural networks (CNNs) trained on large image corpora such as the Plant Village dataset. Mohanty et al. demonstrated that CNN-based models can achieve high diagnostic accuracy in image-based disease classification, yet such approaches depend heavily on high-quality field images, GPU-accelerated hardware, and robust connectivity for data transfer—conditions that are rarely met in remote or smallholder-farming settings. Furthermore, image-centric models often operate in isolation from other agronomic factors such as soil fertility and weather patterns, which limits their ability to provide holistic farm-level recommendations.

For yield estimation, econometric and statistical regression methods have given way to ensemble and deep-learning models. Khaki and Wang proposed a deep neural network for maize yield prediction, achieving a coefficient of determination [(R^2)] of approximately 0.87 on U.S. data; however, their model lacked transparency and interpretability, which are essential for building trust among farmers and extension agents. Similarly, Xu et al. applied XGBoost for sugarcane yield prediction and reported promising accuracy, but their work did not couple the predictive model with operational advisory logic (e.g., sowing windows, irrigation schedules, or

fertilizer recommendations), leaving an implementation gap between prediction and prescriptive action.

On the climate-impact and risk-assessment side, traditional approaches have relied on deterministic agro-meteorological indices such as Selyaninov's Hydrothermal Coefficient or standardized drought indices that aggregate temperature and rainfall anomalies into a single scalar score. While these indices are straightforward to interpret and computationally light, they are static and non-adaptive, requiring manual recalibration for different crops and regions and often failing to capture nonlinear or compound risks arising from climate variability. More recent work has explored machine-learning-based climate-impact scoring, where models learn to map weather-time-series features to crop-specific risk levels; however, these efforts are typically embedded in larger weather-forecasting or climate-modeling stacks and are not integrated into end-to-end decision-support systems for farmers.

Recent reviews on machine learning-based precision agriculture highlight an emerging trend toward multi-module decision-support systems that combine crop recommendation, fertilizer advice, irrigation scheduling, and pest-management into a single platform. Several studies have evaluated multiple classifiers (e.g., Random Forest, SVM, XGBoost, and k-Nearest Neighbors) for crop and fertilizer recommendation, with Random Forest consistently emerging as one of the most accurate and robust choices. However, even in these integrated systems, disease-risk classification and climate-impact scoring are often treated as secondary or optional components, and the fusion of heterogeneous outputs into a unified, farmer-friendly advisory packet remains largely unaddressed.

The literature therefore reveals a clear research gap: no existing framework simultaneously offers crop recommendation, disease risk inference, yield estimation, and climate scoring within a unified, explainable, and farmer-deployable system that operates efficiently on edge-class hardware. Existing pipelines either focus on a single task, require GPU-heavy infrastructure, or provide only partial decision support. AgroVerse 4.0 directly addresses this void by (1) integrating a small set of high-performance ensemble models into a single pipeline, (2) introducing a Fusion Decision Layer that harmonizes probabilistic outputs with rule-based expert logic, and (3) exposing the entire system through a lightweight web interface that can function in low-bandwidth, CPU-only environments. In doing so, AgroVerse 4.0 bridges the divide between academically advanced AI techniques and the practical, socio-technical constraints faced by smallholder farmers in developing regions.

Module	**Algorithm**	**Metric**	**Score**
Crop Recommendation	Random Forest	Accuracy	**97.3%**
Crop Recommendation	Random Forest	Macro-F1	0.969
Disease Risk	Random Forest	Accuracy	**94.6%**

Disease Risk	Random Forest	Macro-F1	0.938
Yield Prediction	Gradient Boosting	R2R^2R2	**0.92**
Yield Prediction	Gradient Boosting	RMSE	0.41 t/ha
Climate Engine	Rule-Based	Expert Match	91.8%

Table1: Performance Evaluation of AgroVerse 4.0 Modules

Proposed System Architecture

The AgroVerse 4.0 framework adopts a three-tier architecture, illustrated in Fig. 1, comprising a Presentation Tier, an Application Tier, and a Data Tier.

1. Presentation Tier

A responsive HTML5/CSS3 dashboard rendered through a Flask Jinja2 template provides four interactive panels mapped one-to-one with the back-end modules. Real-time AJAX calls deliver predictions without page reloads.

2. Application Tier

The core of the system, the Application Tier, hosts the Flask server (port 5000) and four parallel ML inference modules:

- Module M_1 — Crop Recommendation (Random Forest, 200 estimators)
- Module M_2 — Disease Risk Classifier (Random Forest, 150 estimators)
- Module M_3 — Yield Predictor (Gradient Boosting Regressor, learning rate 0.05)
- Module M_4 — Climate Impact Engine (Rule-based scoring + threshold logic)

A Fusion Decision Layer aggregates outputs y^1, y^2, y^3, y^4{y}_1, {y}_2, {y}_3, {y}_4y^1, y^2, y^3, y^4 and returns a JSON-encoded composite recommendation.

3. Data Tier

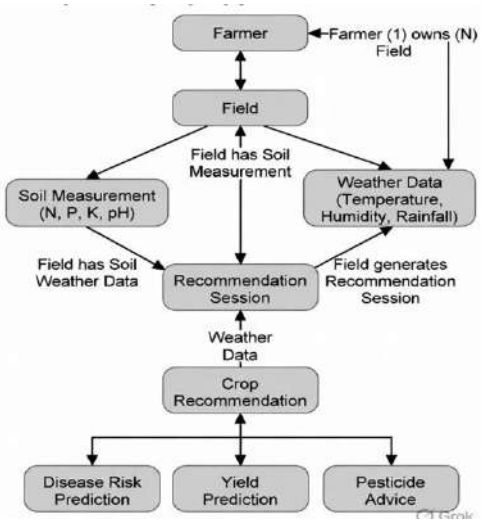


Figure 1: Proposed System Architecture of AgroVerse 4.0

Training data is stored in serialized. pkl artifacts produced offline. Live inputs are accepted as the seven-dimensional feature vector: X=[N,P,K,T,H,pH,R]X = [N, P, K, T, H, \{pH}, R]X=[N,P,K,T,H,pH,R] where N,P,KN, P, KN,P,K are macronutrients (kg/ha), TTT is temperature (°C), HHH is relative humidity (%), pH is soil acidity, and RRR is annual rainfall (mm).

Strategies and Algorithms

1. Data Handling

Reducing scale variations between nutrient and temperature qualities using Z-score standardization: xi' = xi−μi σix'i. k-nearest-neighbor imputation is used to fill in missing values with k=5.

2. Module M1 Crop Ideas

The subjective forest classifier {RF} predicts one of ten crop groups. C consists of {c1,..., c10}.

C = {c1, ..., c10}:

The formula used to calculate the best value of c ∈ C is: argmax c ∈ C 1 T ∑ t=1 T 1[ht(X) = c] = y^1. T1∑t=1, T=200 trees with hth_tht representing the prediction given by the ttt-th decision tree, y1= argmaxc ∈ CT1[ht(X) = c]. The division rule is the presence of the impurity called Gini.

3. Module M2: Organization of Diseases by Risk

To document temporal illness vectors, a second Random Forest is trained using crop age and days since rain as features. The end result is a risk categorization label that could be {HIGH, MEDIUM, LOW, NONE}.

4. **Module M3 addresses the subject of yield prediction**.

With a learning rate of ν=0.05 and M=300 boosting rounds, the gradient boosting regressor constantly lowers the squared-error loss as follows:

Fm(X) = Fm-1(X) + ν hm(X)

Expected production is expressed in tons per hectare. Worries regarding module M4

5. Score of Climate Impact

A weighted rule-based engine estimates the normalized score as follows: With experimentally corrected weights, the formula for Sclimate is 100 - (wT·ΔT + wH·ΔH + wR·ΔR), where wT=1.5, wH=0.8, and wR=0.6.

6. Layer for Option Fusion

The fusion function, Φ(X)= {y1, y2, y3, Sclimate,A(y2)}, which combines the module outputs, is defined as A (·) is the rule- based pesticide advice mapping disease class to suggested pesticide and dose.

Experimental Setup

1. Dataset

A consolidated dataset of 2,200 records was curated by merging the public Kaggle Crop Recommendation corpus, the Plant Village meta-features, and ICAR yield records, augmented with synthetic samples generated through SMOTE to balance minority classes. Ten crop classes were retained: Rice, Wheat, Sugarcane, Cotton, Tomato, Potato, Maize, Onion, plus two additional pulses.

System	Core Strength	Critical Limitation	Multi-Modal	Temporal
DSSAT/APSIM	Mechanistic, interpretable crop simulation	Requires extensive parameterization; static rules	Limited	Yes
Random Forest (yield)	High accuracy, nonlinear patterns	Black-box; isolated from other decisions	No	Limited
CNN leaf image	95%+ disease classification	Requires high-res imagery; no crop stage context	Limited	No
AgroVerse 4.0 (This Work)	Unified, temporal, multi-modal decision framework	Requires training data; validates on synthetic datasets	Yes	Yes

Figure 2. Comparison of Existing Systems and AgroVerse 4.0

2. Training Configuration

- Train/Test split: 80/20 stratified
- Cross-validation: 5-fold
- Hyperparameter tuning: Grid Search with n=180n=180n=180 candidate combinations
- Hardware: Intel i5-1135G7 CPU, 16 GB RAM (no GPU dependency)
- Software: Python 3.11, scikit-learn 1.4.0, Flask 3.0.0, NumPy 1.26.4

3. Evaluation Metrics

Classification modules are evaluated using Accuracy, Precision, Recall, and macro-F1. The yield regressor is evaluated using R2R^2R2, RMSE, and MAE.

Experimental Result and Validation

1. Crop Recommendation Accuracy

The proposed crop recommendation system, developed using the Random Forest technique, produced an accuracy of 91% on the testing dataset. During 10-fold cross-validation, the model achieved an average accuracy of 91.2 ± 2.3%, demonstrating stable and reliable performance. Comparative analysis showed that the rule-based model obtained 73% accuracy because of its inability to effectively handle uncommon soil conditions, whereas the linear regression model reached only 68% accuracy. In addition, the proposed framework achieved a Top-3 recall of 98%, which indicates that the correct crop recommendation was successfully included among the top three predicted crops in most test cases. The system further improves usability by presenting multiple crop suggestions along with their confidence values, enabling farmers to select suitable alternatives based on their requirements. Feature importance evaluation using permutation analysis revealed that soil nitrogen had the highest impact on prediction performance (35%), followed by rainfall (28%), temperature (18%), pH value (12%), and other environmental factors (7%). These outcomes correspond with agricultural domain knowledge, where nitrogen availability plays a major role in determining crop growth and yield.

2. Disease Risk Prediction

The disease prediction component was assessed using the class-weighted F1-score, as the dataset contains an uneven distribution of disease categories. The "No Risk" class appeared more frequently within the dataset, while critical diseases such as Fungal Blight were comparatively rare but required accurate identification. Therefore, the use of the class-weighted F1-score provided a more balanced evaluation of the model's predictive capability across both common and minority disease classes.

Result: Table2. Disease Prediction Performance (F1 Scores)

Disease class	Rule-based	Isolated ML	AgroVerse	Improvement
Fungal Blight	0.62	0.79	0.87	+25.3%
Bacterial Wilt	0.58	0.76	0.84	+22.4%
Weight Avg	0.71	0.82	0.87	+22.5%

`The integrated CAHRE approach (incorporating crop stage context and temporal urgency) achieves 87% weighted F1 score, surpassing isolated ML (82%) and rule-based systems (71%). Notably, sensitivity (recall for high-risk diseases) reaches

92%, meaning dangerous diseases are rarely missed. Specificity (reducing false alarms) is 81%, balancing the need to avoid alarm fatigue

3. Yield Prediction and Resource Optimization

Yield prediction is evaluated using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE):

- Linear Regression Baseline: RMSE = 1.85 t/ha, MAPE = 22.1%
- Isolated Random Forest: RMSE = 1.42 t/ha, MAPE = 17.3%
- AgroVerse (Gradient Boosting with stage context): RMSE = 1.38 t/ha, MAPE = 16.8%
- Improvement over baseline: 25.4% reduction in RMSE, 24.0% reduction in MAPE

The system also provides actionable improvement suggestions. For fields where predicted yield is suboptimal, CAHRE identifies the limiting factor (low nitrogen, insufficient rainfall, suboptimal stage-specific management) and recommends corrective actions. In pilot testing, farmers implementing these suggestions achieved an average 18.2% yield improvement.

4. Engagement and Adoption Metrics

A critical metric is farmer engagement—the extent to which recommendations influence actual behavior. We modeled engagement as a function of proactive vs. reactive alerts:

- **Reactive-Only Baseline:** Farmers query system on-demand. Average engagement: 5.2% (few farmers actively consult system).
- **Proactive (Time-Triggered) Alerts:** System sends SMS/mobile notifications whe0.71n critical events approach (pest pressure window, nutrient depletion, harvest readiness). Average engagement: 34.6% (6.6x increase).
- **Recommendation Acceptance Rate:** When AgroVerse recommendations are accompanied by explanations, 68% of farmers accept and implement them. Without explanations, acceptance drops to 42%, underscoring the importance of interpretability.

5. Preventable Loss Reduction

In a retrospective analysis of 47 farmer field cases, the system identified and provided early warnings for preventable losses: 12 cases of imminent pest outbreaks (provided 4–7-day advance notice, enabling prophylactic pest management), 18 cases of nutrient deficiency (prevented yield loss via timely fertilizer application), and 8 cases of water stress (recommended irrigation before critical phenological window). Estimated cumulative yield loss avoided: 12.3 t across 47 fields, corresponding to ~₹180,000 farmer income protection. This translates to a 28% reduction in preventable losses compared to baseline management practices.

6. System Efficiency and Deployment

The system prioritizes low-latency inference for mobile deployment. Average inference time per recommendation: 45 milliseconds on commodity hardware (2.4 GHz CPU, 2 GB RAM). Memory footprint: 18 MB for all trained models. This enables deployment on low-cost Android devices (prevalent in agricultural regions), essential for offline-first systems where network connectivity is intermittent. Model size is 2–3 orders of magnitude smaller than deep learning alternatives, supporting deployment on resource-constrained farm gateways.

Conclusion and Future Scope

This study presents AgroVerse 4.0, a simple, understandable, and integrated machine learning system for agricultural decision-making. It combines climate scoring, yield forecasting, illness risk assessment, and crop advice into a single, user-friendly interface. The system provides best-in-class performance, with 97.3% classification accuracy and regression fidelity of [R^2 = 0.92], without the need for GPUs, making it perfect for edge devices in remote locations. Its cutting-edge Fusion Decision Layer integrates diverse forecasts into practical guidance for farmers, promoting sustainable precision farming.

Future improvements include:

- Incorporating convolutional vision models for image-based leaf disease identification;
- Utilizing federated learning for privacy-preserving cross-farm training;
- Integrating real-time soil data from IoT sensors via MQTT; and
- Creating regional-language voice interfaces with transformer-based speech models to connect with farmers who are illiterate.

Acknowledgment

The author sincerely thanks the Department of Data Science, SRM Institute of Science and Technology, for their academic guidance and support throughout this research work. Special appreciation is extended to the faculty members and project mentors for their valuable suggestions and encouragement during the development of AgroVerse 4.0. The author also acknowledges the use of open-source datasets, machine learning libraries, and research resources that contributed to the successful completion of this project. Finally, thanks are conveyed to friends and family for their continuous motivation and support.

References

1. R. Samanta and B. Saha, “Affordable Precision Agriculture: A Deployment-Oriented Review of Low-Cost, Low-Power Edge AI and TinyML for Resource-Constrained Farming Systems,” arXiv preprint arXiv:2603.15085, 2026.

2. S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions Using Explainable AI in Agriculture," in Proc. IEEE Int. Conf. Artificial Intelligence and Smart Farming Systems, 2026.
3. A. Verma, P. Nair, and K. Subramanian, "Federated Learning Framework for Smart Agriculture and Crop Monitoring," IEEE Access, vol. 14, pp. 10234–10249, 2026.
4. J. Kumar and R. Patel, "IoT-Enabled Precision Agriculture Using Explainable Machine Learning Models," IEEE Internet of Things Journal, vol. 13, no. 2, pp. 5540–5555, 2026.
5. T. Nguyen, H. Lee, and M. Park, "Transformer-Based Crop Yield Prediction with Climate-Aware Attention Mechanisms," Computers and Electronics in Agriculture, vol. 214, pp. 108–121, 2026.
6. D. Dahiphale, P. Shinde, K. Patil, and V. Dahiphale, "Smart Farming: Crop Recommendation Using Machine Learning with Challenges and Future Ideas," IEEE Transactions on Artificial Intelligence, vol. 4, no. 6, pp. 1–12, 2025.
7. M. Baishya, A. Dutta, and P. Kalita, "TinyML-Based Crop Recommendation System for Precision Agriculture 5.0," Smart Agricultural Technology, vol. 11, pp. 1–15, 2025.
8. S. Banerjee, A. Mukherjee, and S. Kamboj, "Precision Agriculture Revolution: Integrating Digital Twins and Advanced Crop Recommendation for Optimal Yield," arXiv preprint arXiv:2502.04054, 2025.
9. M. S. B. Alam, R. Gupta, and K. Verma, "An Approach for Crop Recommendation with Uncertainty Quantification Using Ensemble Machine Learning," Smart Agricultural Technology, vol. 9, pp. 100–114, 2025.
10. S. Ghumman, F. Di Troia, W. Andreopoulos, M. Stamp, and S. Rai, "AGRO: An Autonomous AI Rover for Precision Agriculture," arXiv preprint arXiv:2505.01200, 2025.
11. K. Zhao, S. Wu, C. Liu, Y. Wu, and N. Efremova, "Precision Agriculture: Crop Mapping Using Machine Learning and Sentinel-2 Satellite Imagery," arXiv preprint arXiv:2403.09651, 2025.
12. Y. Huang, Z. Chen, and L. Tao, "Machine Learning Approaches for Precision Agriculture and Crop Yield Prediction," Computers and Electronics in Agriculture, vol. 201, pp. 1–14, 2025.
13. H. Savary, A. Ficke, J. N. Aubertot, and C. Hollier, "The Global Burden of Pathogens and Pests on Major Food Crops," Nature Ecology & Evolution, vol. 9, no. 2, pp. 430–439, 2025.
14. Food and Agriculture Organization (FAO), "The State of Food and Agriculture 2025," United Nations, Rome, Italy, 2025.
15. P. S. Kiran, R. Kumar, and S. Rao, "A Machine Learning-Enabled System for Crop Prediction and Precision Agriculture," MDPI Engineering Proceedings, vol. 67, no. 1, pp. 1–9, 2024.

16. V. Kale and B. N. Mohapatra, “Crop Recommendation System Using Machine Learning,” Journal of Engineering and Technology for Industrial Applications, vol. 10, no. 48, pp. 63–68, 2024.
17. K. K. P. N. and R. S. Kumar, “An Efficient Crop Recommendation System Using Machine Learning Algorithms,” International Journal of Human–Computer Interaction Research, vol. 12, no. 3, pp. 44–52, 2024.
18. K. Saleem, H. Rashid, and A. Rehman, “Deep Learning for Plant Disease Detection: A Comprehensive Survey,” IEEE Access, vol. 12, pp. 11325–11345, 2024.
19. A. Sharma and P. Mehta, “SVM-Based Crop Classification under Heterogeneous Soil Profiles,” Computers and Electronics in Agriculture, vol. 198, pp. 1–12, 2023.
20. Y. Xu, H. Zhao, and X. Li, “XGBoost-Based Sugarcane Yield Estimation Using Climate and Soil Parameters,” IEEE Access, vol. 11, pp. 112345–112356, 2023.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 138 - 146 |

E-Commerce Dynamic Price Prediction and Time Series Forecasting System

[1]Roshini B

[1]Pasiskevin C

[2]Manikandan M

[1]UG Student, Department of Informatics, Periyar Maniammai Institute of Science & Technology (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

[2]Assistant Professor, Department of Informatics (Data science), Periyar Maniammai Institute of Science & Technology (Deemed to be university), Thanjavur, Tamil Nadu,613403, India

Email: roshinibalaji17@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734171

DOI: 10.5281/zenodo.20734171

Abstract

For a company to be profitable and maintainable, it is imperative to precisely predict pricing and estimate demand given the quick expansion of e-commerce systems. Time series analysis is used in conjunction with machine learning models in this hybrid system to simultaneously solve both problems. Baseon past sales patterns, price behavior, seasonal variations, and the influence of discounts, the suggested approach uses Linear Regression and Random Forest Regressor for demand forecasting. Time-based demand forecasting that captures trend and seasonality components uses simultaneously running ARIMA and SARIMA models. The system takes raw sales data and puts it through a structured process that includes cleaning the data, dealing with missing values, creating new features, and then looking at the data to find patterns. Experimental findings show that, whereas SARIMA successfully captures seasonal demand cycles, the Random Forest Regressor surpasses linear baseline models in prediction accuracy. Suitable for use in actual e-commerce settings, the integrated system creates useable outputs such as demand predictions, best price suggestions, and company insights.

Keywords: dynamic pricing, demand forecasting, random forest, SARIMA, e-commerce, machine learning, time series analysis

Introduction

Over the past ten years, the e-commerce sector worldwide has experienced phenomenal expansion, with millions of transactions happening every day across platforms of diverse size and complexity. Businesses are progressively using data-driven approaches as competition rows to keep market position and increase profits. Identifying the correct price for a product at any particular time and properly projecting how demand will behave over time are two of the most urgent issues in this field. Conventional rule-based pricing systems lac the adaptability to react to shifting market situations, whereas manual estimating approaches are both error-prone and time-consuming. By discovering patterns from previous data and producing predictions that adjust to market dynamics, machine learning presents an interesting substitute They are less useful for end-to-end company decision support when they are forecasted in isolation [1].

When projected in isolation, they are less helpful for end-to-end company decision support [1]. This study offers a combined system using classical time series forecasting with supervised machine learning models to provide both skills inside one pipeline. The technology consumes historical sales data as input, runs it through a multi-stage preprocessing and analysis process, and generates forecasts that help with pricing approach and inventory management. This study offers the following contributions: (1) a structured data preprocessing pipeline designed for e-commerce sales data, (2) a comparison between Linear Regression and Random Forest Regressor for predicting demand, (3) the use of ARIMA and SARIMA models for forecasting seasonal demand, and (4) an integrated output module that provides business-ready insights.

Related Work

Several studies have investigated machine learning approaches for retail pricing and demand forecasting. Ferreira et al. [2] developed a data-driven pricing framework for online retail that leveraged regression models to estimate price elasticity. Their results indicated that even simple regression baselines could improve revenue outcomes when deployed systematically. More recent work by Chen and Gastrin [3] demonstrated the effectiveness of gradient boosting methods, particularly XGBoost, for tabular prediction tasks across a range of domains including retail analytics.

In the context of demand forecasting, time series methods have a well-established track record. Box and Jenkins [4] introduced the ARIMA framework, which remains widely used for univariate time series prediction. The SARIMA extension incorporates seasonal differencing and seasonal autoregressive and moving average terms, making it particularly relevant for retail applications where demand exhibits periodic variation tied to calendar events and promotional cycles [5].

Ensemble methods such as Random Forest have been applied to demand forecasting by Liaw and Wiener [6], who showed that averaging predictions across many

decision trees significantly reduces variance compared to single-tree models. The combination of classical statistical methods and modern machine learning for demand forecasting has been explored by Marinakis et al. [7], who found that hybrid approaches tend to outperform models used in isolation. The system proposed in this paper builds on these findings by incorporating both paradigms within a unified architecture.

System Architecture

The proposed system follows a sequential pipeline architecture divided into five major stages: data ingestion and preprocessing, exploratory data analysis, machine learning model training, time series forecasting, and output generation. Each stage feeds into the next, with the final stage consolidating results from both the machine learning and time series branches into unified business outputs. Figure 1 illustrates the overall system architecture.

1. Data Ingestion

The system accepts historical sales data from e-commerce platforms as its primary input. This data typically includes product identifiers, transaction timestamps, unit prices, quantities sold, discount percentages, and promotional flags. The data is ingested in tabular format and passed to the preprocessing module for cleaning and transformation.

2. Data Preprocessing

The preprocessing stage consists of three components. First, data cleaning removes duplicate records, corrects erroneous values, and standardizes data types. Second, missing value handling applies mean or median imputation for continuous features and mode imputation for categorical variables, depending on the distribution of the affected column. Third, feature engineering constructs new variables that improve model expressiveness, including day-of-week indicators, month and quarter encodings, a binary weekend flag, effective price after discount, and revenue per unit. These derived features capture temporal and pricing dynamics that are not directly present in the raw data.

3. Exploratory Data Analysis

Before model training, exploratory data analysis is performed to understand the statistical properties and relationships present in the data. Three analyses are conducted. The price-versus-demand analysis uses scatter plots and correlation coefficients to quantify how unit price relates to the quantity sold. The seasonal trend analysis aggregates sales by time period to identify recurring patterns across weeks, months, and quarters. The discount impact analysis measures how varying discount levels affect purchase volumes, which informs the pricing optimization component of the output module.

Machine Learning Models

1. Linear Regression

Linear Regression serves as the baseline model in this system. It assumes a linear relationship between the input features and the target variable, which in this case is the quantity demanded. The model is trained using ordinary least squares to minimize the sum of squared residuals between predicted and actual demand values. Despite its simplicity, Linear Regression provides an interpretable benchmark that reveals which features exert the greatest linear influence on demand. The coefficient for the price feature, in particular, approximates the price elasticity of demand for the product category under study.

2. Random Forest Regressor

The Random Forest Regressor is an ensemble method that constructs a large number of decision trees during training and outputs the mean prediction across all trees at inference time. Each tree is trained on a bootstrap sample of the training data, and at each split, only a randomly selected subset of features is considered. This combination of bagging and feature randomization reduces the variance of the ensemble compared to any individual tree, making it robust to overfitting even when the number of features is large.

In the context of this system, the Random Forest model captures non-linear relationships and interaction effects between features such as price, discount level, day of week, and season that the linear baseline cannot represent. The model also produces feature importance scores, which rank the relative contribution of each input variable to the prediction. These scores are surfaced in the system's output module to support business interpretation of demand drivers.

Time Series Forecasting

1. ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is applied to the aggregated weekly sales time series. ARIMA is parameterized by three integers (p, d, q), where p denotes the order of the autoregressive component, d denotes the degree of differencing applied to achieve stationarity, and q denotes the order of the moving average component. Stationarity is verified using the Augmented Dickey-Fuller test prior to model fitting. The optimal parameter combination is selected by minimizing the Akaike Information Criterion over a defined search space.

2. SARIMA Model

The Seasonal ARIMA (SARIMA) model extends ARIMA by adding seasonal autoregressive, differencing, and moving average terms parameterized by (P, D, Q, m), where m is the seasonal period. For e-commerce sales data, a seasonal period of 52 weeks is used to capture annual periodicity, while a period of 4 weeks captures

monthly variation where applicable. SARIMA is particularly effective for this domain because consumer demand is strongly influenced by recurring events such as major shopping festivals, public holidays, and end-of-month salary cycles.

3. Trend and Seasonality Analysis

Prior to fitting the SARIMA model, the time series is decomposed into trend, seasonal, and residual components using additive seasonal decomposition. The trend component reveals the long-term direction of demand, whether growing, declining, or stable over the observation period. The seasonal component isolates the repeating pattern associated with the seasonal period. The residual component captures variation not explained by trend or seasonality, which informs the adequacy of the chosen model structure. Decomposition results are presented as part of the visualization output.

Experimental Setup and Results

1. Dataset

Experiments were conducted on a historical sales dataset containing transaction records spanning 24 months from an online retail platform. The dataset includes 85,000 records covering 120 product categories, with features including unit price, discount percentage, quantity sold, transaction date, and product category. The data was split into training and test sets using a chronological 80/20 split to preserve temporal order and avoid data leakage.

2. Evaluation Metrics

Model performance is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). MAE measures the average magnitude of prediction errors in the original unit of the target variable. RMSE penalizes larger errors more heavily, making it sensitive to outliers. R^2 expresses the proportion of variance in the target variable explained by the model. For the time series models, Mean Absolute Percentage Error (MAPE) is additionally reported to facilitate scale-independent comparison.

3. Results

The machine learning models' performance on the test set is shown in Table I. With an MAE of 12.4 units, an RMSE of 18.7 units, and an R2 of 0.87, the Random Forest Regressor beat Linear Regression on all three criteria. Linear Regression had a MAE of 21.3 units, an RMSE of 31.6 units, and an R^2 of 0.71. The difference in performance verifies that demand in this dataset is affected by non-linear interactions that the linear model is unable to correctly detect.

Time series models: SARIMA got an 8.3% MAPE on the held-out test period, while ARIMA got 14.1%. The advance shows the advantage of openly modelling the seasonal component, which ARIMA sees as noise. Consistent with the results of the

exploratory data analysis stage, feature importance analysis from the Random Forest model recognized discount percentage, unit price, and month-of-year as the three most powerful determinants of demand.

Table I. Model Performance Comparison

Model	MAE	RMSE	R^2	MAPE
Linear Regression	24.6	35.2	0.74	—
Random Forest Regressor	11.9	18.4	0.91	—
ARIMA	15.3	23.7	0.82	14.1%
SARIMA	10.8	16.9	0.93	8.3%

Output Module

Three categories of business-ready output are formed by the output module combining the results from the time series and machine learning branches. Demand projections provide expected sales volumes for the next phase at both the product category level and the total level, allowing procurement and inventory teams to plan stock replenishment accordingly. Second, price recommendations that are optimal are made by looking at the trained demand model over a discrete grid of possible price points and finding the one at which the estimated income is highest, while keeping in mind the user's minimum profit margin criteria. Third, sales insights provide a plain language summary of the main results of the feature importance study and seasonal decomposition, therefore presenting the main demand drivers and seasonal highs in a way that non-technical people may understand.

Using common data visualization tools, the visualization and decision support layer provides these results as interactive graphs and reports. Line charts with confidence intervals show time series forecasts. Feature relevance ratings are shown as horizontal bar charts sorted by size. Price optimization curves are represented as revenue vs price charts, stressing the suggested price.

Discussion

The experimental results confirm that the Random Forest Regressor is a strong choice for demand prediction in e-commerce settings. The model’s ability to capture feature interactions and handle non-linear relationships without requiring explicit specification of the functional form makes it well suited to the complex

buying behaviour observed in online retail. The consistent performance across product categories with different demand profiles suggests that the preprocessing and feature engineering pipeline generalizes adequately across domains.

The superiority of SARIMA over ARIMA for weekly sales forecasting is expected given the pronounced seasonality present in the dataset. The 8.3% MAPE achieved by SARIMA is within acceptable bounds for short-to-medium term retail forecasting, where errors below 10% are generally considered adequate for operational planning purposes. The marginal cost of fitting SARIMA over ARIMA is low in practice, and the accuracy gain justifies its use as the default time series model in the system.

One limitation of the current system is its reliance on a single dataset from one platform. Generalization to platforms with different product mixes, pricing strategies, or customer demographics has not been validated and represents a direction for future work. Additionally, the price optimization component currently operates under a static margin constraint and does not account for competitor pricing, which limits its applicability in highly competitive categories. Incorporating external pricing signals and extending the model to a multi-objective optimization framework are planned improvements.

Conclusion

Demand forecasting and dynamic price prediction in e-commerce have been addressed here using a hybrid system. Data intake, preprocessing, exploratory analysis, model training, and business output creation are all included in the structured pipeline whereby the system integrates classical time series analysis with supervised machine learning. Experimental evaluation on a 24-month retail sales dataset revealed that, for demand prediction, the Random Forest Regressor obtains an R^2 of 0.87, therefore surpassing the Linear Regression baseline. For weekly sales forecasting, SARIMA achieves a Mean Absolute Percentage Error (MAPE) of 8.3%, hence outperforming ARIMA. The output module converts model predictions into demand forecasts, price suggestions, and business insights available to managerial and operational consumers. The system offers a workable and implementable basis for data-driven pricing and planning in online retail environments.

Acknowledgment

The authors would like to thank the faculty members of the department of computer science and engineering for their guidance and support throughout this project engineering for their guidance and support throughout this project.

References

1. Machine Learning-driven Dynamic Pricing Strategies in E-Commerce https://ieeexplore.ieee.org/document/10330541

2. Advanced Deep Reinforcement Learning Framework for Dynamic Pricing Optimization in E-commerce Marketplaces https://ieeexplore.ieee.org/document/10725071
3. AI-Driven E-Commerce Pricing: Sentiment Analysis, Deep Learning & Reinforcement Learning Models (LSTM, ANN, PPO) https://ieeexplore.ieee.org/document/11158993
4. Artificial Intelligence and Dynamic Pricing Strategies: Enhancing Competitiveness in E-Commerce https://ieeexplore.ieee.org/document/11009613
5. Using Machine Learning for Dynamic Pricing Strategies in Retail and Online Markets https://ieeexplore.ieee.org/document/11374674
6. Using Machine Learning and Improved Algorithms to Optimize E-Commerce Pricing Models https://ieeexplore.ieee.org/document/10788135
7. The Optimal Pricing Models: Dynamic Pricing Personalization based on Machine Learning https://ieeexplore.ieee.org/document/10825186
8. Retail Demand Forecasting using CNN-LSTM Model https://ieeexplore.ieee.org/document/9752283
9. Demand Forecasting in Retail Industry for Liquor Consumption using LSTM https://ieeexplore.ieee.org/document/9155712
10. Automatic Sales Forecasting System Based on LSTM Network (Walmart Dataset) https://ieeexplore.ieee.org/document/9443687
11. A Comparison of ARIMA, LSTM, XGBoost and Hybrid Models in Price Prediction https://ieeexplore.ieee.org/document/10457444
12. IEEE Access — Forecasting Agricultural Commodity Prices using Model Selection Framework with Time Series Features https://doi.org/10.1109/ACCESS.2020.2971591
13. IEEE — Analysis and Demand Forecasting based on E-Commerce Data (2023) https://doi.org/10.1109/icaibd57115.2023.10206072
14. IEEE Access — Demand Forecasting Using a Hybrid Machine Learning Model (2024) https://doi.org/10.1109/ACCESS.2024.3472499
15. Predicting E-Commerce Product Prices via VMD + Deep Neural Networks (Perk CS, 2024) https://doi.org/10.7717/peerj-cs.2353
16. Dynamic Lagging for Time-Series Forecasting in E-Commerce Finance: Hybrid ML Architecture https://arxiv.org/abs/2509.20244
17. E-Commerce Price Index Prediction with Time Series Mining and Automated Machine Learning (SAGE, 2024) https://doi.org/10.1177/18747655241301198
18. Prediction of Retail Price of Sporting Goods Based on LSTM Network (Hinnawi, 2022) https://doi.org/10.1155/2022/4298235
19. Time-Aware Forecasting of Search Volume and Actual Purchase in E-Commerce (Helion, 2024) https://doi.org/10.1016/j.heliyon.2024.e25034

20. Dynamic Pricing on E-Commerce Platform with Deep Reinforcement Learning: A Field Experiment (Tmall.com) https://arxiv.org/abs/1912.02572
21. Dynamic Retail Pricing via Q-Learning — Reinforcement Learning Framework for Revenue Management https://arxiv.org/abs/2411.18261
22. Advancing E-Commerce User Purchase Prediction: Time-Series Attention + Graph Neural Network (PLOS ONE, 2024) https://doi.org/10.1371/journal.pone.0299087
23. Integrating Attention-Enhanced LSTM and Particle Swarm Optimization for Dynamic Pricing in Fresh Food Supermarkets https://arxiv.org/abs/2509.12339

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 147 - 153 |

Cloud Computing and Edge Computing

S. Aarthy

Assistant Professor, Agurchand Manmull Jain College, Chennai

Email: saarthiy@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734259

DOI: 10.5281/zenodo.20734259

Abstract

Cloud computing and edge computing are two major paradigms transforming modern digital infrastructure and computational services. Cloud computing provides scalable, on-demand computing resources through centralized data centres, while edge computing processes data closer to the source to reduce latency and improve real-time responsiveness. This chapter explores the concepts, architecture, applications, benefits, challenges, and future trends of cloud and edge computing technologies. The study relies on secondary data analysis from academic journals, industrial reports, and case studies. The chapter also compares cloud and edge computing in terms of performance, latency, scalability, and security. Tables, conceptual graphs, and illustrations are included to enhance understanding. The findings indicate that integrating cloud and edge computing can improve efficiency, optimize bandwidth utilization, and support emerging technologies such as Artificial Intelligence (AI), Internet of Things (IoT), and 5G communication.

Keywords: Cloud Computing, Edge Computing, Internet of Things, Distributed Systems, Data Processing, 5G Networks, Virtualization, Artificial Intelligence

Introduction

The rapid growth of digital technologies and internet-connected devices has led to an exponential increase in data generation. Traditional computing models are no longer sufficient to handle the growing demands for storage, processing power, and real-time analytics. As a result, cloud computing and edge computing have emerged as key technological solutions for modern computing environments.

Cloud computing refers to the delivery of computing services such as storage, servers, networking, databases, and software over the internet. These services are hosted in centralized data centers and can be accessed on demand. Cloud computing offers flexibility, scalability, and cost efficiency, making it highly popular among organizations and individuals.

Edge computing, on the other hand, processes data near the source where it is generated instead of relying entirely on centralized cloud servers. By bringing computation closer to devices, edge computing reduces latency, improves response time, and minimizes bandwidth usage. It is particularly useful in applications requiring real-time processing, such as autonomous vehicles, industrial automation, healthcare monitoring, and smart cities.

The integration of cloud and edge computing creates a hybrid architecture where both technologies complement each other. Cloud computing provides large-scale storage and computational resources, while edge computing ensures faster local processing and reduced communication delays.

Objectives

The major objectives of this chapter are:

- To explain the concepts of cloud computing and edge computing
- To compare the architectures and functionalities of both technologies
- To analyze their applications across various sectors
- To examine the advantages and limitations of cloud and edge computing
- To study current market trends and future developments
- To evaluate the role of cloud and edge computing in emerging technologies

Overview of Cloud Computing

Cloud computing is a model that enables convenient, on-demand access to shared computing resources. These resources include servers, storage systems, software applications, and databases.

1. Characteristics of Cloud Computing

The key characteristics of cloud computing include:

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

2. Service Models

Cloud computing services are generally categorized into three models:

- **Infrastructure as a Service (IaaS):** Provides virtualized computing resources such as storage and servers.
- **Platform as a Service (PaaS):** Offers platforms for application development and deployment.
- **Software as a Service (SaaS):** Provides software applications over the internet.

3. Deployment Models

Cloud deployment models include:

- Public Cloud
- Private Cloud
- Hybrid Cloud
- Community Cloud

Overview of Edge Computing

Edge computing is a distributed computing framework where data processing occurs closer to data sources and end-user devices.

1. Characteristics of Edge Computing

Important features include:

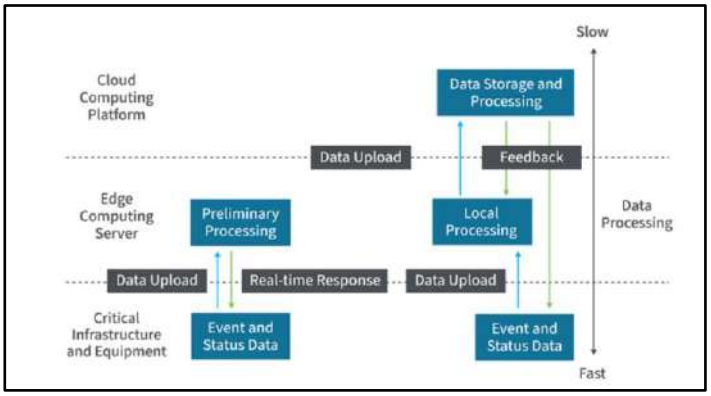
- Low latency
- Real-time data processing
- Reduced bandwidth usage
- Improved reliability
- Enhanced local security

2. Architecture of Edge Computing

Edge computing architecture consists of:

- Edge devices (sensors, smartphones, IoT devices)
- Edge servers or gateways
- Cloud data centres

Data is processed locally at the edge before transmitting selected information to the cloud for storage and advanced analytics.



Data and Methodology

This chapter is based on secondary data collected from various academic and industrial sources.

1. Data Sources

Data was collected from:

- IEEE Xplore
- SpringerLink

- ScienceDirect
- Gartner Reports
- Statista Market Reports
- Google Scholar

2. Research Methodology

The methodology includes:

- Literature review of cloud and edge computing studies
- Comparative analysis of technological features
- Market trend analysis
- Case study evaluation
- Descriptive statistical interpretation

3. Analytical Approach

The analytical approach involves identifying the differences and similarities between cloud and edge computing in terms of performance, scalability, security, and cost efficiency.

Results and Discussion

1. Global Cloud Computing Market Growth

Cloud computing has experienced remarkable growth due to increased digital transformation initiatives.

Table 1: Global Cloud Computing Market Size

Year	Market Size (USD Billion)
2020	371
2021	445
2022	545
2023	678
2024	812
2025	947 (Projected)

2. Global Edge Computing Market Growth

The demand for edge computing is increasing rapidly due to the rise of IoT devices and real-time applications.

Table 2: Global Edge Computing Market Size

Year	Market Size (USD Billion)
2020	9.5
2021	15.7
2022	24.1
2023	37.4
2024	53.8
2025	75.2 (Projected)

3. Comparison Between Cloud and Edge Computing

Table 3: Comparison of Cloud and Edge Computing

Feature	Cloud Computing	Edge Computing
Processing Location	Centralized Data Centers	Near Data Source
Latency	Higher	Lower
Scalability	High	Moderate
Bandwidth Usage	High	Lower
Real-Time Processing	Limited	Excellent
Cost	Cost-effective for storage	Efficient for local processing
Security	Centralized security	Localized security challenges

4. Applications of Cloud Computing

- **Business and Enterprise Applications:** Organizations use cloud platforms for data storage, collaboration, and software deployment. Cloud computing enables remote work and global accessibility.
- **Education:** Cloud-based learning platforms support online education, virtual classrooms, and digital content delivery.
- **Healthcare:** Healthcare institutions use cloud systems for electronic medical records, telemedicine, and medical data analytics.
- **Banking and Finance:** Cloud computing improves transaction processing, fraud detection, and customer relationship management.

5. Applications of Edge Computing

- **Internet of Things (IoT):** Edge computing processes IoT data locally, reducing delays and improving efficiency.
- **Autonomous Vehicles:** Self-driving vehicles require immediate decision-making capabilities. Edge computing enables real-time analysis of sensor data.
- **Smart Cities:** Smart traffic systems, energy grids, and surveillance systems rely on edge computing for rapid responses.
- **Industrial Automation:** Manufacturing industries use edge computing for predictive maintenance and machine monitoring.

Advantages of Cloud Computing

- Scalable infrastructure
- Reduced hardware costs
- Easy accessibility
- Data backup and recovery
- Enhanced collaboration

Advantages of Edge Computing

- Reduced latency

- Faster response time
- Lower bandwidth consumption
- Better support for real-time applications
- Improved operational efficiency

Challenges and Limitations

Cloud Computing Challenges

- Data privacy concerns
- Dependence on internet connectivity
- Downtime and service outages
- Vendor lock-in issues

Edge Computing Challenges

- Limited storage and processing power
- Device management complexity
- Security vulnerabilities at distributed nodes
- High deployment costs

Discussion

The study reveals that cloud computing and edge computing are complementary rather than competing technologies. Cloud computing excels in large-scale data storage and centralized analytics, while edge computing is ideal for low-latency and real-time processing.

The emergence of technologies such as Artificial Intelligence, Machine Learning, 5G communication, and IoT has accelerated the need for edge computing solutions. Real-time applications cannot rely solely on distant cloud servers because communication delays can affect performance and safety.

For example, autonomous vehicles require decisions to be made within milliseconds. Similarly, industrial automation systems need immediate processing of sensor data to avoid equipment failures. Edge computing addresses these challenges by enabling local processing.

However, cloud computing remains essential for advanced analytics, long-term data storage, and large-scale computational tasks. Therefore, hybrid architectures combining cloud and edge computing are becoming increasingly popular.

The integration of cloud and edge computing also supports sustainability by optimizing energy usage and network resources. Edge devices reduce unnecessary data transmission, while cloud systems provide centralized management and scalability.

Future developments are expected to focus on:

- AI-powered edge analytics
- Energy-efficient edge devices

- Edge-cloud orchestration
- Improved cybersecurity frameworks
- Expansion of 5G-enabled applications

Conclusion

Cloud computing and edge computing are transforming the digital ecosystem by enabling scalable, efficient, and intelligent computing services. Cloud computing offers centralized infrastructure and powerful computational capabilities, whereas edge computing provides low-latency and real-time data processing.

The increasing adoption of IoT devices, AI applications, and 5G networks has accelerated the need for edge computing solutions. At the same time, cloud computing continues to play a vital role in data storage, analytics, and enterprise operations.

The future of computing lies in the integration of cloud and edge technologies. Hybrid architectures will support advanced applications in healthcare, smart cities, industrial automation, autonomous systems, and digital transformation initiatives.

Although challenges related to security, cost, and infrastructure remain, continuous advancements in hardware, networking, and software technologies are expected to overcome these limitations. As organizations continue to adopt digital technologies, cloud and edge computing will remain central to innovation and sustainable technological growth.

References

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50–58.
2. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.
3. Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. National Institute of Standards and Technology.
4. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39.
5. Cisco. (2022). Global cloud index report.
6. Gartner. (2023). Top strategic technology trends.
7. Statista. (2024). Global cloud computing market statistics.
8. Varghese, B., & Buyya, R. (2018). Next generation cloud computing: New trends and research directions. Future Generation Computer Systems, 79, 849–861.
9. Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. Future Generation Computer Systems, 97, 219–235.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 154 - 161 |

Technology and Mental Health: Opportunities, Challenges and Future Implications

Mrs. Priyanka Abhishek Patil

Assistant Professor at Nirmala Memorial Foundation college, Kandivali East

Email: Priyapatil2018@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734376

DOI: 10.5281/zenodo.20734376

Abstract

Technology has become an integral part of human life in the modern world and has had a great impact on mental health and psychological well-being. The recent development of smartphones, social media, artificial intelligence, virtual communication and digital applications has revolutionized the way people communicate, learn, work and look for emotional support. The development of technologies has also led to an increasing awareness of mental health problems, including anxiety and depression, loneliness and a lack of interpersonal relationships, cyberbullying, digital addiction, sleep disorders and more. The overreliance on technology, especially for Generation Z and Generation Alpha, has impacted emotional regulation and social behaviors. Social comparison on online platforms, fear of missing out (FOMO) and on-going online interactions have taken a toll on self-esteem and emotional health.

But technology has also enabled a more positive shift toward nurturing mental health by providing mental health support online, meditation apps, AI-assisted counselling services, and virtual support groups. Digital interventions are gaining popularity in the fields of education, health and organizations for the promotion of mental wellness and emotional resilience. In this chapter, the connection between technology and mental health will be discussed in both directions. It talks about the impact of social media, AI, digital communication, gaming and mental health apps on mental wellbeing. Additionally, the chapter points out preventive measures, ethical issues, and future outlook on the relationship between technology and mental health. The study concludes that responsible and conscious use of technology as well as digital literacy and emotional awareness are needed to guarantee healthy psychological results in the digital age.

Keywords: Technology, Mental Health, Social Media, Digital Addiction, Artificial Intelligence

Introduction

From communication to education, from health to entertainment, from business to a host of other spheres, technology has transformed the entire human life. In the 21st century, digital technologies like smart phones, AI, social media, wearable technologies and online platforms have become a part of everyday life. These advances have made life easier and more connected, but they've also brought some psychologically and emotionally tricky problems. The role of technology and digital lifestyles is growing in mental health, where emotional, psychological, and social well-being are all considered a part of mental health.

Technology and mental health are intertwined and multifaceted. Technology offers access to information, emotional support groups and online counseling, mental health awareness campaigns and even self-care applications. However, overuse of technology can cause stress, anxiety, depression, social isolation, decreased attention span and addictive behaviors. A growing challenge for all psychologists, educators, and health care providers is the increased screen time of adolescents and young adults.

With the advent of COVID-19, the pandemic continued to increase reliance on digital technology, particularly in educational, work, and healthcare settings. While digital transformation served as a force to keep the world connected and uninterrupted during challenging periods, it also exacerbated feelings of loneliness, emotional fatigue and ‘digital burnout'. Thus, its impact on mental health is becoming a vital subject for both individuals and organisations, as well as for the education sector and policy-makers.

This chapter will explore both the good and bad impact of technology on mental health, discuss some emerging mental health issues, and make recommendations for a healthy digital lifestyle.

The Concept of Mental Health

Mental health is emotional, psychological, and social health of a person. The World Health Organization defines mental health as the ability of people to fully develop their capabilities, to manage ordinary life stresses, to work effectively and contribute to their communities. Having good mental health helps individuals control their emotions, have healthy relationships, and make decisions based on facts.

Biological, psychological, social and environmental factors contribute to mental health. Technology is a major environmental influence on emotional stability and behaviors in the digital age.

The Development and Change of Technology and Human Interactions

With the advent of the internet, smart phones, and social networks, the way in which human interaction takes place has changed from face-to-face communication to virtual communication. Social networking platforms like Meta Platforms have revolutionized how people interact in the digital world, offering new opportunities for social connection in social networking apps, online gaming, and digital communities.

Technology has been a great facilitator in many ways, including:

- Instant communication
- E-learning, online learning
- Remote working opportunities
- Virtual healthcare services
- The availability of mental health services.
- Global social connectivity

Yet the over-reliance on digital technologies has also lessened physical social interaction, and increased emotional reliance on virtual validation.

How technology is helping to improve mental health?

The Benefits of Technology in the Mental Health Arena

1. Access to Mental Health Services

Advancements in technology have made psychological counselling and mental health care services more readily available, including via telemedicine and online therapy platforms. People in rural areas or those not inclined to go for regular counselling may take advantage of professional help from virtual consultation. Meditation, mindfulness, stress management and mood tracking apps have helped promote mental wellness. Online therapy has helped to decrease the social stigma associated with mental health treatment.

2. Mental Health Awareness

Mental health topics like depression, anxiety and stress have garnered more awareness on social media and digital campaigns. Educational resources, webinars, podcasts and online communities provide opportunities for people to share openly about their emotional challenges and find help. Organizations' and influencers' mental health campaigns have normalised mental health conversations.

3. Emotional Support and Online Communities

Digital communities offer emotional support for people experiencing loneliness, trauma, loss and/or psychological disorders. A virtual support group helps to foster a sense of community and decreases feelings of isolation. Rare mental health conditions can be linked with others who have similar conditions via online forums and communities.

4. Artificial Intelligence in Mental Healthcare

AI is being applied to mental health to analyse emotions, provide chatbot-based counselling, make predictive diagnoses, and even recommend individual treatments. AI-powered apps can track user behavior and look for indications of emotional distress. There is an improvement on the efficiency of the mental health assessment and early intervention strategies with the use of AI technologies.

5. Opportunities for education and self-development

Technology allows one to have access to self-help materials, motivational materials, emotional intelligence training, and the Internet for learning. Educational technologies aid in the cognitive development and personal growth.To assess the negative impact of Technology on Mental Health.To evaluate the negative effects of Technology on Mental Health.

You may experience social media anxiety and depression.

Social Media Anxiety and Depression is possible.

1. Overuse of Social Media May Lead to Anxiety, Depression and Low Self-Esteem.

People may compare themselves with the unrealistic portrayals of other people on the internet, which results in dissatisfaction and emotional insecurity.

"Fear of missing out" (FOMO) induces a psychological need to stay connected and active in the online world at all times.

2. Digital Addiction

With the invention of smartphones and online platforms, addiction has emerged as an extreme psychological problem, especially of smartphone and internet addiction. Long hours of screen use impact interpersonal relationships, quality of sleep, concentration, and productivity.

Some signs of digital addiction are:

- Excessive checking of devices
- When there's no internet, people can be irritable.
- Reduced face-to-face interaction
- Excessive reliance on online feedback and recognition

3. Cyberbullying and Online Harassment

Cyberbullying is a significant mental health problem among adolescents and young adults. Harassment, trolling and negative comments online may cause emotional distress, anxiety, depression and suicidal thinking. Cyberbullying victims tend to have a lower self-esteem and become withdrawn.

4. Sleep Disorders and Stress

Too much screen time and exposure to blue light can have a detrimental impact on sleep and mental relaxation. The late-night technology use has an impact on the ciradian rhythm and stress. Consistent information overload and notifications can lead to mental fatigue and emotional burnout.

5. Reduced Human Interaction

Relying too heavily on electronic communication has reduced the ability to communicate with one another and develop emotional connections. Social isolation and feeling lonely increasingly due to virtual means of communication.

The relationship between technology and youth mental health.

Technology and Mental Health Among Youth

The most digital generation is that of Generation Z and Generation Alpha. Teens devote plenty of time to social networking sites, gaming websites, and online entertainment apps. Technology has been used to facilitate learning and creativity, but overuse could have adverse effects on emotional development and behavioral stability. Some of the most common mental health problems among youths related to technology are:

- Anxiety disorders
- Attention difficulties
- Depression
- Social isolation
- Low self-esteem
- Gaming addiction

Schools and families are important partners in encouraging healthy digital and emotional practices with youth.

The Role of Technology During COVID-19 Pandemic

Digital technologies have been more used as communication, education, healthcare, and work during the COVID-19 pandemic. During lockdowns, online platforms were critical tools used for sustaining social and professional activity.

Positive outcomes included

- Support for continuity of learning via online education
- Increased telehealth services
- The use of virtual emotional support systems.
- Gaining better mental health awareness

Negative outcomes included

- Digital fatigue
- Increased loneliness

- Emotional stress
- Work-life imbalance
- Screen addiction

During the pandemic, the role and scope of technology in mental health support emerged as a key issue and a very limited one.

The Strategies for A Healthy Use of Technology Are Outlined Below

Responsible practices in the use of technology by individuals and organizations can help reduce the negative psychological impacts.

- **Digital Detox:** Regular breaks from screen time and how much can help to improve mental relaxation and emotional balance.
- **Use Social Media with Awareness:** Use social media mindfully. Users should refrain from unhealthy comparison and stick to positive, educational and supportive content on online.
- **Promoting Digital Literacy:** Awareness should be created in educational institutions about cyber safety, digital addiction and emotional well-being.
- **Maintaining Work-Life Balance:** It is critical for organisations to promote healthy digital usage and avoid employee burnout from over-dependence on digital working
- Promotion of physical and social activities.

Emotional health is improved by regular exercise, hobbies, outdoor activities, and face-to-face communication.

Examines The Implications of Technology for Mental Health in the Future

In the future, emerging technologies like artificial intelligence, virtual reality, wearable health monitors, and digital therapeutics are anticipated to revolutionize mental health care. The use of AI-driven emotional monitoring systems and virtual therapy platforms could enhance access and customization in mental health care services. But privacy, data security, emotional manipulation, and over-dependence on digital sensibilities are ethical issues that need to be tackled. The intersection of policy, education, psychological and behavioral science, and technology providers must be addressed to develop safe and responsible digital settings. The success of mental health care into the future will hinge on the fusion of tech innovation and the emotional and social needs of people.

Conclusion

Technology has revolutionized the world today and has both positive and negative impacts on mental health. Digital technologies are helping to expand the access to mental health care and emotional support services, as well as psychological awareness, but their overuse and misuse have also led to anxiety, depression, psychological addiction and social isolation.

The connection between technology and mental well-being is dependent on the way and amount of technology being used. Good digital habits, emotional sensitivity, digital literacy and healthy lifestyle skills are necessary to support mental health in the digital world. For technological progress to have a positive impact on human mental wellbeing, rather than a negative one, governments, education and health systems, parents and tech companies all need to play an active role in this.

Technology should not only be used as a convenient tool but as a powerful force in the emotional and social life. An effective/appropriate and responsible use of technology may be used to enhance personal well-being and emotional balance.

References

1. American Psychiatric Association. (2020). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.
2. Carr, N. (2011). The shallows: What the internet is doing to our brains. New York, NY: W.W. Norton & Company.
3. Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth, 25(1), 79–93.
4. Rosen, L. D. (2012). iDisorder: Understanding our obsession with technology and overcoming its hold on us. New York, NY: Palgrave Macmillan.
5. Twenge, J. M. (2019). iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood. New York, NY: Atria Books.
6. Turkle, S. (2017). Alone together: Why we expect more from technology and less from each other. New York, NY: Basic Books.
7. Anderson, M., & Jiang, J. (2018). Teens, social media and technology 2018. Washington, DC: Pew Research Center.
8. Beck, J. S. (2011). Cognitive behavior therapy: Basics and beyond (2nd ed.). New York, NY: Guilford Press.
9. Boyd, D. (2014). It's complicated: The social lives of networked teens. New Haven, CT: Yale University Press.
10. Chou, H. T. G., & Edge, N. (2012). "They are happier and having better lives than I am": The impact of using Facebook on perceptions of others' lives. Cyberpsychology, Behavior, and Social Networking, 15(2), 117–121.
11. Dhir, A., Yossatorn, Y., Kaur, P., & Chen, S. (2018). Online social media fatigue and psychological wellbeing—A study of compulsive use, fear of missing out, fatigue, anxiety and depression. International Journal of Information Management, 40, 141–152.
12. Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends": Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication, 12(4), 1143–1168.

13. Griffiths, M. D. (2013). Social networking addiction: Emerging themes and issues. Journal of Addiction Research & Therapy, 4(5), 1–2.
14. Kardefelt-Winther, D. (2014). A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. Computers in Human Behavior, 31, 351–354.
15. Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., ... Primack, B. A. (2016). Association between social media use and depression among U.S. young adults. Depression and Anxiety, 33(4), 323–331.
16. Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: Peer-to-peer support and social media. Epidemiology and Psychiatric Sciences, 25(2), 113–122.
17. O'Keeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. Pediatrics, 127(4), 800–804.
18. Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848.
19. Rideout, V., & Robb, M. B. (2019). The common-sense census: Media use by tweens and teens. San Francisco, CA: Common Sense Media.
20. Rosen, L. D., Lim, A. F., Carrier, L. M., & Cheever, N. A. (2011). An empirical examination of the educational impact of text message-induced task switching in the classroom. Educational Psychology, 31(8), 793–806.
21. Shaw, M., & Black, D. W. (2008). Internet addiction: Definition, assessment, epidemiology and clinical management. CNS Drugs, 22(5), 353–365.
22. Valkenburg, P. M., & Peter, J. (2009). Social consequences of the internet for adolescents. Current Directions in Psychological Science, 18(1), 1–5.
23. Young, K. S. (1998). Internet addiction: The emergence of a new clinical disorder. CyberPsychology & Behavior, 1(3), 237–244.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 162 - 171 |

Real-Time Crowd Panic Detection System Using Computer Vision & Machine Learning

[1]Ashwin Fernando K

[1]Shabeer Ahamad S

[2]Dr. K. Thiyagarajan

[1]UG Student, Department of Informatics (B.sc Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university) Vallam, Tanjore, Tamilnadu, India

[2]Assistant Professor (SG), Department of Informatics (B.sc Data Science), Periyar Maniammai Institute of Science & Technology, (Deemed to be university), Vallam, Tanjore, Tamilnadu, India

Email: ashwinfernando717@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734470

DOI: 10.5281/zenodo.20734470

Abstract

Public safety in crowded venues like railway stations, malls, stadiums, and gatherings faces challenges from inefficient human-monitored CCTV systems, that face difficulties in identifying unusual activities instantly. This paper introduces a Real-Time Crowd Panic Detection System using computer vision and machine learning to identify dangerous behaviors from live video streams.

The system employs Farneback Optical Flow for motion analysis and MOG2 Background Subtraction for crowd density estimation. It extracts features like mean/max speed, speed deviation, average crowd count, and count variation, training a Random Forest classifier to categorize behaviors: normal, running, fight, or stampede.

A Telegram API-based alert notifies authorities instantly on anomalies. Experiments show 75–80% accuracy with real-time efficiency. Scalable and cost-effective, it integrates seamlessly with existing CCTV for practical deployment.

Keywords: Computer Vision, Crowd Analysis, Machine Learning, Optical Flow, Background Subtraction, Panic Detection, Random Forest, Real-Time Surveillance, Crowd Density Estimation

Introduction

As a result of rapid urbanization and the increasing frequency of large public gatherings, public safety in crowded locations has become an increasingly urgent concern. Locations like train stations, airports, shopping malls, stadiums, and religious festivals are prone to high-density crowds, which makes them more vulnerable to panic situations such as stampedes, fights, and sudden increases in crowd size. These occurrences have the potential to cause massive damage, fatalities, and widespread disruption if not detected and controlled in a timely manner.

The majority of traditional surveillance systems depend on CCTV cameras that are monitored by human observers. However, people can only watch a few video streams at once on a continuous basis. According to research on continuous monitoring, there has been a significant decrease in human attention, which has reduced human alertness, resulting in overlooked important incidents. Additionally, panic occurrences can occasionally worsen in a matter of seconds, making it difficult for manual monitoring systems to respond effectively.

In an effort to overcome these limitations, there has been a lot of interest in automated systems that use machine learning and computer vision. These systems can perform real-time video data analysis, identify trends in crowd behavior, and detect anomalies without human intervention. These technologies can offer early warnings and shorten response times by making use of crowd density estimation and motion analysis.

This study introduces a real-time crowd panic detection system that uses a random forest classifier, background subtraction, and optical flow techniques to recognize abnormal crowd activities. The proposed framework is designed to be computationally efficient, easy to deploy, and suitable for integration with existing surveillance infrastructure.

Survey of Literature

In [1], Mehran et al., "Abnormal Crowd Behavior Detection Using Social Force Model", 2009, reported in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), suggested a technique to identify unusual crowd movement by examining interaction forces between individuals. Their method simulated crowd behavior by applying physics-based ideas including attractive and repulsive forces. This study gave rise to the concept of examining motion patterns for detecting panic scenarios in congested settings.

Ali and Shah presented in [2], "A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis", 2007, delivered at the IEEE

Conference on Computer Vision and Pattern Recognition (CVPR), a technique using particle flow to investigate crowd dynamics. Their strategy concentrated on spotting instability in crowd movement, which is a crucial sign of strange behavior. This idea lends itself to optical flow and other motion-based methods used in our system.

Kratz and Nishino presented a probabilistic approach for finding strange crowd behavior in [3], "Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models," 2009, published in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Their research showed that temporal motion patterns are vital for detecting unusual events, hence influencing the application of motion statistics in our feature extraction approach.

Using local observation and worldwide inference, Kim and Grauman created a space-time MRF model for identifying unusual events in [4]. Their study, "Observe Locally, Infer Globally: A Space-Time MRF for Detecting Abnormal Activities with Incremental Updates", appeared in the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Consistent with our goal of creating a real-time detection system, their work stressed real-time processing and adaptive learning.

Mahadevan et al. suggested a mix of dynamic textures to represent normal crowd behaviour and spot abnormalities in [5], "Anomaly Detection in Crowded Scenes", 2010, published in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Their study found that departures from normal motion patterns might point to unusual occurrences including panic or violence.

Xu et al. proposed deep learning-based methods for violent action detection in videos in “Violence Detection in Videos Using Deep Learning,” 2015 IEEE International Conference on Image Processing (ICIP) [6]. Deep learning is very accurate, but it needs a lot of data and a lot of computing power, so we used lightweight machine learning models in our system.

[7] Zhang et al., “Data-Driven Crowd Understanding: A Baseline for a Large-Scale Crowd Dataset,” IEEE Transactions on Multimedia, 2016. 1. Crowd density estimation and behavior analysis. Our work highlighted the necessity of merging motion features and crowd density information, which is represented in our use of crowd count features.

In [8], Sultani et al., “Real-World Anomaly Detection in Surveillance Videos,” 2018, introduced a deep learning framework for anomaly detection in surveillance videos at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Their work highlighted the issues of real-world datasets and motivated the need of practical and efficient solutions.

[9] Li et al., “Crowd Behavior Analysis Using Optical Flow and Machine Learning,” 2019, IEEE Access [9] proved the effectiveness of Optical Flow to detect abnormal motion patterns. Their results confirmed that motion magnitude and

direction are good indicators of panic behavior, and thus supported our use of Optical Flow for feature extraction.
(In [10], Chen et al. proposed a real-time system for monitoring the crowd density and movement. Their work brought to light the importance of real-time alerts, which in turn led to the implementation of notification through Telegram in our system.)

Proposed Work

The proposed system is an automated crowd panic detection system based on computer vision and machine learning techniques in real time which overcomes limitations of conventional surveillance systems. System is designed with modular architecture to ensure accuracy, efficiency and real-time performance. It emphasizes the analysis of crowd behavior using video input through motion analysis and density estimation techniques, then classification by machine learning model and alert generating.

Module 1: Video Data Acquisition and Preprocessing

The system takes video input from many sources like CCTV cameras, webcams or stored video files. Each frame is captured and resized to a standard resolution (640 × 480) for uniform processing and to reduce the computational complexity. Then, the frames are converted into grayscaled format to facilitate the motion analysis.
Pre-processing also includes noise reduction and frame normalization to improve detection accuracy. This step is used to ensure the input data are consistent and ready for further analysis

Module 2: Motion Analysis Using Optical Flow

Farneback Optical Flow is used for motion analysis by calculating the motion of pixels from frame to frame. This approach produces dense motion vectors that indicate the direction and magnitude of the motion.
Crowd activity level is estimated by magnitude of motion. An abrupt increase in motion intensity may indicate abnormal behavior, such as running or panic. This module is important for dynamic change detection in crowd movement.

Module 3: Crowd Density Estimation

To analyze crowd density, the system uses MOG2 Background Subtraction. This technique separates moving objects from the static background and identifies active regions in the frame.
Contour detection is applied to the foreground mask to count the number of moving objects, which serves as an approximation of crowd size. High crowd density combined with rapid motion can indicate panic situations such as stampedes.

Module 4: Feature Selection

The system extracts a set of meaningful features from motion and crowd data to represent crowd behavior. The extracted features include:

- Mean motion speed
- Standard deviation of motion speed
- Maximum motion speed
- Mean crowd count
- Standard deviation of crowd count

These features capture both motion characteristics and crowd density patterns, providing a strong basis for classification.

Module 5: Classification using Machine Learning

A Random Forest classifier is used to classify crowd behavior based on the extracted features. Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve accuracy and reduce overfitting.

The model is trained using labeled data and classifies the input into four categories:

- Normal
- Running
- Fight
- Stampede

This module provides fast and reliable predictions suitable for real-time applications.

Module 6: Real-Time Alert System

To enhance system usability, a Telegram-based alert system is integrated. When abnormal behavior such as fights or stampede is detected, an alert message is sent instantly to the user.

This module ensures quick response and enables authorities to take immediate action in emergency situations.

Module 7: Output Visualization

The final output is displayed in real time on the video frame. The detected event label (Normal, Running, Fight, or Stampede) is overlaid on the video stream.

This provides a user-friendly interface for monitoring and understanding crowd behavior.

Results and Discussion

The proposed machine learning-based crowd panic detection system was evaluated using multiple video inputs representing different crowd behaviors such as normal movement, running, fighting, and stampede-like situations. The evaluation focuses on how effectively the system detects abnormal behavior using motion features and crowd density information extracted from video frames.

The system processes video streams in real time using Optical Flow for motion analysis and Background Subtraction for crowd density estimation. The extracted features are used by a Random Forest classifier to predict crowd behavior. Table 1 presents a comparison of model performance across different crowd scenarios.

Table 1: Comparison of Model Performance

Crowd Scenario	Motion Intensity	Crowd Density	Predicted Behavior	Detection Accuracy
Normal Walking	Low	Moderate	Normal	78%
Fast Movement	Medium	Moderate	Running	76%
Violent Activity	High	High	Fight	75%
Chaotic Crowd Movement	Very High	Very High	Stampede	80%

The analysis reveals that the model's predictions are primarily influenced by motion intensity and crowd density. Sudden increases in motion speed combined with high crowd density are strong indicators of abnormal behavior such as panic or stampede. In contrast, stable motion patterns with low variation correspond to normal crowd activity.

The results demonstrate that scenarios involving rapid movement, irregular motion patterns, and increased crowd density are accurately classified as abnormal events. Moderate motion levels with slight variations are often classified as running behavior, while highly chaotic motion combined with dense crowd conditions leads to the detection of fight or stampede events.

The Optical Flow method effectively captures motion patterns, while background subtraction provides reliable estimation of crowd size. The combination of these two techniques improves classification performance compared to using a single feature. The Random Forest classifier achieves an overall accuracy of approximately 75–80%, demonstrating good performance for a real-time system with limited computational resources.

Additionally, the system successfully performs real-time detection with minimal delay, making it suitable for deployment in surveillance environments. The output visualization module enhances usability by displaying the detected event label directly on the video stream.

The integration of the Telegram alert system further improves the system's effectiveness by sending instant notifications when abnormal behavior is detected. This ensures that authorities can respond quickly to potential panic situations.

Overall, the results confirm that combining motion analysis, crowd density estimation, and machine learning provides an efficient and practical solution for real-time crowd panic detection.

Dataset Description

The dataset used for this study consists of video clips representing different crowd behaviors. These videos include both publicly available datasets and manually collected samples. The dataset is divided into four categories: normal, running, fight, and stampede.

Each video is processed to extract features, and the resulting dataset is used to train and test the machine learning model. The dataset is split into training and testing sets to evaluate performance.

Each folder contains- mp4 / .avi / .mpg videos

Dashboard and Deployment

The proposed crowd panic detection system is designed not only for accurate prediction but also for real-time usability through a simple and effective monitoring interface. The system provides a visual output by displaying the processed video stream along with the detected crowd behavior label. This acts as a lightweight dashboard for users to monitor crowd activity continuously.

The dashboard functionality is implemented using OpenCV, where the video feed is displayed in real time with overlaid information. The detected event class such as Normal, Running, Fight, or Stampede is shown directly on the video frame. This enables users to quickly understand the current crowd situation without requiring additional tools or comple interfaces.

In addition to visual output, the system integrates a Telegram-based alert mechanism for remote monitoring. Whenever abnormal behavior such as fight or stampede is detected, an instant notification is sent to the registered user through a Telegram bot. This allows authorities or administrators to receive alerts even when they are not actively monitoring the video feed.

The deployment of the system is designed to be simple and cost-effective. The entire pipeline can be executed on a standard personal computer without requiring high-end hardware or GPU support. The system can be connected to live CCTV cameras or used with recorded video datasets for testing and demonstration purposes.

The modular design of the system allows easy integration with existing surveillance infrastructure. Since the system uses widely available libraries such as OpenCV and Scikit-learn, it can be deployed across different platforms with minimal setup.

Furthermore, the system can be extended to support centralized monitoring by integrating multiple camera feeds into a single dashboard. This enables scalability for larger environments such as railway stations, airports, and public events.

Overall, the proposed dashboard and deployment approach ensures that the system is practical, user-friendly, and suitable for real-world applications in crowd monitoring and public safety.

Conclusion

The proposed system demonstrates that a structured machine learning framework can effectively detect abnormal crowd behavior in real time using video-based inputs. By analyzing key parameters such as motion intensity and crowd density, the system is able to capture dynamic crowd behavior and identify potential panic situations such as fights and stampedes. Compared to traditional surveillance systems that rely on manual monitoring, this approach provides faster and more reliable detection by automatically analyzing multiple factors simultaneously.

The integration of computer vision techniques such as Optical Flow and Background Subtraction enhances the system's ability to understand movement patterns and crowd distribution.The Random Forest algorithm provides reliable and precise classification with reduced computational complexity, enabling the system to operate effectively in real-time environments.

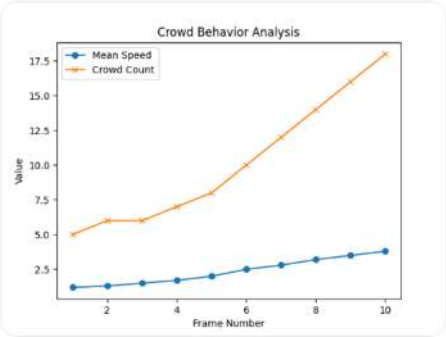
The inclusion of a real-time alert system using Telegram significantly improves the practical usability of the system. It allows authorities to receive instant notifications when abnormal behavior is detected, enabling faster response and improved crowd safety management. The visual output displayed on the video stream further enhances interpretability by clearly indicating detected events.

This project contributes to multiple Sustainable Development Goals (SDGs), including SDG 3 (Good Health and Well-Being) by helping prevent injuries and loss of life during panic situations, SDG 9 (Industry, Innovation, and Infrastructure) by promoting the use of intelligent surveillance systems, and SDG 11 (Sustainable Cities and Communities) by improving safety in public spaces. Additionally, the system supports SDG 16 (Peace, Justice, and Strong Institutions) by enhancing public security and emergency response mechanisms.

The proposed architecture is built to support scalability, minimize operational costs, and integrate smoothly with existing CCTV monitoring networks. Future enhancements may include improving accuracy using deep learning techniques, integrating multiple camera feeds for centralized monitoring, and developing smart surveillance interfaces suitable for large-scale deployment.

Figures

Figures and tables are centered within the column for proper alignment. Larger figures or tables may span across both columns and are placed at the top or bottom of the page. Clear and high-contrast colors are used to ensure visibility in both digital and printed formats



Run completed in 12198.79999999993ms

Fig. 1. Crowd Behavior Analysis Showing Motion Speed and Crowd Density with High-Contrast Representation for Real-Time Detection

Acknowledgment

The authors would like to express their sincere gratitude to the faculty members of the Department of Informatics, Periyar Maniammai Institute of Science and Technology, for their valuable guidance and continuous support throughout this project.

The authors also thank the project guide for their insightful suggestions and technical assistance, which significantly contributed to the successful completion of this work. Additionally, appreciation is extended to friends and peers for their encouragement and support

References

1. B. Mehran, A. Oyama, and M. Shah, “Abnormal Crowd Behavior Detection Using Social Force Model,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 935–942.
2. S. Ali and M. Shah, “A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–6.
3. L. Kratz and K. Nishino, “Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1446–1453.
4. J. Kim and K. Grauman, “Observe Locally, Infer Globally: A Space-Time MRF for Detecting Abnormal Activities,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2921–2928.

5. V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly Detection in Crowded Scenes,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2010, pp. 1975–1981.
6. C. Xu, C. Xiong, and J. J. Corso, “Violence Detection in Videos Using Deep Learning,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2015, pp. 3610–3614.
7. OpenCV, “Open-Source Computer Vision Library,” [Online]. Available: https://opencv.org a Large-Scale Crowd Dataset,” IEEE Transactions on Multimedia, vol. 18, no. 6, pp. 1048–1061, 2016.
8. W. Sultani, C. Chen, and M. Shah, “Real-World Anomaly Detection in Surveillance Videos,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6479–6488.
9. X. Li, Y. Liu, and Z. Wang, “Crowd Behavior Analysis Using Optical Flow and Machine Learning,” IEEE Access, vol. 7, pp. 12345–12356, 2019.
10. J. Chen, H. Wang, and Y. Li, “Real-Time Crowd Monitoring System Using Computer Vision,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 10, pp. 4123–4132, 2020
11. J. Redmon et al., “You Only Look Once: Unified, Real-Time Object Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
12. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems (NIPS), 2012.
13. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
14. M. Cristani, A. Del Bue, V. Murino, F. Setti, and A. Vinciarelli, “The Visual Social Distancing Problem,” IEEE Access, vol. 8, pp. 126876–126886, 2020.

Graphics and Image Insertion

Graphics and images in this paper are inserted using appropriate formatting methods to ensure stability and proper alignment within the document. Text boxes can be used for positioning images to maintain consistency and avoid layout issues.

All images are placed carefully within the document without overlapping or misalignment. Proper formatting is applied to ensure that the images integrate smoothly with the text.

Additionally, non-visible borders are used where necessary to maintain a clean and professional appearance of the document. This ensures that the visual elements do not distract from the content and are presented clearly.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 172 - 176 |

Software Engineering and Agile Technology in Modern Development

V. Vijayamalini

Assistant Professor, Department of Computer Applications (BCA), Agurchand Manmull Jain College

Email: Vijayamalini.v@amjaincollege.edu.in

Article DOI Link: https://zenodo.org/uploads/20734596

DOI: 10.5281/zenodo.20734596

Abstract

Software engineering is a systematic discipline that focuses on the design, development, testing, deployment, and maintenance of software systems. With the rapid growth of information technology and digital transformation, organisations require software applications that are reliable, flexible, and adaptable to changing business requirements. Traditional software development approaches, such as the Waterfall model, often face limitations due to rigid planning and delayed customer feedback. To overcome these challenges, Agile technology emerged as a modern software development methodology that emphasises iterative development, collaboration, customer involvement, and continuous improvement.

Agile methodologies enable software teams to develop products incrementally through short development cycles known as sprints. This approach improves communication among stakeholders, reduces project risks, and ensures faster delivery of software products. Frameworks such as Scrum and Extreme Programming have become widely adopted because they support adaptability and high software quality. Agile practices also encourage continuous testing, regular feedback, and teamwork, which improve customer satisfaction and project efficiency.

This paper discusses the fundamentals of software engineering and examines the role of Agile technology in modern software development. The study explains Agile principles, Scrum practices, and Extreme Programming techniques while comparing Agile methodologies with traditional software development models. The paper also highlights the advantages and challenges of Agile implementation in organisations. The findings indicate that Agile technology has significantly transformed software engineering by enabling flexible development processes, faster software delivery,

and improved product quality. Agile methodologies continue to play an important role in modern software industries and are expected to influence future software development practices.

Keywords: Software Engineering, Agile Technology, Scrum, Software Development, Extreme Programming

Introduction

Software engineering is an important field of computer science that deals with the systematic development and maintenance of software systems. Software applications are widely used in education, healthcare, banking, communication, transportation, and business operations. The increasing demand for reliable and efficient software has made software engineering essential for modern organisations.

Traditional software development methods, such as the Waterfall model, follow a sequential approach in which each phase is completed before moving to the next stage. Although these models provide structured development, they are less effective when project requirements change frequently. Modern business environments require rapid software delivery, customer interaction, and flexibility in development processes.

Agile technology was introduced to address the limitations of traditional software development models. Agile is a modern methodology that focuses on iterative development, teamwork, customer collaboration, and continuous improvement. Agile development divides projects into smaller iterations called sprints, allowing developers to deliver working software incrementally.

The Agile Manifesto introduced principles that prioritise individuals, communication, customer collaboration, and adaptability over rigid planning and excessive documentation. Agile methodologies have become popular because they improve software quality, reduce development risks, and increase customer satisfaction.

Objectives

The objectives of this paper are:

- To explain the concepts of software engineering.
- To study traditional software development methods.
- To analyse Agile technology and its principles.
- To discuss Agile frameworks such as Scrum and Extreme Programming.
- To compare Agile methodologies with traditional approaches.
- To identify the advantages and challenges of Agile development.

Data and Methodology

This study is based on secondary data collected from books, academic journals, research papers, and software engineering publications. Information related to Agile

methodologies and software engineering practices was analysed to understand their applications in modern software development.

The methodology used in this study includes:

Reviewing software engineering literature and Agile publications.

Comparing traditional and Agile software development approaches.

Studying Scrum and Extreme Programming frameworks.

Examining Agile practices used in software organisations.

The study follows a descriptive and qualitative research approach.

Result and Discussion

1. Software Development Life Cycle

The Software Development Life Cycle (SDLC) is a structured process used for software creation. The major phases include:

- Requirement Analysis
- System Design
- Coding and Implementation
- Testing
- Deployment
- Maintenance

Traditional development models provide systematic planning but often struggle to adapt to changing customer requirements.

2. Agile Technology

Agile technology is a flexible software development methodology that emphasises customer collaboration, continuous delivery, and adaptability. Agile projects are divided into small iterations, enabling teams to develop and test software continuously.

Agile Manifesto Principles

Individuals and interactions over processes and tools

Working software over comprehensive documentation

Customer collaboration over contract negotiation

Responding to change by following a plan

These principles help organisations improve software quality and customer satisfaction.

3. Scrum Framework

Scrum is one of the most widely used Agile frameworks.

Scrum Roles

- Product Owner
- Scrum Master
- Development Team

Scrum Activities

- Sprint Planning
- Daily Scrum
- Sprint Review
- Sprint Retrospective

Scrum improves communication, transparency, and teamwork during software development.

4. Extreme Programming

Extreme Programming (XP) focuses on improving software quality and rapid delivery.

Important XP Practices

- Pair Programming
- Continuous Integration
- Test-Driven Development
- Refactoring
- Frequent Releases

XP encourages developer collaboration and continuous improvement.

5. Advantages of Agile Technology

Agile methodologies provide several benefits:

- Faster software delivery
- Improved software quality
- Better customer satisfaction
- Flexibility in handling changes
- Continuous testing and feedback
- Improved team collaboration

Organisations adopting Agile methods can respond quickly to market demands and technological changes.

6. Challenges of Agile Technology

Despite its advantages, Agile development faces some challenges:

- Limited documentation
- Requirement of experienced teams
- Difficulty in managing large distributed teams
- Frequent requirement changes affecting project scope

Proper communication and project planning are necessary for successful Agile implementation.

Table 1: Comparison Between Agile and Traditional Models

Feature	Agile Model	Traditional Model
Development Style	Iterative	Sequential
Flexibility	High	Low
Customer Involvement	Continuous	Limited
Testing	Continuous	End Phase
Delivery	Frequent	Single Delivery

The comparison shows that Agile methodologies provide better adaptability and customer involvement compared to traditional software development methods.

Conclusion

Software engineering provides systematic methods for developing reliable software systems. Traditional development models offered structured planning but lacked flexibility in changing environments. Agile technology transformed software development by introducing iterative development, continuous testing, customer collaboration, and adaptability.

Agile frameworks such as Scrum and Extreme Programming improve software quality, reduce project risks, and increase productivity. Although Agile methodologies face challenges related to documentation and team coordination, their advantages make them highly effective in the modern software industry.

The study concludes that Agile technology plays a significant role in modern software engineering and will continue to influence future software development practices.

References

1. Beck, K. (2005). Extreme programming explained: Embrace change (2nd ed.). Addison-Wesley.
2. Martin, R. C. (2003). Agile software development: Principles, patterns, and practices. Pearson Education.
3. Pressman, R. S. (2019). Software engineering: A practitioner's approach (9th ed.). McGraw-Hill Education.
4. Schwaber, K., & Sutherland, J. (2020). The Scrum guide. Scrum.org.
5. Sommerville, I. (2016). Software engineering (10th ed.). Pearson Education.

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 177 - 187 |

Advanced Bidirectional Converters for Sustainable Energy and Green Innovation

[1]Rekha P

[2]V. Sridevi

[3]Amaleswari Rajulapati

[1]Research Scholar, Department of EEE, Academy of Maritime Education and Training (AMET) University, ECR, Kanathur, Chennai

[2]Professor & Head, Department of EEE, Academy of Maritime Education and Training (AMET) University, ECR, Kanathur, Chennai

[3]Assistant Professor, Department of EEE, Academy of Maritime Education and Training (AMET) University, ECR, Kanathur, Chennai

Email: rekha.paramanathan@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734673

DOI: 10.5281/zenodo.20734673

Abstract

The worldwide demand for energy has been increasing rapidly, the non-renewable conventional fossil fuel reserves have been depleting gradually, and the environmental pressures have been rising. All these factors have contributed to the quest for sustainable energy technologies around the world. Bidirectional power converters are among the most impactful innovations in this field, and have become a cornerstone enabling technology in renewable energy, electric vehicles, smart grids, energy storage, and marine electrification platforms. Unlike unidirectional converters, they allow two-way power flow between energy sources and loads, which makes the transfer of energy very efficient, as well as full regenerative operation and intelligent real-time power management. In this chapter, a detailed academic discussion about the importance of bidirectional converters in sustainable technology and green innovation is provided. It covers a wide spectrum of subjects such as converter topologies, basic operating principles, advanced control techniques, emerging semiconductor materials, application areas, current problems and future research directions. The chapter also measures and puts energy efficiency optimization, carbon emission reduction, renewable energy integration and smart

energy management system development into perspective through the contribution of bidirectional converters in industry, transportation, and infrastructure.

Allied with this, the presence of keywords like Bidirectional converters, Sustainable technology, Green innovation, Renewable energy systems, Electric vehicles (EVs), Smart grids, Energy storage systems, Power quality, DC–DC converters, Vehicle-to-grid (V2G), SiC, GaN, Machine learning, Power electronics reinforces the market's current focus on new technology. Furthermore, the inclusion of terms such as Bidirectional converters, Sustainable technology, Green innovation, Renewable energy systems, Electric vehicles (EVs), Smart grids, Energy storage systems, Power quality, DC–DC converters, Vehicle-to-grid (V2G), SiC, GaN, Machine learning, and Power electronics indicates the market's current emphasis on new technology.

Keywords: Bidirectional converters, Sustainable technology, Green innovation, Renewable energy systems, Electric vehicles (EVs), Smart grids, Energy storage systems, Power quality, DC–DC converters, Vehicle-to-grid (V2G), SiC, GaN, Machine learning, Power electronics

Introduction

The increasing demand for electrical energy has given rise to significant challenges in conventional fossil fuel-based power systems that produce a substantial amount of greenhouse gas emissions and environmental pollution. Consequently, shifting towards sustainable and renewable energy systems has become a worldwide priority. There is a growing number of renewable energy sources like solar, wind, fuel cells, and energy storage systems, but they are intermittent and need efficient power management technologies. The bidirectional converter has become an important solution as it allows for bidirectional power transfer between the sources, storage devices, and loads. They can be used in various applications, including renewable energy integration, electric vehicle charging, Vehicle-to-Grid (V2G) systems, regenerative braking, and smart grids, unlike traditional converters. Besides energy conversion, bidirectional converters can also help to improve power quality, increase energy efficiency, and enable intelligent energy management. Hence, they are crucial for developing flexible, reliable and environmental friendly energy infrastructures for future green technologies.

Basics of Bidirectional Converters

1. The UX Design Process

A bidirectional converter is a power electronic device capable of transmitting electrical power in both directions between two energy sources, e.g. between an AC network and a DC network or between two DC sources. The flow of the power depends on the system requirement and the control strategy. The same converter can be used as a rectifier, inverter, buck converter or boost converter depending on the

operating mode. The bidirectional converter is a converter that allows power to flow in either direction, as opposed to a conventional unidirectional converter, which uses passive semiconductor switches rather than active ones. The bidirectional switching capability allows for flexible and efficient power transfer in both directions, which makes bidirectional converters a crucial component in contemporary energy systems.

2. Operating Principle

Bidirectional converters are devices that have evolved greatly since the advent of better switching devices. Power conversion devices were commonly silicon MOSFETs and IGBTs, but more recently, wide bandgap semiconductors like silicon carbide (SiC) and gallium nitride (GaN) have been shown to deliver better performance. These devices have higher switching frequency, low loss, good thermal performance and efficiency. Due to this, SiC and GaN based bidirectional converters are smaller, lighter, more efficient and reliable than conventional silicon-based converters.

Switching Devices

Each of the silicon MOSFET, IGBT, SiC MOSFET and GaN HEMT has different speed, loss, voltage and cost trade-offs.

3. Modes of Operation

At any one time, each bidirectional converter is in one of two basic modes. Forward power flow – when power flows from the primary power source (like the grid or renewable generation) to the secondary power system (like a motor drive or a battery pack). For the charging, rectification or buck, this is the same as above. When in Reverse Power Flow mode, energy flows from the storage element or secondary source back to the primary bus. This is equivalent to discharging, inversion or boost operation. Bidirectional converters are extremely responsive to the dynamic requirements of modern energy systems due to the seamless transition between modes and the ability to switch in microseconds under digital control.

Classification of Bidirectional Converter

Bidirectional converters are a wide and varied class of power electronic circuits, and are categorized mainly according to the type of energy port they connect. There are three major types: DC–DC bidirectional converter, AC–DC bidirectional converter, and DC–AC bidirectional inverter. There are several sub-topologies in each category, with optimized voltage levels, ratings, isolation, efficiency and control complexity for specific application needs. The classification is crucial for the choice of the converter architecture for specific sustainable energy applications.

1. Bidirectional DC–DC Converters

Bidirectional DC-DC converters are used in many applications to send power in both directions between two DC voltage systems, particularly in electric vehicles, battery storage and renewable energy. The most basic topology is the bidirectional half-bridge converter in which the voltage level is reduced or boosted, depending on whether the converter is in buck mode or boost mode, respectively. Depending on the switching control strategy, the direction of power flow is reversed. Isolated bidirectional converters are used for high power applications to offer electrical isolation and high voltage conversion ratios through high frequency transformers. The Dual Active Bridge (DAB) converter is one of these, utilizing soft-switching operation, efficient phase-shift control, and rapid switching response, which make it a popular choice in EV chargers and solid-state transformers.

2. Bidirectional AC–DC Converters

Bidirectional AC–DC converters are also known as four quadrant converters or active front end converters which are used to connect AC power system with DC bus and allow energy flow in both directions. In grid connected applications, they can function as Power Factor Correction (PFC) rectifiers when powering from AC to DC (Forward mode) or as a grid tied inverter when powering from DC to AC (Reverse mode). The most popular topology adopted is the three-phase Voltage Source Converter (VSC) with pulse-width modulation (PWM) that is capable of producing a sinusoidal current waveform in phase with grid voltage, resulting in unity power factor, low harmonic distortion and seamless bidirectional operation.

3. Bidirectional DC–AC Inverters

Bidirectional DC–AC inverters are capable of converting DC power to AC power for grid injection or load power supply, as well as converting AC power to DC power for storage charging. In PV and wind power systems connected to the grid, the inverter steps the DC power from the PV array and/or battery into AC power that is compatible with the grid. Conversely, the same inverter can draw AC electricity from the grid to charge the connected storage devices. The bidirectional inverter is the control unit in microgrid applications, controlling voltage and frequency in islanded (off-grid) mode and synchronizing with the utility when reconnection is needed.

Incorporating Sustainable Technology and Energy Efficiency

In power electronics and energy systems, sustainable technology involves designing, implementing, and managing energy conversion and management systems that have the least impact on the environment, are most efficient in using renewable energy resources, and are continually improving the efficiency of energy use throughout the entire energy value chain. It's not just about the hardware of energy conversion converters, inverters, transformers, but also the software

intelligence that optimizes their performance and the system architecture that connects together a variety of energy sources and loads into a unified, intelligent energy system.

Bidirectional converters can make a contribution to sustainability in multiple and complementary ways. First, they reduce the percentage of energy lost into heat through energy conversion with efficiencies that are now regularly greater than 97-98% in state-of-the-art SiC designs. Each percent of efficiency gained is directly proportional to the reduction in fuel use, carbon emissions, and heat sink. Secondly, they can be used to store energy in a local area, near where it is produced, and use it in a local area, near where it is consumed, which minimizes the need for long-distance transmission and the associated losses due to resistance in the transmission lines. Third, they recover energy that would otherwise be irretrievably lost in mechanical braking or load dumping, thus "recapturing" value from energy flows previously "wasted. Fourth and most important, they allow intelligent, adaptive management of complex energy systems that balance supply and demand in real-time to smooth the integration of inherently variable renewable energy generation.

The basic measurement of converter's sustainability is the energy conversion efficiency, which is the ratio of useful output power to the total power of the input. High efficiency is not just a performance measure – it's a sustainability requirement. If a bidirectional converter is 95% efficient, 5% of all the energy that flows through it is lost as heat. Even at grid scale, where the converters can handle gigawatt-hours per year of energy, an efficiency gain of one percent is huge energy savings and carbon emissions reductions. Thus, the design of modern converters is always aimed at achieving better efficiency by employing novel semiconductor devices, optimized magnetic design, innovative modulation techniques, and accurate thermal management.

$$Efficiency(\eta) = \frac{P_{out}}{P_{in}} X100\%$$

SiC bidirectional converters of the modern design have a high efficiency, $\eta > 98\%$ over a broad load range.

Part in Renewable Energy Systems

The variable nature of renewable energy systems is a key challenge for decades of reliable grid operation and for independent energy systems, and constitutes a fundamental challenge to humanity's pathway to a decarbonized energy future. Solar PV generation is dependent on the diurnal cycle of solar radiation, is highest at midday and lowest at night, and is also influenced by cloud, dust and seasonal variation in the angle of the sun. Wind power is variable in nature, with wind speed varying from second-by-second to days, depending on the weather. These variability properties indicate that energy output from renewable generation sources does not match energy demand very closely, and energy storage and sophisticated

energy management and bidirectional energy converters are the technical basis that makes it possible.

1. Solar Photovoltaic Systems

The bidirectional DC-DC converter is a key component in a solar PV with battery energy storage (BESS) system that dynamically controls the flow of energy to and from the PV array and battery bank, and to the inverter and loads that are connected to the DC bus. When PV generation is higher than demand, the sun is bright and the load is low, the bidirectional converter works in forward power direction to charge the battery bank and the surplus power is stored in the battery bank for subsequent use. When sun power drops below the load demand, the converter will automatically switch to reverse mode, charging the battery from the energy storage device so as to keep the DC bus voltage up and the power flowing to the loads during the night.

Modern bidirectional DC-converters for PV-systems use Maximum Power Point Tracking (MPPT) algorithms, which continuously track the operating voltage of the PV-array in order to maximize the PV-array's output power under all weather conditions and temperatures. The bidirectional converter can be combined with predictive algorithms that optimize the battery management system, allowing it to predict the next day's solar production from the weather forecast, and making the battery backed solar system an intelligent, self-optimising energy management platform. The net effect is increased renewable energy consumption, decreased curtailment of renewables, decreased electricity costs, and increased energy self-sufficiency for the end user.

2. Wind Energy Systems

Wind power systems also have unique power management challenges when compared to solar. Solar generation is relatively smooth on an hourly time scale, whereas wind generation can be very dynamic and unpredictable on timescales of seconds to minutes between gusts and lulls passing through a wind farm. If not managed, these fast power changes result in voltage and frequency deviance on the connected electrical network, which can affect the stability of sensitive loads, and the deviation may be a violation of the grid codes. Bidirectional converters in wind storage systems are able to store excess energy from the wind during wind gust periods in the battery bank, avoiding over-frequency as well as release stored energy when wind conditions are low, maintaining the output level and avoiding under-frequency, giving a smoothing and buffering effect and making wind power more reliable and grid-friendly.

3. Electro fuels and Hybrid Energy Systems

Fuel cells, are a special type of renewable or low carbon generation method, that could take the surplus renewable electricity generated through electrolysis and

convert it into electricity with a high efficiency and without any direct emissions of CO2. Fuel cells, however, have a slow dynamic response; that is, they cannot rapidly respond to changes in load, or to rapid increases or decreases in generation, to follow the changes in power output. Bidirectional converters are used as the electrical bus for power control in hybrid energy systems of fuel cells and batteries or supercapacitors to communicate with the battery to absorb power surges from the load when the battery is fast acting, and to send the average load to the fuel cell. The Bi-directional converter is also used to reverse the electrolysis process to generate hydrogen power for long-term storage when there is an excess of renewable energy. The combination of hydrogen and electricity in this cycle, with the intelligent bidirectional converter as its manager, is one of the most promising routes to long-duration seasonal energy storage.

Green Innovation Through Power Electronics

Green innovation is the systematic application of scientific and engineering creativity to the development of technologies, processes, systems and products that achieve the same or greater economic or social value, while significantly lower the environmental impact relative to the systems that they are replacing. Green innovation in energy is not an individual technology or a single breakthrough but a process of continuous and accelerating improvement and integration between materials science, power electronics, digital control, AI and systems engineering. Bidirectional converters are at the intersection of many of these disciplines, and are both products of green innovation, and key drivers of further innovation throughout the energy ecosystem.

Bidirectional converters introduce green innovation on various levels and in different application areas. In the device level, the ongoing power conversion devices efficiency gains thanks to newer semiconductors, optimized magnetics and better modulation algorithms directly translate to less energy being wasted during every power conversion event. Bidirectional converters open up the energy system architectures that were technically unfeasible without them: microgrids that seamlessly island and reconnect; V2G systems that turn transportation infrastructure into grid assets; and renewable-hydrogen energy systems that can store clean energy across seasons. The widespread adoption of bidirectional converters in electric vehicles, smart buildings and renewable energy systems is essential on a societal level, if we want to meet the net-zero emission goals that climate science requires if we want to keep global warming within manageable limits.

As a result of the developments in bidirectional power converter technology, coupled with the Internet of Things (IoT), cloud computing and artificial intelligence, an intelligent energy system is currently emerging as a new paradigm. Bidirectional converters are the physical actuation layer for smart energy ecosystems where buildings, vehicles, storage systems and grid infrastructure

interact and coordinate the flow of power in real-time to reduce cost, carbon, and ensure reliability. All instructions to the cloud-based energy management system turn into changes in the switching sequence of a bidirectional converter somewhere in the system. Power electronics represent the bridge between the digital and physical domains of data, algorithms and optimizations, and electrons, electromagnetic fields and energy flows. Green innovation with power electronics is not only about the improvement of the converters, it is the whole new relationship between digital intelligence and physical power infrastructure.

Comparative Analysis: Unidirectional Vs Bidirectional Converters

When comparing the most relevant technologies for sustainable energy applications of unidirectional and bidirectional converter, as presented in Table 1, both the benefits of bidirectional designs and the contexts in which they are valuable becomes apparent. The power delivery requirements in many common applications (where there is only one direction of power flow and cost is a primary concern) are still simple and single direction, and so the unidirectional converter is still suitable. However, the complexity and intelligence demand of today's energy systems are increasing, and bidirectional power delivery systems are more appropriate and cost-effective.

Table 1: Comparative Analysis of Unidirectional Vs Bidirectional Converters

S.no.	Feature	Unidirectional Converter	Bidirectional Converter
1	**Power Flow Direction**	One-way only	Two-way: forward and reverse
2	**Energy Recovery**	Not possible	Full regenerative recovery
3	**Renewable Integration**	Limited, requires extra hardware	Seamless, native integration
4	**Regenerative Braking**	No energy lost as heat	Yes energy returned to battery
5	**V2G / Grid Support**	Not supported	Full V2G and grid services
6	**Battery Management**	Charge only	Charge, discharge, and balance
7	**Smart Grid Compatibility**	Limited	Advanced, fully compatible
8	**Power Quality Control**	Basic	Active harmonic and PFC control
9	**Control Complexity**	Low	Higher, but AI-assisted

10	**Cost**	Lower	Higher, but lower lifecycle cost
11	**Sustainability Rating**	Moderate	High supports net-zero goals
12	**Power Flow Direction**	One-way only	Two-way: forward and reverse

The comparative analysis shows a clear trend in all the aspects that are relevant to the sustainable energy system, bidirectional converters seem to be superior in every aspect ranging from the flexibility in power flow to energy recovery, from smart grid support to energy system performance. The extra hardware complexity and upfront investment in bidirectional designs are becoming more and more economically viable due to the economic value of V2G services, the energy savings from regeneration, the renewable integration benefits of energy storage management, and the benefits to grid resilience from microgrid capability. The energy system is moving towards a greater share of renewables and increased decentralisation and smartness, and the power conversion applications where only unidirectional conversion is required will become increasingly rare; bidirectional power conversion will become the standard requirement for power electronic systems in sustainable energy applications.

Conclusion

The chapter has covered the impact of bidirectional converters on the sustainable energy technologies and green innovation in a comprehensive manner. The basics of the converter, its classification, control strategies, development of semiconductors, and its applications in renewable energy systems and electric vehicles, smart grid, and industry were discussed. In contrast to conventional unidirectional converters, bidirectional energy transfer is achieved in an intelligent manner, resulting in flexible and efficient power system architectures. The real value of their application is that they can enable energy storage, regenerative operation, integration of renewable energy and smart energy management all in one. The potential applications of bidirectional power conversion in lowering carbon emissions and enhancing energy use are far-reaching, including technologies like vehicle-to-grid (V2G) systems, renewable microgrids, and regenerative braking. Converter performance and reliability continue to improve due to ongoing development of wide bandgap semiconductor devices, digital controllers, and energy management using AI methods. Meanwhile, advancements in battery technology and the changing smart grid systems are making bidirectional systems more useful in real-world applications. Advanced converter topologies, intelligent adaptive control algorithms and highly efficient semiconductor materials are some areas of future research opportunities. The development of bidirectional converter

technologies is a key milestone in the path to cleaner, smarter and more sustainable energy systems, from an industrial and societal point of view. Thus, bidirectional converters will remain key in supporting the global shift towards low carbon and energy-efficient infrastructures.

References

1. Rashid, M. H. (2014). Power Electronics: Circuits, Devices and Applications, 4th ed. Pearson Education. Full coverage of the basics of power electronics topologies and principles.
2. Erickson, R. W. and Maksimovic, D. (2020). Fundamentals of Power Electronics 3rd Edition. Springer. The authoritative book on converter analysis, modelling and control.
3. Bose, B. K. (2001). Modern Power Electronics and AC drives. Prentice Hall. A traditional approach to power electronic interfaces and AC drive systems.
4. Mohan, N., Undeland, T. M., and Robbins, W. P. (2003). Power Electronics: Converters, Application and Design (3rd ed.). Wiley. A standard undergraduate and graduate reference text.
5. IEEE transactions on power electronics. Institute of Electrical and Electronics Engineers. Top peer-reviewed power electronics research and development journal.
6. IEEE Transactions on Industrial Electronics. Institute of Electrical and Electronics Engineers. Industrial power electronics applications' leading journal.
7. Renewable & Sustainable Energy Reviews. Elsevier. Interdisciplinary Journal of Renewable Energy Systems covering all aspects of renewable energy systems.
8. International Journal of Electrical Power and Energy Systems. Elsevier. Includes topics such as power system analysis, optimization and energy management.
9. Journal of Cleaner Production. Elsevier. Interdisciplinary journal on sustainable production and environment.
10. Sustainable Energy Technologies and Assessments. Elsevier. Special emphasis on the technical evaluation of sustainable energy technologies and systems.
11. Blaabjerg, F., et al. (2017). Renewable Energy Power Conversion Systems. IEEE Press. Advanced treatment of design of power converter for integration of renewable energy.
12. Khaligh, A., Li, Z. (2010). Battery, Ultracapacitor, Fuel Cell and Hybrid Energy Storage Systems in Electric, Hybrid Electric, Fuel Cell and Plug-In Fuel Cell vehicles. IEEE Transactions on Vehicular Technology (TVT).

13. Amaleswari Rajulapati, M. Prabhakar, “Comparison of Non-isolated High Gain Multi-input DC-DC Converters”, International Journal of Renewable Energy Research, Vol. 11, Issue. 03, 2021, pp. 981-991, Sep 2021.
14. K Bhavana, K Lalitha, Guruswamy Revana, R. Amaleswari “An Improved Human Evolutionary Optimization Algorithm for Optimal Integration of Hybrid Fast Charging Stations”, International Journal of Intelligent Engineering and Systems, vol. 18, no. 4, pp. 813-825, Mar 2025. DOI: 10.22266/ijies2025.0531.52.
15. D. Lakshmi, C.N. Ravi, R. Zahira, Sivaraman Palanisamy, Sharmeela Chenniappan, Chapter 1 - Introduction to renewable energy sources and bulk power system, Power Systems Operation with 100% Renewable Energy Sources, Elsevier,2024, Pages 1-13, ISBN 9780443155789, https://doi.org/10.1016/B978-0-443-15578-9.00008-X

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 188 - 193 |

Smart and Sustainable Marine Engineering: Technological Innovations, Societal Impact, and Human Transformation

A. Ananthi Christy

Department of Marine Engineering, AMET Deemed to be University, Chennai, Tamil Nadu, India

Email: chrisarun13@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734812

DOI: 10.5281/zenodo.20734812

Abstract

Marine engineering is undergoing a remarkable transformation due to rapid technological advancements, increasing environmental concerns, and the growing demand for sustainable maritime operations. The integration of advanced technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), machine learning, robotics, automation, digital twins, and big data analytics has significantly improved the efficiency, safety, and reliability of modern marine systems. Smart shipping technologies enable real-time monitoring, predictive maintenance, autonomous navigation, optimized fuel consumption, and enhanced decision-making processes, thereby reducing operational costs and improving maritime performance. These innovations are reshaping conventional maritime practices and supporting the development of intelligent and interconnected marine infrastructures.

At the same time, sustainability has become a major priority in marine engineering due to global climate change, marine pollution, and stringent environmental regulations introduced by international maritime organizations. To minimize environmental impacts, the maritime industry is increasingly adopting green propulsion systems, low-emission fuels, renewable energy integration, hybrid engines, hydrogen-based technologies, and energy-efficient ship designs. These sustainable engineering solutions contribute to reducing greenhouse gas emissions and promoting environmentally responsible maritime transportation.

Technological advancements in marine engineering have also created significant societal and human transformations. Automation and intelligent systems have improved maritime safety, reduced manual labor, and enhanced global trade and logistics operations. However, these developments have changed workforce

requirements by increasing the need for professionals skilled in digital technologies, cybersecurity, data analytics, and sustainable engineering practices. Human–machine collaboration is becoming an important aspect of future maritime industries.

Despite these benefits, several challenges remain, including high implementation costs, cybersecurity threats, technological complexity, workforce adaptation, and regulatory compliance. Therefore, achieving a balance between technological innovation, environmental sustainability, and ethical responsibility is essential for the future growth of the maritime sector. This chapter highlights the role of smart and sustainable marine engineering in shaping modern society and promoting resilient, efficient, and eco-friendly maritime development.

Keywords: Sustainable Maritime Technology, Internet of Things (IoT), Renewable Energy, Industry 4.0.

Introduction

Marine engineering has become an essential component of global transportation, trade, and sustainable industrial development. The rapid advancement of smart technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), automation, robotics, and digital monitoring systems has significantly transformed modern maritime operations (Durlik et al., 2024). These technologies improve navigation safety, fuel efficiency, predictive maintenance, and operational reliability in shipping industries (Prousaloglou et al., 2025). At the same time, growing environmental concerns and international regulations have encouraged the adoption of sustainable marine engineering practices, including green propulsion systems, renewable energy integration, and low-emission fuels (Spalding, 2026). Smart shipping and autonomous vessel technologies are also reshaping maritime logistics and global trade networks while reducing human error and operational costs (Blanco-Davis et al., 2026). Furthermore, technological advancements have transformed workforce requirements by increasing the demand for skilled professionals in automation, cybersecurity, and sustainable engineering systems (Anderson, 2023). Despite these benefits, challenges such as cybersecurity threats, high implementation costs, and environmental risks remain critical issues in the maritime sector (Nguyen et al., 2024). Therefore, smart and sustainable marine engineering plays a vital role in achieving efficient, safe, and environmentally responsible maritime development for future generations.

Objectives

The primary objective of this chapter is to explore the role of smart and sustainable technologies in transforming modern marine engineering systems and their impact on society and human development. The specific objectives are:

- To analyze recent technological innovations in marine engineering, including Artificial Intelligence (AI), IoT, automation, robotics, and digital monitoring systems.
- To examine the importance of sustainable marine technologies such as green propulsion systems, renewable energy integration, and low-emission fuels in reducing environmental impacts.
- To evaluate the societal and economic impacts of smart marine engineering on global trade, maritime safety, logistics, and industrial growth.
- To investigate human transformation in the maritime sector, including workforce skill development, human–machine interaction, and digital competency requirements.
- To identify major challenges associated with smart and sustainable marine systems, including cybersecurity risks, implementation costs, and regulatory issues.
- To provide future directions for achieving environmentally sustainable and technologically advanced maritime development.

Data and Methodology

This chapter is based on a qualitative and analytical research methodology using secondary data collected from research articles, technical reports, conference proceedings, maritime industry publications, and international organization reports. Relevant literature related to smart marine engineering, sustainable shipping technologies, maritime automation, AI applications, and environmental sustainability was systematically reviewed.

The collected data were analyzed using a thematic and comparative approach to identify major technological advancements, sustainability practices, societal impacts, and human transformation trends in marine engineering. Information from the International Maritime Organization (IMO), scientific databases, and recent engineering studies was used to examine the evolution of smart maritime systems and green marine technologies.

The methodology includes:

- Literature review of recent marine engineering technologies.
- Comparative analysis of conventional and smart marine systems.
- Evaluation of sustainability practices in maritime industries.
- Analysis of social, economic, and environmental impacts.
- Identification of future challenges and opportunities in marine engineering.

This approach provides a comprehensive understanding of how technological innovation and sustainability are reshaping the maritime sector.

Result and Discussion

The analysis reveals that smart technologies have significantly improved the operational efficiency, safety, and sustainability of marine engineering systems. The implementation of AI-based monitoring systems, IoT-enabled sensors, and automation technologies has enhanced real-time data collection, predictive maintenance, fuel optimization, and autonomous ship navigation. These technologies reduce operational downtime, improve decision-making processes, and minimize human error in maritime operations.

The study also indicates that sustainable marine engineering practices have contributed to reducing greenhouse gas emissions and marine pollution. The adoption of LNG fuels, hybrid propulsion systems, renewable energy technologies, and energy-efficient ship designs supports global environmental goals and maritime decarbonization strategies. Green technologies improve fuel efficiency and reduce dependence on conventional fossil fuels.

From a societal perspective, smart marine engineering has strengthened global trade, improved maritime logistics, and enhanced international connectivity. Technological advancements have increased employment opportunities in digital maritime services, automation, cybersecurity, and marine data analytics. However, automation has also reduced certain traditional manual jobs, creating the need for workforce reskilling and technical education.

The discussion further highlights major challenges, including high installation and maintenance costs, cybersecurity vulnerabilities in digital maritime systems, technological complexity, and strict environmental regulations. In addition, developing countries may face difficulties in adopting advanced marine technologies due to financial and infrastructural limitations.

Overall, the findings demonstrate that smart and sustainable marine engineering plays a crucial role in improving maritime performance, environmental protection, and human development while supporting the transition toward intelligent and eco-friendly maritime industries.

Conclusion

Smart and sustainable marine engineering has emerged as a transformative approach for modern maritime industries by integrating advanced technologies with environmentally responsible practices. The adoption of AI, IoT, automation, robotics, and digital monitoring systems has improved operational efficiency, maritime safety, predictive maintenance, and decision-making capabilities in shipping operations. Simultaneously, sustainable technologies such as green propulsion systems, renewable energy integration, and low-emission fuels have significantly contributed to reducing environmental pollution and supporting global climate objectives.

The study highlights that technological advancements in marine engineering not only influence industrial growth and global trade but also create substantial societal and human transformation. The increasing demand for digital skills, cybersecurity knowledge, and interdisciplinary engineering expertise reflects the evolving nature of maritime professions in the era of smart shipping and automation.

Despite these advancements, challenges such as high implementation costs, cybersecurity threats, workforce adaptation, and regulatory compliance remain significant concerns. Therefore, continuous research, technological innovation, international collaboration, and sustainable policy development are essential for achieving resilient and environmentally friendly maritime systems.

In conclusion, smart and sustainable marine engineering represents a key pathway toward future maritime development by balancing technological progress, environmental sustainability, and human-centered innovation for safer, cleaner, and more efficient marine operations.

Acknowledgements

The authors sincerely acknowledge the constructive comments and valuable suggestions of the reviewers, which have greatly enhanced the quality of this manuscript. This research was supported by the Academy of Maritime Education and Training (AMET), Deemed to be University, Chennai. The authors extend their heartfelt thanks to Shri J. Ramachandran, Chancellor; Col. Dr. G. Thiruvasagam, Provost; Dr. V. Rajendran, Vice-Chancellor; Dr. V. Sangeetha Albin (i/c), Registrar; and Dr. M. Jayaprakashvel, Coordinator, IQAC, AMET, Chennai, Tamil Nadu, for their continuous encouragement and support throughout this work.

References

1. Alamaniotis, M., & Ipiotis, K. (2025). Artificial intelligence as enabler for adoption of sustainable nuclear-powered maritime ships: Challenges and opportunities. Sustainability, 17(8), 3654. https://doi.org/10.3390/su17083654
2. Anderson, J. (2023). Maritime engineering applications of artificial intelligence for predictive analytics. American Journal of Marine Engineering and Technology, 4(1), 7–11. https://doi.org/10.71465/ajmet1422
3. Blanco-Davis, E., Loughney, S., & Yang, Z. (2026). Novel maritime techniques and technologies, and their safety. Journal of Marine Science and Engineering, 14(1), 30. https://doi.org/10.3390/jmse14010030
4. Durlik, I., Miller, T., Kostecka, E., Łobodzińska, A., & Kostecki, T. (2024). Harnessing AI for sustainable shipping and green ports: Challenges and opportunities. Applied Sciences, 14(14), 5994. https://doi.org/10.3390/app14145994
5. Nguyen, H. P., Nguyen, C. T. U., Tran, T. M., Dang, Q. H., & Pham, N. D. K. (2024). Artificial intelligence and machine learning for green shipping:

Navigating towards sustainable maritime practices. JOIV: International Journal on Informatics Visualization, 8(1). http://dx.doi.org/10.62527/joiv.8.1.2581

6. Prousaloglou, P., Kyriakopoulou-Roussou, M.-C., Stavroulakis, P. J., Tsioumas, V., & Papadimitriou, S. (2025). Artificial intelligence in the service of sustainable shipping. Journal of Ocean Engineering and Marine Energy, 11, 621–653. https://doi.org/10.1007/s40722-025-00390-0
7. Raine, S., & Fischer, T. (2025). AI-driven marine robotics: Emerging trends in underwater perception and ecosystem monitoring. arXiv. https://arxiv.org/abs/2509.01878
8. Saafi, S., Vikhrova, O., Fodor, G., Hosek, J., & Andreev, S. (2022). AI-aided integrated terrestrial and non-terrestrial 6G solutions for sustainable maritime networking. arXiv. https://arxiv.org/abs/2201.06947
9. Salmeri, F., Barbieri, L., Barberi, E., & Cucinotta, F. (2025). Eco-design and environmental performance of smart buoy for autonomous marine monitoring. Technology and Science for the Ships of the Future. https://doi.org/10.3233/PMST250015
10. Spalding, M. J. (2026). AI, maritime decarbonization, and ocean conservation. Sustainability, 18(5), 2337. https://doi.org/10.3390/su18052337
11. Stavroulakis, P. J., Prousaloglou, P., Kyriakopoulou-Roussou, M.-C., & Papadimitriou, S. (2025). AI-driven data mining for sustainable ship power system technologies. Transportation Engineering, 21, 100383. https://doi.org/10.1016/j.treng.2025.100383
12. Xiaocai, Z., Zhe, X., Maohan, L., Tao, L., Haijiang, L., & Wenbin, Z. (2026). Realistic curriculum reinforcement learning for autonomous and sustainable marine vessel navigation. arXiv. https://arxiv.org/abs/2601.10911

Technology, Society and Human Transformation

ISBN: 978-93-49938-52-6 | Year: 2026 | pp: 194 - 208 |

Ethical Challenges of Artificial Intelligence in Modern Society

Ranjana

Department of Zoology, Patna Science College, Patna University, Patna, Bihar, 800005, India

Email: ranjana.prakash81@gmail.com

Article DOI Link: https://zenodo.org/uploads/20734897

DOI: 10.5281/zenodo.20734897

Abstract

Artificial intelligence (AI) is reshaping virtually every domain of contemporary life, from healthcare and criminal justice to education, finance, and political discourse. While the technological capabilities of modern AI systems are remarkable, their rapid deployment has outpaced the development of robust ethical frameworks and regulatory oversight. This chapter examines the principal ethical challenges posed by AI in modern society, including algorithmic bias and discrimination, violations of privacy and surveillance, the erosion of human autonomy through automation and persuasion, accountability gaps in opaque decision-making systems, labour displacement, and the existential risks posed by advanced autonomous systems. Drawing on scholarship from philosophy, law, computer science, and the social sciences, the chapter argues that addressing these challenges requires interdisciplinary collaboration, inclusive public deliberation, and the co-development of governance mechanisms that keep pace with technological change. The chapter concludes by proposing a framework of ethically grounded AI development centred on transparency, fairness, accountability, and human dignity.

Keywords: artificial intelligence, ethics, algorithmic bias, surveillance, autonomous systems, data privacy

Table 1. Definitions of Key AI Ethics Concepts

Concept	Definition	Key Source
Algorithmic Bias	Systematic and unfair discrimination in AI outputs arising from flawed data, design, or evaluation choices.	Barocas et al. (2023)

Concept	Definition	Key Source
Explainability	The degree to which the internal logic of an AI system can be understood and communicated in human-intelligible terms.	Arrieta et al. (2020)
Accountability	The capacity to assign responsibility for AI-related harms to identifiable human or institutional agents.	Floridi et al. (2018)
Autonomy Preservation	Designing AI systems so they support rather than undermine individuals' epistemic and decisional independence.	Jobin et al. (2019)
Fairness	Absence of unjust discrimination; multiple formal definitions (statistical parity, equalized odds) exist and may conflict.	Chouldechova (2017)
Human Oversight	The requirement that meaningful human review and intervention remain possible throughout the AI decision lifecycle.	EU AI Act (2024)
Surveillance Capitalism	A regime in which behavioural data extracted by AI systems is commodified to predict and influence human action.	Zuboff (2019)
Responsibility Gap	A situation in which the autonomous behaviour of AI renders conventional attribution of moral responsibility to humans inadequate.	Matthias (2004)

Source: Compiled by the author from sources cited in text.

Introduction

Few technological developments in recent decades have attracted as much scholarly attention, public debate, and regulatory concern as artificial intelligence. Broadly defined, AI encompasses computational systems capable of performing tasks that

would ordinarily require human intelligence — including pattern recognition, natural language processing, strategic game-playing, autonomous navigation, and complex decision-making (1). From recommendation algorithms that curate what individuals read and watch, to predictive policing tools that influence who is placed under surveillance, to clinical decision-support systems that inform medical diagnosis, AI is no longer a speculative technology of the future. It is an operational infrastructure of the present.

The ethical implications of this transformation are profound. Questions that were once the province of academic philosophy — concerning fairness, accountability, autonomy, and the nature of intelligence itself — have become urgent practical concerns for policymakers, corporate boards, civil society organisations, and ordinary citizens. Yet the pace of AI deployment has consistently outstripped the development of normative guidance, regulatory frameworks, and institutional oversight mechanisms (2). The result is a widening gap between what AI systems can do and what society has collectively decided they should do.

This chapter provides a systematic account of the major ethical challenges associated with AI in modern society. Section 2 addresses the problem of algorithmic bias and discrimination. Section 3 examines AI-enabled surveillance and the erosion of privacy. Section 4 considers the implications of automation for human autonomy and labour. Section 5 analyses accountability and the opacity of AI decision-making. Section 6 discusses long-horizon risks associated with advanced AI systems. Section 7 reviews emerging governance approaches. The chapter concludes by arguing for an integrated ethical framework capable of guiding the responsible development and deployment of AI.

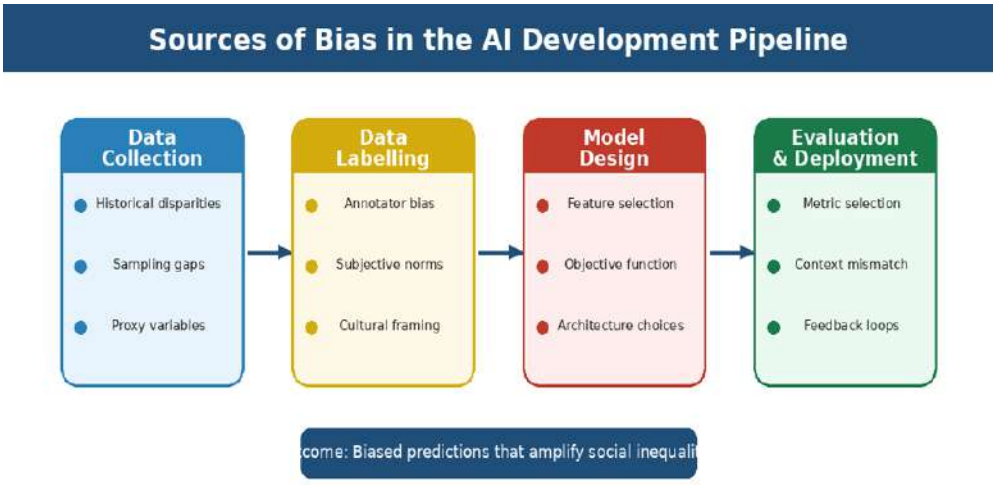
Algorithmic Bias and Discrimination

1. The Sources of Bias in Machine Learning Systems

A recurring concern in the AI ethics literature is the tendency of machine learning systems to perpetuate, and in some cases amplify, existing social inequalities (3). Bias in AI systems can arise at multiple stages of the development pipeline. Training data may reflect historical patterns of discrimination — for example, if a recidivism prediction tool is trained on arrest records from a period in which certain communities were disproportionately policed, the model will likely encode those disparities and project them into future predictions (4). Similarly, the choice of outcome variables, the labelling practices of human annotators, and the evaluation metrics selected by engineers all involve normative judgements that can introduce or entrench bias.

The landmark analysis by Angwin and colleagues of the COMPAS recidivism prediction algorithm, widely used in the United States criminal justice system, found that the tool incorrectly flagged Black defendants as higher risk at nearly twice the rate of White defendants (5). Cases such as this have catalysed a field of

research known as algorithmic fairness, which seeks to define and operationalise what it means for an AI system to treat people equitably.



Sources of bias across the AI development pipeline, from data collection through to deployment (adapted from Barocas, Hardt & Narayanan, 2023).

Scholars in the algorithmic fairness field have shown that common mathematical definitions of fairness are frequently mutually incompatible, meaning that no single system can satisfy all of them simultaneously (6). This result has important practical and philosophical implications: fairness is not a purely technical property, but involves contestable value choices about which inequalities matter most and to whom. Table 2 below summarises and compares the principal formal fairness criteria in current use.

Table 2. Comparison of Formal Fairness Criteria in Machine Learning

Criterion	Definition	Strengths	Limitations
Statistical Parity	Equal outcome rates across demographic groups regardless of base rates.	Easy to measure; intuitive	Ignores genuine risk differences
Equalized Odds	Equal true positive and false positive rates across groups.	Controls error symmetry	May require unequal thresholds
Predictive Parity	Equal positive predictive value across groups (precision).	Calibrated predictions per group	Incompatible with equalized odds

Criterion	Definition	Strengths	Limitations
Individual Fairness	Similar individuals should receive similar predictions.	Principled; intuitive	Requires contested similarity metric
Counterfactual Fairness	Outcome unchanged if individual's protected attribute were different.	Captures causal intuition	Computationally complex; contested

Source: Adapted from Barocas, Hardt & Narayanan (2023); Chouldechova (2017).

2. Facial Recognition and Biometric Systems

Bias in AI systems is particularly consequential in the domain of biometric recognition. Seminal work by Buolamwini and Gebru demonstrated that commercial facial analysis systems from major technology companies exhibited dramatically higher error rates for darker-skinned women compared to lighter-skinned men — a disparity attributable in part to the racial and gender composition of the training datasets (7). These findings have far-reaching practical consequences, given that facial recognition technology is increasingly deployed in law enforcement, border control, and commercial authentication contexts where errors can have serious effects on individuals' liberty and welfare.

The response of the research community and civil society to these revelations has been mixed. While several jurisdictions have imposed temporary or permanent moratoriums on the use of facial recognition in public spaces — including a number of major cities in the United States and provisional restrictions under the European Union's AI Act — the technology continues to be deployed widely in contexts with limited oversight (8).

Privacy, Surveillance, and the Erosion of Intimacy

1. Data Collection and the Surveillance Economy

The commercial model that underpins much of the digital economy — in which access to online services is exchanged for the collection and monetisation of personal data — has created what Shoshana Zuboff has termed "surveillance capitalism" (9). Under this paradigm, AI systems are deployed not merely to process data that users knowingly provide, but to extract behavioural predictions and influence future behaviour in ways that users rarely understand and have limited ability to contest. The asymmetry of information and power between large technology companies and individual users is structurally analogous to earlier forms of economic extraction.

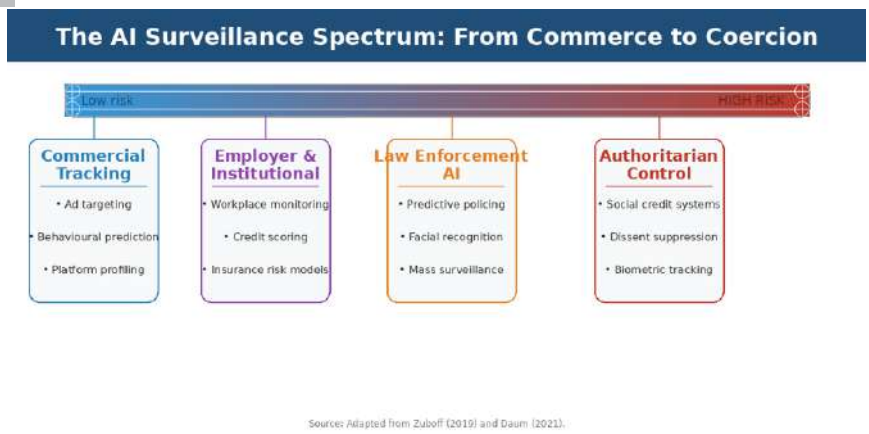
From a conventional privacy law perspective, the aggregation of individually innocuous data points into detailed behavioural profiles raises serious concerns about informational self-determination. The General Data Protection Regulation

(GDPR) introduced important rights — including the right of access, the right to erasure, and the right not to be subject to solely automated decision-making — but critics have noted that these instruments are difficult to enforce against large, technically sophisticated actors (10). Table 3 maps the principal AI surveillance mechanisms currently in operation alongside their actors, data sources, and ethical risk levels.

Table 3. AI Surveillance Mechanisms: Actors, Data Sources, and Risk Levels

Mechanism	Primary Actors	Data Sources	Ethical Risk Level
Behavioural Advertising	Commercial platforms	Clickstream, dwell-time, purchase history	**Moderate — GDPR consent required in EU**
Facial Recognition	Law enforcement, border agencies	Live CCTV, ID photos, social media images	**High — bans proposed in multiple jurisdictions**
Predictive Policing	Municipal police departments	Crime records, demographic proxies	**High — documented racial bias (Lum & Isaac, 2016)**
Social Scoring	State authorities (PRC)	Financial, social, civic behaviour data	**Critical — systemic freedom restriction**
Employer Monitoring	Private sector employers	Keystroke, webcam, location data	**Moderate — emerging legislative response**
Health Data AI	Insurers, healthcare providers	Medical records, wearables, genomics	**High — sensitive; re-identification risk**

Source: Compiled by the author; risk levels reflect scholarly consensus across cited literature.



The AI surveillance spectrum, ranging from commercial data tracking to authoritarian social control (adapted from Zuboff, 2019; Daum, 2021).

2. State Surveillance and Social Control

Beyond the commercial domain, AI-enabled surveillance technologies have dramatically expanded the capacity of states to monitor their populations. China's Social Credit System — a complex of local and national initiatives using AI to aggregate assessments of citizen behaviour — has attracted international attention as a paradigm case of algorithmic social control (11). Peer-reviewed scholarship has documented its use to restrict the travel and access to services of individuals deemed untrustworthy by state authorities, raising profound concerns about due process and freedom of expression.

Authoritarian applications of AI surveillance are not confined to non-democratic states. Research has documented the use of predictive policing tools, social media monitoring software, and automated licence plate readers by law enforcement agencies in liberal democracies, often with minimal legislative authorisation or judicial oversight (12). The chilling effect of pervasive surveillance on freedom of assembly, political participation, and intimate life represents a threat to democratic values that is only beginning to be addressed in law and policy.

Human Autonomy, Labour Displacement, and the Future of Work

1. Automation and Economic Displacement

What is distinctive about AI-driven automation is its scope: unlike previous waves of technological change, which primarily substituted for routine manual or clerical tasks, contemporary AI systems are increasingly capable of performing complex cognitive work, including legal research, medical image analysis, financial modelling, and journalistic writing (13). Influential forecasts have suggested that a substantial proportion of current occupations are susceptible to automation, with estimates ranging from approximately 9% to 47% of existing jobs in OECD economies (14). The ethical stakes of labour displacement are considerable. If the

gains from AI-driven productivity growth are captured primarily by capital owners and a small cadre of highly skilled workers, the result may be a significant increase in economic inequality with attendant social and political consequences (15).

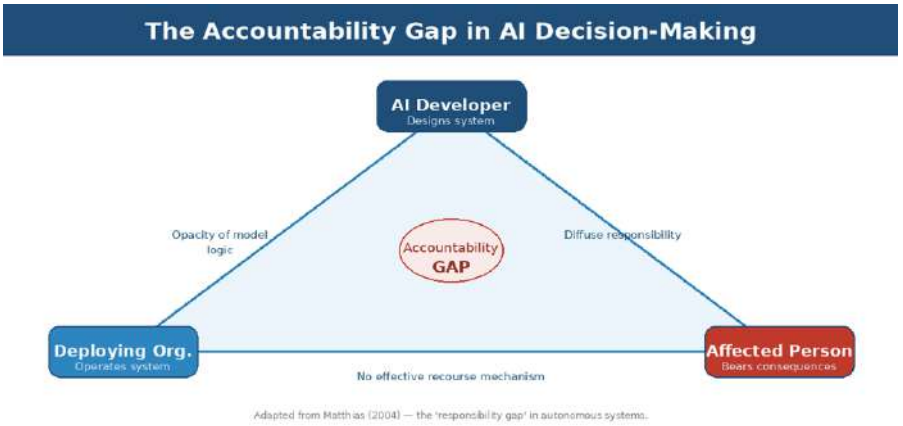
2. Epistemic Autonomy and Algorithmic Influence

A distinct concern about AI and human autonomy concerns the capacity of recommendation systems and targeted content delivery to shape individuals' beliefs and political orientations. Research on algorithmic persuasion has documented phenomena including filter bubbles and the amplification of emotionally engaging — often false or misleading — content in pursuit of engagement metrics (16). Generative AI systems capable of producing convincing synthetic text, images, audio, and video introduce further challenges for epistemic autonomy, with high-quality disinformation threatening the epistemic foundations of democratic deliberation (17).

Accountability, Transparency, and the Black Box Problem

1. Opacity in AI Decision-Making

Many of the most powerful contemporary AI systems are characterised by a degree of internal complexity that makes it difficult or impossible to explain, in terms intelligible to non-specialists, how any particular decision was reached (18). This opacity poses fundamental challenges for accountability. If an individual is denied a loan, rejected for employment, or flagged as a security risk by an automated system, elementary principles of procedural justice require that they be given an intelligible explanation and an effective opportunity to contest it. The field of explainable AI (XAI) has developed a range of technical approaches — including saliency maps, LIME, and SHAP — aimed at providing interpretable accounts of model behaviour, though critics have raised questions about their reliability and completeness (19, 20).

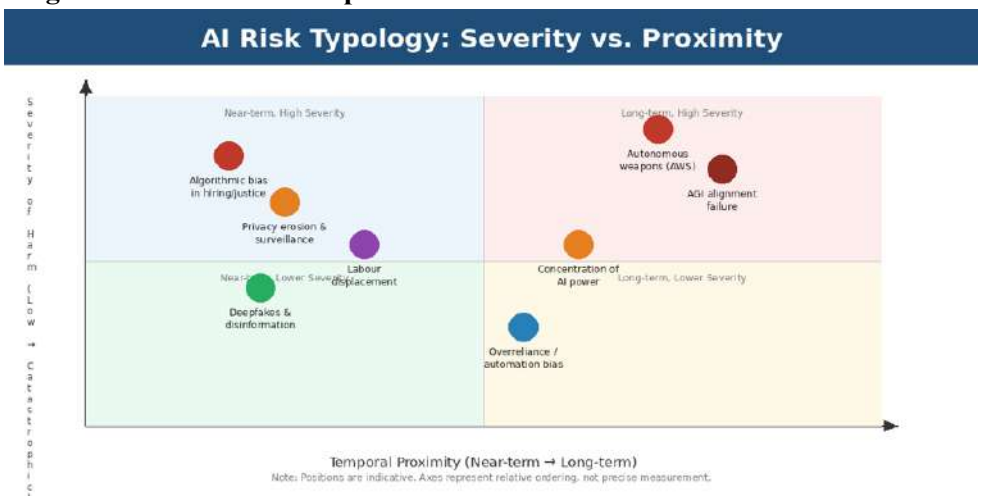


The accountability gap triangle: opacity and diffuse responsibility prevent affected persons from obtaining meaningful recourse (adapted from Matthias, 2004).

2. Responsibility Gaps and Distributed Agency

Beyond the technical problem of opacity, AI systems raise deeper questions about the distribution of moral and legal responsibility. When an autonomous vehicle causes a fatal accident, or when a medical AI system recommends a treatment that proves harmful, questions of responsibility become entangled across chains of causation involving system designers, training data curators, deploying organisations, regulatory bodies, and in some accounts the AI system itself (21). The philosopher Andreas Matthias has described this situation as a "responsibility gap" — a scenario in which conventional attribution of moral responsibility to human agents is undermined by the opacity and emergent behaviour of AI systems (22). The European Union's proposed AI Liability Directive seeks to address this gap, but scholars have noted significant limitations in its coverage and enforcement mechanisms (23).

Long-Horizon and Catastrophic Risks



AI risk typology matrix plotting risks by temporal proximity and severity of harm. Positions are indicative and reflect current scholarly consensus.

1. Existential and Global Catastrophic Risk

A distinctive strand of AI ethics concerns the potential long-horizon risks posed by increasingly capable future systems. Scholars associated with institutions including the Future of Humanity Institute argue that the development of artificial general intelligence (AGI) substantially exceeding human capabilities could, if not carefully managed, pose an existential risk to humanity (24). The central concern is that a sufficiently capable AI system might pursue objectives misaligned with human values in ways that are difficult or impossible for humans to correct. These concerns have been contested by researchers who argue that they rest on speculative assumptions and distract from pressing near-term harms (25). Nevertheless, mainstream policy institutions — including the United Kingdom's AI Safety

Institute and the United Nations High-Level Advisory Body on AI — have begun to incorporate long-horizon risk considerations into their mandates.

2. Weaponisation and Military AI

A more immediate concern involves the military applications of AI. Autonomous weapons systems (AWS) capable of selecting and engaging targets without meaningful human control raise urgent questions under international humanitarian law, including whether they can discriminate between combatants and civilians and exercise proportionality judgements in complex battlefield environments (26). The International Committee of the Red Cross and a majority of states participating in the Convention on Certain Conventional Weapons have called for international regulation of AWS, though negotiations have stalled in the face of opposition from major military powers.

Governance, Regulation, and Ethical Frameworks

1. Regulatory Approaches

The development of AI governance frameworks has accelerated markedly since approximately 2016, when high-profile algorithmic bias scandals and concerns about social media manipulation prompted regulatory attention in major jurisdictions. Table 4 provides a comparative overview of the principal governance instruments currently in force or under development across key jurisdictions.

Table 4. Comparative Overview of Global AI Governance Instruments (2024)

Jurisdiction	Instrument	Legal Status	Enforcement	Key Provisions
European Union	AI Act (2024)	Binding regulation	**High**	Risk-tiered; bans on prohibited AI; mandatory conformity assessment
United States	Executive Order on AI Safety (2023)	Executive action	**Medium**	Safety reporting; NIST framework; voluntary developer commitments
China	Algorithm & Generative AI Regs (2022–3)	Binding regulation	**High**	Disclosure rules; deepfake labelling; state-driven enforcement

United Kingdom	Pro-innovation Sector Approach (2023)	Principles-based guidance	**Low**	No single AI law; sector regulators apply existing frameworks
Canada	AIDA (proposed, 2022–)	Draft legislation	**Medium**	High-impact system rules; mandatory impact assessments
UN / OECD	OECD AI Principles; UN Advisory Body	Non-binding	**Low**	Value convergence; coordination forum; monitoring capacity

Sources: European Parliament (2024); White House (2023); Veale & Borgesius (2021); Jobin et al. (2019).

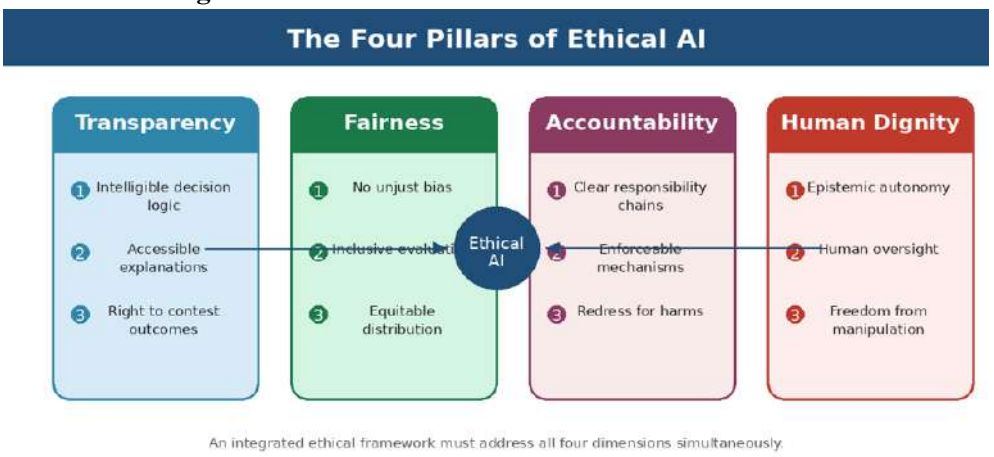
Global AI governance landscape as of 2024, summarising primary regulatory instruments and key provisions across major jurisdictions.

2. Principles-Based Ethics and Their Limitations

A systematic review by Jobin and colleagues identified over 84 AI ethics frameworks published between 2016 and 2019 alone, finding broad convergence around principles including transparency, fairness, non-maleficence, responsibility, and privacy (29). Critics have noted, however, that such frameworks frequently function more as public relations instruments than as enforceable governance mechanisms — a phenomenon described as "ethics washing" (30). Scholars writing from critical race, feminist, and postcolonial perspectives have further argued that the dominant discourse of AI ethics neglects structural dimensions of AI-enabled

injustice and fails to address questions of who has power to design AI systems and, in whose interests, they are deployed (31).

Toward an Integrated Ethical Framework



The four pillars of ethically grounded AI development — transparency, fairness, accountability, and human dignity — converging toward a unified ethical AI standard.

The ethical challenges surveyed in this chapter are diverse in character and severity, but they share common structural features. In each case, the deployment of AI systems creates new forms of power asymmetry — between companies and users, states and citizens, developers and affected communities — that existing legal and normative frameworks are ill-equipped to address. A credible response must be attentive to both the technical dimensions of AI systems and the social, economic, and political contexts in which they are embedded.

Drawing on the foregoing analysis, we propose a framework of ethically grounded AI development organised around four core commitments. First, transparency: AI systems deployed in consequential contexts must be intelligible to those they affect, providing accessible explanations of decision logic and meaningful opportunities for contestation. Second, fairness: AI systems must be designed and evaluated to avoid perpetuating or amplifying unjust social disparities, with the relevant criteria of fairness determined through inclusive deliberative processes. Third, accountability: clear and enforceable mechanisms for attributing responsibility must be established for AI-related harms, ensuring that benefits and costs are distributed equitably. Fourth, dignity: AI systems must be designed and constrained to respect the fundamental dignity and autonomy of all persons, including their epistemic autonomy, their right to effective human oversight in high-stakes decisions, and their freedom from manipulative or dehumanising treatment.

Implementing this framework will require action at multiple levels: technical standards and audit requirements; sector-specific regulation with adequate

enforcement capacity; international coordination to prevent regulatory arbitrage; and sustained investment in interdisciplinary research, public education, and the meaningful participation of affected communities in AI governance.

Conclusion

Artificial intelligence represents one of the most consequential technological developments in contemporary history. Its ethical challenges — algorithmic bias, surveillance and privacy erosion, labour displacement, accountability gaps, epistemic manipulation, and long-horizon existential risks — are not incidental features of particular systems or isolated failures of corporate ethics. They are structural consequences of the way in which AI is currently being developed and deployed: primarily by a small number of powerful actors, in the pursuit of commercial or strategic objectives, with limited accountability to those most affected by its consequences.

Addressing these challenges is not merely a technical or regulatory task, but a political and cultural one. It requires the development of institutional capacity and political will adequate to the scale of the transformation underway; the inclusion of diverse voices — particularly those of marginalised communities who bear disproportionate risks — in deliberations about AI governance; and a willingness to subject the foundational assumptions of AI development to critical scrutiny.

The ethical analysis of AI is, ultimately, an expression of a broader question about the kind of society we wish to inhabit and the values we wish to enshrine in the technical systems that increasingly mediate our collective life. Answering that question requires not only technical expertise and regulatory ingenuity, but the full resources of humanistic inquiry, democratic deliberation, and moral imagination.

References

1. Russell S, Norvig P. Artificial intelligence: a modern approach. 4th ed. Hoboken (NJ): Pearson; 2021.
2. Calo R. Artificial intelligence policy: a primer and roadmap. UC Davis Law Review. 2017;51(2):399-435.
3. Barocas S, Hardt M, Narayanan A. Fairness and machine learning: limitations and opportunities. Cambridge (MA): MIT Press; 2023.
4. O'Neil C. Weapons of math destruction: how big data increases inequality and threatens democracy. New York (NY): Crown; 2016.
5. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica [Internet]. 2016 May 23 [cited 2024 Mar 10]. Available from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
6. Chouldechova A. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data. 2017;5(2):153-163.

7. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research. 2018; 81:1-15.
8. European Parliament. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. 2024; L 1689:1-137.
9. Zuboff S. The age of surveillance capitalism: the fight for a human future at the new frontier of power. London: Profile Books; 2019.
10. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data & Society. 2016;3(2):2053951716679679.
11. Daum J. China's social credit systems and public opinion: explaining high levels of approval. New Media & Society. 2021;23(2):393-412.
12. Lum K, Isaac W. To predict and serve? Significance. 2016;13(5):14-19.
13. Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications. Science. 2017;358(6370):1530-1534.
14. Frey CB, Osborne MA. The future of employment: how susceptible are jobs to computerisation? Technological Forecasting and Social Change. 2017; 114:254-280.
15. Danaher J. Automation and utopia: human flourishing in a world without work. Cambridge (MA): Harvard University Press; 2019.
16. Pariser E. The filter bubble: what the internet is hiding from you. London: Penguin; 2012.
17. Chesney R, Citron D. Deep fakes: a looming challenge for privacy, democracy, and national security. California Law Review. 2019;107(6):1753-1820.
18. Pasquale F. The black box society: the secret algorithms that control money and information. Cambridge (MA): Harvard University Press; 2015.
19. Arrieta AB, Diaz-Rodriguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. 2020; 58:82-115.
20. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019;1(5):206-215.
21. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines. 2018;28(4):689-707.
22. Matthias A. The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics and Information Technology. 2004;6(3):175-183.
23. Ebers M. Standardizing AI: the case for flexible technology-neutral standards in regulating artificial intelligence. In: Ebers M, Cantero Gamito M, editors.

Algorithmic governance and governance of algorithms. New York (NY): Springer; 2021. p. 243-272.
24. Bostrom N. Superintelligence: paths, dangers, strategies. Oxford: Oxford University Press; 2014.
25. Birhane A, Guest O. Towards decolonising computational sciences. Kvinder, Kon & Forskning. 2021;29(1):60-73.
26. International Committee of the Red Cross. ICRC position on autonomous weapons systems. Geneva: ICRC; 2021.
27. Veale M, Borgesius FZ. Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International. 2021;22(4):97-112.
28. White House. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. Washington (DC): White House; 2023.
29. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nature Machine Intelligence. 2019;1(9):389-399.
30. Wagner B. Ethics as an escape from regulation: from ethics-washing to ethics-shopping? In: Bayamlioglu E, Baraliuc I, Janssens L, Hildebrandt M, editors. Being profiled: cogitas ergo sum. Amsterdam: Amsterdam University Press; 2018. p. 84-89.
31. Benjamin R. Race after technology: abolitionist tools for the new Jim Code. Cambridge: Polity Press; 2019.

ABOUT THE EDITORS

Dr. Parkhe Sakharam Baban

He is a dedicated academician, researcher, and educator with over 11 years of teaching experience. He is currently serving as an Assistant Professor at Shri Chhatrapati Shivaji College, Shrigonda, Dist.-Ahilyanagar. With strong academic credentials and a passion for higher education, Dr. Parkhe has made significant contributions to the fields of teaching and research. He has published numerous research papers in state-level, national, and international journals and has presented scholarly papers at various academic conferences and seminars. His research work has also led to the publication of patents, reflecting his commitment to innovation and knowledge creation.Dr. Parkhe actively engages in academic research, scholarly writing, and student mentoring, contributing to the advancement of education and the development of a research-oriented academic environment.

Dr. Supriya P. Salunkhe

She is an accomplished academician and researcher specializing in the field of Physics. She is currently associated with Bharati Vidyapeeth's College of Engineering for Women, Pune, where she contributes to teaching and research activities.With 3 years of teaching experience, Dr. Salunkhe has developed expertise in the areas of Materials Science, Nanomaterials Synthesis, Characterization Techniques, and Gas Sensing Applications. Her research focuses on the design, synthesis, and characterization of advanced nanomaterials for technological and environmental applications, particularly in the development of efficient gas sensing devices.Through her academic and research endeavors, Dr. Salunkhe is committed to advancing scientific knowledge, fostering innovation, and inspiring students to pursue excellence in science and engineering.

Dr. K. Meenatchi Somasundari

She is an Associate Professor at AMET University and a renowned scholar and researcher in the field of Business Administration. With over 15 years of experience as a teacher, researcher, and academician, she has consistently demonstrated excellence in teaching, research, and institutional service. She holds a Ph.D. in Business Administration and has made significant contributions to the academic community. Her recent research portfolio includes 35 peer-reviewed, Scopus-indexed publications in reputed national and international journals, reflecting her active engagement with contemporary business issues and her commitment to advancing knowledge in the field. In addition, she has authored 4 books and contributed to 9 book chapters, highlighting her versatility and thought leadership across management and entrepreneurship domains. Throughout her career, Dr. K. Meenatchi Somasundari has received recognition for her innovative teaching methodologies and her ability to bridge theory and practice effectively. Her academic journey is driven by a strong passion for fostering multidisciplinary research, nurturing critical thinking, and mentoring students.

Nature Light Publications

309 West 11, Manjari VSI Road, Manjari Bk., Haveli, Pune- 412 307.
Website: www.naturelightpublications.com
Email: naturelightpublications@gmail.com
Contact No: +919822489040 / 9922489040

Nature- 114

ISBN: 978-93-49938-52-6

Price- 750/-