LKP Academy

In Association with

Since 2023

Leilani Katie Publication and Press

Reg.No: UDYAM-TN-12-0099209

(Registered under MSME) - Recognized by Government of India.

International Conference on

Outstanding Research, Business Innovation and Technology (ORBIT – 2026)

30th June, 2026

Chief Editor

Dr.Suganya Bharathi.S

Assistant Professor,

School of Law,

SRM Institute of Science and Technology,

Kattankulathur, Chennai, Tamil Nadu, India.

Book Title: International Conference on Outstanding Research, Business Innovation and Technology (ORBIT – 2026)

Chief Editor: Dr.Suganya Bharathi.S

Edition: I

Book Size: B5 Size

ISBN: 978-93-6348-139-8

Publisher: Leilani Katie Publication and PRESS, Madurai, Tamil Nadu, India.

ABOUT THE LKP ACADEMY

LKP Academy is a reputed academic organization dedicated to promoting quality education, research excellence, and professional development. The academy actively supports academic initiatives such as international conferences, faculty development programs, research publications, academic networking, and student enrichment activities. Through its collaborations with institutions across India and abroad, LKP Academy provides meaningful platforms for knowledge sharing, career advancement, and global academic exposure, helping educators and learners grow in a competitive academic environment.

ABOUT THE LEILANI KATIE PUBLICATION AND PRESS

Leilani Katie Publication and PRESS (LKP) is a professional academic and educational publishing organization dedicated to supporting authors, academicians, researchers, educators, students, and professionals in bringing their knowledge and ideas to a wider audience. The organization publishes textbooks, reference books, research-oriented books, conference proceedings, and professional publications across diverse fields such as Computer Science, Information Technology, Artificial Intelligence, Data Science, Engineering, Management, Commerce, Economics, Science, Education, and other emerging disciplines. LKP focuses on providing quality-oriented and author-friendly publishing services, including manuscript support, editing, book formatting, cover design, ISBN-related assistance, printing, and digital publishing. Through its commitment to academic excellence, innovation, and knowledge dissemination, Leilani Katie Publication and PRESS aims to create meaningful publishing opportunities and contribute to the growth of education, research, and professional learning.

ABOUT THE CONFERENCE

The **International Conference on Outstanding Research, Business Innovation and Technology – ORBIT 2026** is a virtual international conference organized by **LKP Academy** in association with **Leilani Katie Publication and PRESS**. Scheduled for **30 June 2026**, the conference provides a dynamic platform for UG and PG students, research scholars, faculty members, industry professionals, and academicians to present their innovative ideas, research findings, and emerging technological solutions. The conference focuses on promoting outstanding research, business innovation, technological advancement, knowledge sharing, and collaboration among researchers and professionals from diverse disciplines. Participants will have the opportunity to submit and present research papers virtually, interact with researchers and experts, and explore publication opportunities. ORBIT 2026 aims to encourage innovation, strengthen research collaboration, and empower young minds to contribute meaningful ideas toward a better and technologically advanced future.

TABLE OF CONTENTS

AI-DRIVEN SMART CITIES: INTELLIGENT DISASTER RESPONSE AND URBAN RESILIENCE

ID: ORBIT – 002

Dr.Chandrasekaran Venkatesan
Assistant Professor,
Department of Marine Engineering,
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Abstract

In recent years, fast urbanization, variable climate, interrelated infrastructure systems, and increased number and frequency of extreme events compel the move from responsive disaster management to proactive and adaptive resilient planning [cite:1]. Most literature in smart city disaster resilience to date has focused on the enabling technologies and data systems. One review indicated that 72% of the components of smart city resilience considered in the review were based on information technology; clearly the reliance on sensing, analysis and digital coordination has been key to urban resilient governance [cite:1]. In parallel, the rise of AI, machine learning, IoT, Fog computing, predictive analytics are employed to enhance real-time awareness, risk prediction, response coordination and essential urban service continuity [cite:4].

This report discusses how AI-enabled smart city architecture could support intelligent disaster response and long-term urban resilience through combined sensing, edge intelligence, prediction and decision support [cite:2]. The report surveys the AI smart city literature, proposes a layered architecture for disaster-resilient urban system, elaborates on IoT plus Edge AI architecture that supports real-time operational and predicts urban disaster situations using the predictive analytical approaches on traffic, energy, environmental risk and emergency management [cite:6]. The chapter then discusses various implementation obstacles of AI in smart city disaster resilience, including privacy, interoperability, fragmented governance, digital inequity and model trustworthiness, and potential future directions in this field, such as digital twin, multi-modal sensing, locally situated framework and human-centered AI governance [cite:2].

Introduction

The era when smart cities were judged only on their ability to implement digitized infrastructure has come to an end. It now depends more on how well cities can withstand stresses and strains and recover from disruptions. [cite:10]. For city's disaster response capabilities it depends on the ability to integrate data, identify anomalies early, resource effectively and provide decision support to emergency responders operating under uncertainty [cite:4]. Traditional response systems tend to

ISBN: 978-93-6348-139-8

be fragmented, backwards-looking, and far too slow for cascading urban crisis events (e.g., flash flood, fire, heat waves, power outage, chemical spill, transport chaos or public health crisis). [cite:10].

AI allows us to transform these continuous flows of urban data into intelligence for action. [cite:6]. Models built using machine learning predict risk, computer vision systems evaluate damage from aerial imagery, systems built using natural language processing extract situational intelligence from social media and other reports by citizens, and optimization engines help improve emergency evacuation routes and distribution of critical resources. [cite:7]. When coupled with IoT-devices, command-and-control centers and communications infrastructure, they enables cities to move toward anticipatory, adaptive, evidence-based governance. [cite:2].

The focus on building smart-city resilience is particularly relevant in rapidly urbanizing areas which have a greater density of stressed infrastructure, climate vulnerabilities and people leading to exacerbated impacts during crises. [cite:11]. Research to date increasingly focuses on not only the technical capacities of a system but also on it being locally sensitive, ethical, context-specific and having local citizen engagement. [cite:11]. So a system building for smart city resilience should be seen more as a socio-technical framework which involves sensors, analytics, governance and human intelligence than as a simple technological system upgrade. [cite:14]

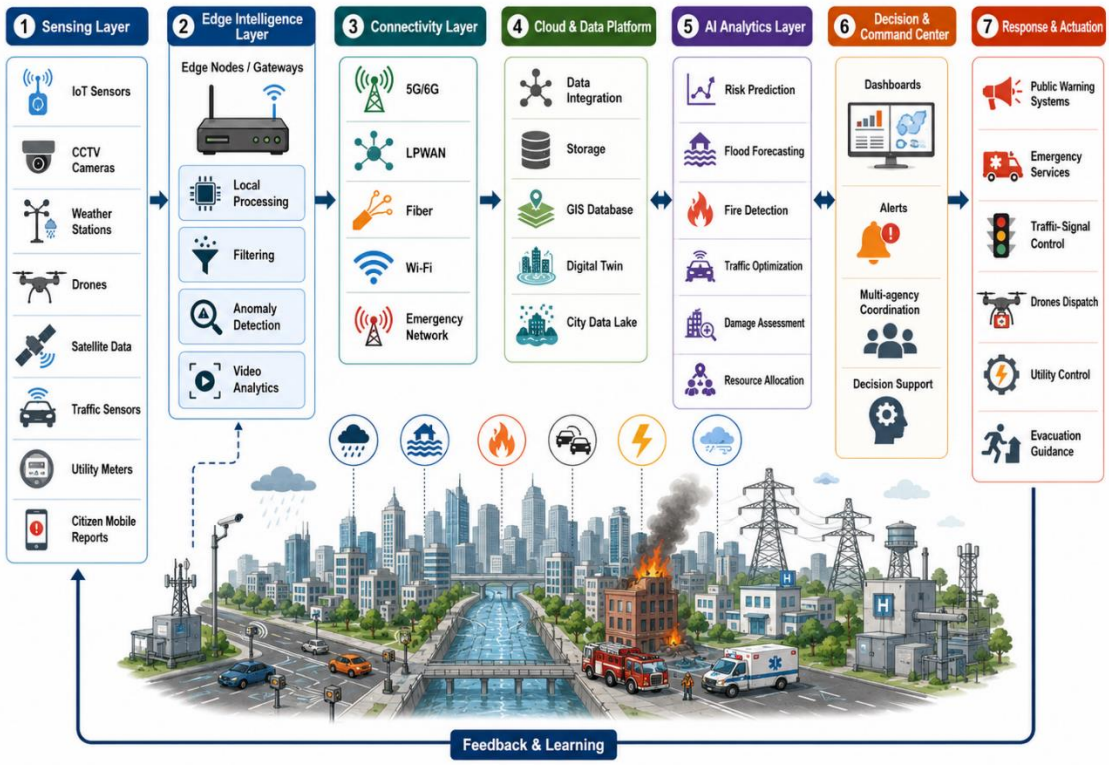


Fig. 1. Smart City Architecture for Disaster Response.

ISBN: 978-93-6348-139-8

Recent literature points toward AI driven smart cities increasingly transitioning from a framework built for service optimization to one which supports disaster resilience at the smart city level with capabilities like pre-event preparedness, rapid response during a crisis and efficient post-disaster recovery. The papers reviewed highlight the importance of IoT, big data analytics, edge computing, and machine learning in building these capabilities which span across different city infrastructures. They also show the need to integrate various disparate systems through interoperable frameworks to create more effective smart city systems which can combine physical infrastructures with emergency governance and citizen communication systems.

A core recurring theme across these research studies is the paradigm shift from a reactive to proactive approach of city governance in terms of disaster management. AI based approaches are continuously being developed to achieve the goal of predicting the risk of crisis, optimizing traffic flows during emergencies, detecting abnormalities in urban systems, and optimizing the resource allocation during disasters. However survey and review papers still consider technology as a piece of the whole picture and not as a singular solution that is needed to enable smart-city resilience.

Combining this information, it can be concluded that intelligent smart city resilience relies on both integrated sensing, analytics and distributed computing infrastructure as well as proper organizational support and control. It will form the backbone of any proposed AI-driven smart city framework which leverages IoT devices, an Edge AI architecture and predictive analytics for enhanced crisis response capability.

Literature Review (AI + Smart Cities)

The existing literature reflects that a clear synergy between AI, IoT, cloud-edge computing and data platforms exist in smart city development [cite:14]. Applications like adaptive traffic management, environmental monitoring, disaster prediction, better administration of civic services, upkeep of infrastructure, and improved decision support to commands centers are increasingly the focus of smart city studies [cite:10]. This trend is being driven by the maturation of smart city sensor networks and the need to manage diverse real-time data streams from various sources [cite:6].

One study has comprehensively reviewed smart city disaster resilience and highlighted how current literature is mainly technology-centered, emphasizing technologies, data sources, and digitally assisted collaboration; social and governance aspects are less structured [cite:1]. This is a critical oversight as disaster resilience not only necessitates detection and forecasting but also institutional preparedness, fairness of service provision, building community trust in the communicated information, and planning for recovery [cite:1]. Emerging research therefore aims at developing an integrated view which links technological systems with community needs, vulnerabilities at the city level and appropriate policy framework design [cite:15].

There are five distinct categories into which AI applications for smart cities can be classified [cite:10]. First, a perception system utilizes sources like cameras, remote sensing, drones, audio sensors and probes to sense the environment in real time

ISBN: 978-93-6348-139-8

[cite:14]. Second, a predictive system uses real-time and historical data for a variety of applications: it forecasts traffic congestion, predicts potential flooding levels, forecasts electricity demand and quality of air and anticipated service failures [cite:12]. Third, a decision support system translates the predictions from predictive systems into actionable strategies like allocating emergency services based on priority of the problem and planning of evacuation plans and distribution of electric load among city resources [cite:15]. Fourth, a citizen facing system employs a network of feedback and social media interactions to gauge response levels during any incident [cite:15]. Fifth, strategic systems are like urban digital twins that use models of cities to assess crisis scenarios and develop preparedness plans prior to any disaster [cite:5].

Edge/fog computing research is particularly relevant in disaster situations, due to its ability to enable localized processing and avoid dependence on the wider cloud network [cite:8]. Several studies show that edge computing systems can significantly improve flood forecasting, fire detection and real-time surveillance by placing processing units closer to sensors, hence decreasing response time and bandwidth consumption and improving service continuity during degraded network connectivity [cite:4]. Mobility literature is a second major area supporting the use of AI for disaster resilience. A recent review finds that AI integration into traffic systems through IoT not only provides emergency responders with appropriate services and reduce traffic delays up to 30% in the studied smart cities, but that smart city control of the mobility networks itself becomes an efficient support system during emergencies [cite:9]. A similar principle can be applied to smart city resource networks like those related to water distribution, waste management, and electric energy, where analytical tools can improve the timing of resource maintenance, prediction of demand and service irregularities. Finally, studies in the realm of governance have also shifted towards predictive and participative rather than solely descriptive systems [cite:15]. AI-enabled smart city governance has found to utilize big data analytics, IoT streams, sentiment analysis and real-time feedback mechanisms in building a more transparent, responsive and inclusive model of decision-making [cite:15]. The effectiveness of these systems, however, is predicated on adequate data quality and standardization, interoperability, and the wise combination of automated processes with human inputs rather than fully robotic systems .

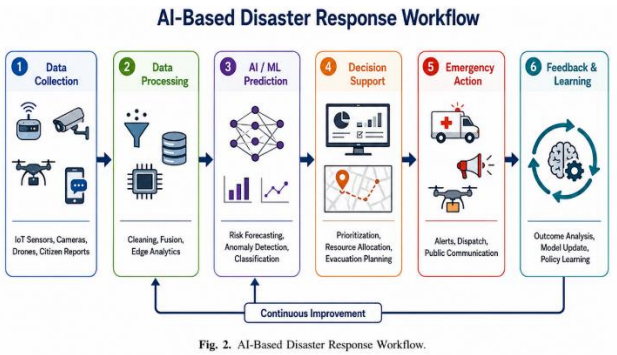


Fig. 2. AI-Based Disaster Response Workflow.

ISBN: 978-93-6348-139-8

Proposed Framework

An AI-driven smart-city disaster response framework should be realized as a multi-layer socio-technical system consisting of sensing layer, edge intelligence layer, communication layer, city data platform, analytical layer, command center decision support and field level actuation. [cite:15]. It is not only to collect abundant data, but the most important thing is to realize the "closed loop" from data collection to actuation based on analyzed and evaluated data and decisions. Signals are perceived, analyzed, prioritized, and put into operation within seconds or minutes of disaster occurrence, depending on its hazard .

The proposed framework is composed of six layers of functionalities:

- Sensing layer: consists of all the deployed sensors that include IOT, camera sensors, weather station, satellite streams, smart meters, structural health sensors, road sensors, wearable devices for field workers, city citizen mobile apps. All of these contribute to the collection of real-time conditions in the city [cite:14].
- Edge intelligence layer: Fog nodes, gateway nodes process data at the edge that includes filtering, event detection, analysis, anomaly identification and prioritization, video detection and so on in order to cut down latency when network traffic becomes very heavy [cite:8].
- Data integration layer: The diverse data streams are aggregated into the "urban operating picture" by adopting the standardized API, message brokers, geographic information system (GIS) databases and data models. [cite:11].
- AI analytics layer: uses forecasting models, risk estimators, image/computer vision algorithms, NLP (natural language processing) tools, optimization models in order to translate these data into prediction and operational recommendation [cite:15].
- Decision and coordination layer: integrated command and control center, emergency operations center, and departmental dashboards to support decision-making and command over the systems [cite:18].
- Response and recovery layer: comprises all actuating systems, warning system, traffic lights, drones, citizen alerting, command center, resource dispatching system, utility control and resource recovery system. This layer closes the loop [cite:15].

There are four principles upon which this framework is designed . First, anticipatory resilience: use the results of the AI analytical layer to act before conditions deteriorate. Second, distributed resilience: individual intelligence can run even when central system fails. Third, interoperable resilience: enabling the mobility, energy, health, and public safety systems to exchange trusted data. Fourth, human-centered resilience: ensuring expert judgments, public alerts and equitable services are still essential in the system [cite:15].

ISBN: 978-93-6348-139-8

The practical deployment path for the city should prioritize high-risk corridors (i.e. Flood prone areas, transportation hubs, industrial districts, dense urban areas) . Integrating data from the sensors, surveillance cameras, emergency communications system and citizens' reports onto a common platform and expand from there to develop digital twins and cross-domain optimization once the data quality and institutional cooperation has improved . This phased approach decreases the implementation risk while achieving improved preparation, faster detection speed and higher response effectiveness [cite:4].

IoT + Edge AI Architecture

IoT + Edge AI architecture is the operational core of smart disaster response since disasters require low-latency analysis, intermittent connectivity tolerance and geographically dispersed awareness [cite:8]. In a typical cloud-centric architecture, raw sensor streams will be pushed upwards for processing and may create communication bottlenecks, delays and dependence on communication links [cite:4]. Edge AI overcomes these weaknesses by moving some inference tasks closer to cameras, sensors, substations, traffic control nodes, neighborhood gateways .

In the device layer, heterogeneous sensors can collect various data including amount of rainfall, height of river, temperature, concentration of smoke, vibration, traffic density, quality of air, structural integrity of infrastructure [cite:6][cite:14]. Cameras and drones add visualization capability to the systems which is used to detect object, crowd, traffic-jammed roads and to estimate degree of damage [cite:1][cite:4]. Citizens also provide various reports through text messages, emails, and cell phone camera pictures and locations [cite:1][cite:15] which complements well the systems that collect information.

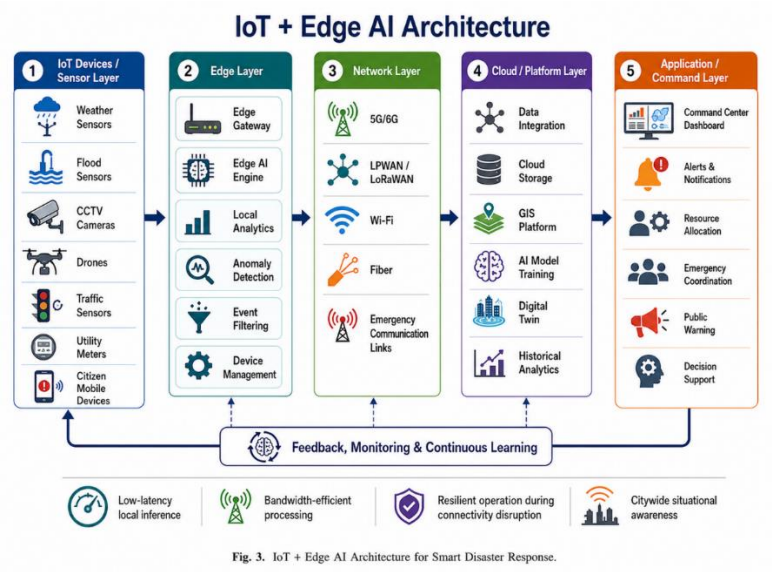


Fig. 3. IoT + Edge AI Architecture for Smart Disaster Response.

ISBN: 978-93-6348-139-8

In the edge layer, embedded processors or small data centers handle time-critical tasks such as filtering, event detection, compression, local thresholding, prioritization of data streams and using models or vision based mini-learning algorithms for inference [cite:8]. For example, the data collected at a road crossing in the flood management may be fed to an embedded processor or mini data center to identify if the road segment is safe, jammed or blocked without depending on network and server connection, so immediate warning to traffic may be given .

In the platform layer, selected aggregated summaries and alerts are pushed up to city management systems, cloud repositories and long term analysis systems . This hybrid model retains the visibility for strategists, while avoiding bandwidth and delay issues and offers federated scalability such that any region or department or a service provider would be able to scale itself while reporting the data to a city-wide information management system [cite:11].

Security and governance must be integral parts of the architecture . Device authenticity, communication encryption, role-based security policies, accountability and reliability must be guaranteed to ensure critical infrastructure safety and public trust . Since sensitive geographic location data, personal information about illness and surveillance data are involved, data privacy must be designed into the system, and data governance practices should be defined for the responsible use of the data [cite:2].

Predictive Analytics for Urban Systems

Turning historic and streaming urban data into predictive intelligence to prepare, prevent, and respond dynamically is the goal of predictive analytics . These predictions support timely response, damage reduction and operational efficiency by forecasting flood likelihood, traffic gridlock, ambulance calls, power surges, water pipe breaches, air pollution events, system failures, or areas of social vulnerability [cite:12]. Predicting when problems are likely to emerge and where-instead of waiting for a service breakdown to occur and for damage to propagate-allows planners and operators to strategically deploy resources .

In urban domains, several families of models are commonly combined depending on the specific problem . Time-series forecasting can be applied to predict demand (e.g., load in the electrical network) or environmental factors such as rainfall intensity or pollution levels [cite:6]. Classification algorithms predict whether an event will escalate into a danger, while clusters and geographical models provide insights into the distribution of risks within a given area, such as how vulnerable a specific neighborhood is, or where an incident is most likely to occur [cite:10]. Image processing can be used for damage evaluation, while natural language processing can facilitate real-time monitoring and interpretation of complaints from citizens, social media messages, or 911 call data .

A successful predictive workflow for urban systems comprises 5 distinct stages . First, data acquisition needs to gather all relevant data types from diverse sources such as sensor networks, government databases, weather feeds, geographical maps, and

ISBN: 978-93-6348-139-8

citizen contributions into a unified platform . Second, preprocessing steps, including imputing missing values, handling outliers, ensuring time synchronization, and enriching the datasets with useful features such as building types, land use and population density, need to be applied [cite:6]. Third, training and validation procedures for selecting and evaluating appropriate models against various criteria such as accuracy, robustness, fairness and transferability over time or different areas are necessary . Fourth, decision integration ensures that the predicted outcomes are used effectively to define risk scores, dispatch instructions, traffic management strategies, or alerts to the public . Fifth, learning from feedback after events occur ensures adaptive improvement of the predictive models [cite:15].

The impact of predictive analytics on urban resilience is demonstrated in multiple application areas. In mobility, AI-assisted decision systems are implemented to minimize transportation delays, ensure the unimpeded transit of emergency vehicles and optimize signal timing to avoid congestion [cite:9]. In utility services, such systems are applied to forecast peak demand for energy or water, to detect unusual patterns in consumption, or for maintenance purposes to prevent failures in water, waste or electricity infrastructure . With regard to disasters, incorporating multiple datasets such as geographical, meteorological and human-reported data can lead to better forecasts, triage of emergency events and resource allocation .

Despite these potential benefits, predictions are not a panacea for urban resilience . The predictive power of models can falter due to the inherent rarity of extreme events or data biases such as lack of data in certain communities; furthermore, institutions need to be able to receive, process and act upon predictions in a timely manner [cite:11]. For this reason, predictive systems should not only be tested and evaluated for their statistical accuracy, but also be thoroughly integrated into the operating procedures of the relevant agencies and be reviewed for their operational effectiveness, interpretability and equity impact .

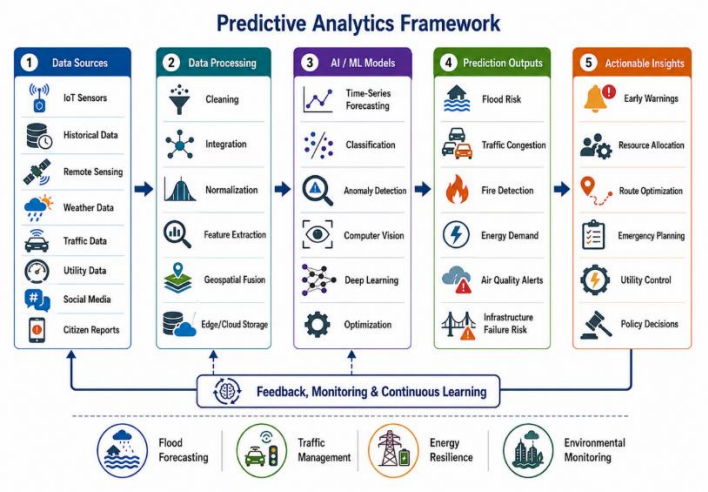


Fig. 4. Predictive Analytics Framework for Smart Urban Resilience.

ISBN: 978-93-6348-139-8

Challenges and Future scope

Interoperability among disparate urban systems is the first challenge. Transport, utility, policing, health and municipality services operate in their own islands with different data models and governance rules resulting in impossible situation awareness during disasters . In absence of standards and trusted exchange mechanisms, AI outputs can be hyper-local and unable to inform any coordinated resilience [cite:15].

Data governance, privacy and ethics are second challenges. Smart city resilience platforms will handle sensitive data generated by cameras, mobile devices and public service databases, raising the specter of surveillance overreach, exclusion, bias and abuse . This leads to necessity of ethical AI principles such as transparency, explainability, accountability, consent-aware data use, protection against discriminatory outcomes [cite:15].

Third is model robustness in the real disaster environments . Real-world hazards are often non-linear, cascading and context specific. Models built on normal operating parameters can perform poorly when faced with extreme events or during the failure of interdependent transport, water and power networks . This strongly justifies hybrid architectures, continuous learning, simulation based tests, and human-in-the-loop designs [cite:15].

Fourth is equity and territorial dimension . A growing body of research warns that tech centric approach to resilience might miss communities with poor access to communication networks, informal infrastructure, lower data visibility, though these communities might be more vulnerable to disasters . Therefore, smart city systems needs to take into account participatory design, local vulnerability mapping, and regionally grounded deployments, rather than assuming a universal solution [cite:11].

Future scope is significant . Urban digital twins will enable scenario-based simulations, such as evacuations, flood spread, interdependence of infrastructure, sequences of reconstruction efforts . AI driven decision support systems are expected to evolve into conversational and context aware agents, providing dynamic resource allocation, retrieval systems and dashboards to assist emergency responders in real-time [cite:7]. Fusion of different data sources like satellite images, remote sensing data, drones, and citizen reports will strengthen early warning and post-disaster damage assessment, particularly if paired with resilient communication networks and local governance.

Fig. 5. Disaster Response Command Center Dashboard.

ISBN: 978-93-6348-139-8

Conclusion

A credible path towards more rapid disaster response and improved urban resilience is offered by AI enabled smart cities through the operational integration of sensing, analytics and governance . Research suggests the integration of IoT, edge intelligence, predictive analytics and AI enabled decision support mechanisms would enable enhanced detection, enhanced situational awareness, improved mobility management and resource deployment not only in case of disaster events, but also in routine urban operations .

This report's proposed framework advocates that resilience cannot be achieved with only algorithms but depends on the enabling technologies, interoperable architecture, responsible data management, human control, community engagement and local context-based implementation of the developed solutions which take into account local disaster vulnerabilities and capabilities. AI will provide the greatest added value when we evolve smart city solutions from stand-alone applications to smart city resilience ecosystems, where its value proposition is more about supplementing civic discretion, transparent coordination, and enhancing operational continuity of services under the pre-, during-, and post-disaster phase .

Fig. 6. Urban Resilience Feedback Loop.

ISBN: 978-93-6348-139-8

Table 1: Comparison of AI Techniques in Disaster Management

AI Technique	Application Area	Key Function	Strengths	Limitations
Machine Learning (ML)	Flood, earthquake, and risk prediction	Pattern recognition from historical data	High accuracy in structured datasets	Needs large labeled datasets [2]
Deep Learning (DL)	Image-based disaster detection (drones, satellites)	Feature extraction from imagery	Excellent for complex visual data	High computational cost [1]
Computer Vision	Damage assessment, rescue mapping	Object detection in images/video	Fast real-time analysis	Sensitive to poor-quality data [3]
GIS + AI Integration	Spatial disaster mapping	Location-based risk analysis	Strong geospatial decision support	Data integration complexity [3]
Cloud AI Systems	Emergency coordination	Centralized data processing	Scalable and collaborative	Latency during network issues [4]
Human-AI Hybrid Systems	Decision-making in emergencies	Augmented decision support	Improves response accuracy	Requires training and trust [6]

Table 2: Smart City Technologies vs Applications

Smart City Technology	Core Function	Key Applications	Role in Urban Systems	Example Use Case
Internet of Things (IoT)	Real-time sensing and data collection	Traffic monitoring, pollution tracking, disaster alerts	Enables continuous urban data flow	Smart traffic sensors for congestion control [1]
Big Data Analytics	Data processing and pattern extraction	Predictive maintenance, urban planning, risk forecasting	Supports data-driven decision-making	Flood prediction from historical + live data [1]

ISBN: 978-93-6348-139-8

Artificial Intelligence (AI)	Intelligent decision-making	Emergency response, resource allocation, safety monitoring	Enhances automation and prediction	AI-based disaster response system [2]
Machine Learning (ML)	Predictive modeling	Disaster forecasting, demand prediction	Improves accuracy over time	Earthquake or flood risk prediction [3]
Cloud Computing	Scalable data storage and processing	Smart governance platforms, emergency coordination	Centralized control and integration	City-wide emergency command systems [4]
GIS (Geographic Information Systems)	Spatial mapping and analysis	Urban planning, evacuation routing	Location intelligence for decision-making	Disaster evacuation route optimization [5]
5G/Edge Computing	Low-latency communication	Real-time emergency alerts, autonomous systems	Enables fast response in crises	Drone-based disaster surveillance [4]

References

1. Author, A. A. (2025). *Leveraging artificial intelligence to enable sustainable urban resilience*. National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC12521657/
2. Author, A. A. (2024). *AI-driven innovation in smart city governance*. Emerald Insight. https://doi.org/10.1108/xxx
3. Author, A. A. (2025). *AI integration to strengthen disaster resilience in smart cities*. ResearchGate.
4. Author, A. A. (2024). *Integrating smart city technologies and urban resilience*. MDPI.
5. Author, A. A. (2024). *Application of image analytics for disaster response in smart cities*. ScholarSpace.
6. Author, A. A. (2022). *Sustainability-oriented innovations in smart cities*. ScienceDirect.
7. Author, A. A. (2024). *Emerging technologies and Industry 4.0–5.0 transition*. Open University Repository.

ISBN: 978-93-6348-139-8

ARTIFICIAL INTELLIGENCE AND STORYTELLING: REIMAGINING AUTHORSHIP AND CREATIVITY IN CONTEMPORARY LITERATURE

ID: ORBIT – 003

Dr. P Femina
Assistant Professor,
Department of English,
Mar Gregorios College of Arts & Science
Chennai, Tamil Nadu, India.

Ms. K Subhashini
Assistant Professor,
Department of English,
Mar Gregorios College of Arts & Science
Chennai, Tamil Nadu, India.

Abstract

The emergence of Artificial Intelligence (AI) has significantly transformed creative industries, challenging traditional concepts of authorship, creativity, and literary production. AI-powered tools are increasingly employed in writing, editing, translation, content generation, and publishing, creating new possibilities for literary creation and dissemination. This paper examines the impact of artificial intelligence on contemporary literature and explores how technological innovation is reshaping the understanding of authorship in the digital age. Drawing upon literary theory, digital humanities, and media studies, the study investigates the relationship between human creativity and machine-generated texts. Through an analysis of Kazuo Ishiguro's *Klara and the Sun* (2021), the paper argues that AI does not merely function as a technological tool but increasingly serves as a collaborative agent influencing narrative structures, language use, and literary production. The study further explores ethical concerns relating to originality, intellectual property, authenticity, and the future of creative writing. Ultimately, it contends that AI is transforming literary culture by encouraging scholars and writers to rethink the nature of creativity and authorship in the twenty-first century.

Keywords: Artificial Intelligence, Digital Humanities, Authorship, Creativity, Literary Innovation, Contemporary Literature

Technological innovation has always influenced literary production, from the invention of the printing press to the rise of digital publishing. The recent development of Artificial Intelligence marks another transformative moment in literary history. AI systems are now capable of generating poems, short stories, essays, and even novels,

ISBN: 978-93-6348-139-8

raising important questions about the nature of creativity and authorship. Storytelling. The growing integration of AI into creative practices has sparked debates among writers, scholars, and publishers regarding the role of human imagination in literary production. This paper explores how AI challenges traditional literary assumptions and examines its implications for contemporary literature.

The relationship between literature and technology has always been dynamic and transformative. From the invention of the printing press to the rise of digital publishing platforms, technological innovations have continuously influenced how literary texts are produced, distributed, and consumed. The emergence of Artificial Intelligence represents the latest and perhaps most significant development in this ongoing evolution. Unlike previous technologies that primarily facilitated literary production, AI can generate content autonomously, thereby challenging traditional assumptions about creativity, authorship, and artistic agency.

Recent advancements in machine learning and natural language processing have enabled AI systems to produce poems, stories, essays, and even full-length novels that resemble human-authored texts. These developments have generated considerable scholarly interest because they blur the distinction between human creativity and computational generation. As Hayles argues, digital technologies are increasingly shaping not only how texts are created but also how they are interpreted and understood (Hayles, 2008).

The growing integration of AI into literary production has prompted important questions concerning originality, authenticity, and intellectual ownership. If an AI system generates a narrative, can it be considered creative? If a human author collaborates with an AI tool, who should be recognized as the author? These questions challenge longstanding literary traditions that associate creativity with human consciousness, imagination, and lived experience.

Contemporary literature has responded to these developments by exploring human-machine relationships through fictional narratives. Kazuo Ishiguro's *Klara and the Sun* (2021) offers a compelling examination of artificial intelligence, empathy, and consciousness through the perspective of an Artificial Friend named Klara. The novel illustrates how AI has become a subject of literary inquiry and a metaphor for broader questions regarding humanity, creativity, and technological dependence.

The rise of AI has contributed to the emergence of Digital Humanities, an interdisciplinary field that combines computational methods with literary scholarship. Franco Moretti's concept of "distant reading" demonstrates how digital tools enable scholars to analyze large bodies of texts in ways previously impossible through traditional close reading practices (Moretti, 2013). Consequently, AI is reshaping not only literary production but also literary criticism and research methodologies. Against this backdrop, this paper explores how artificial intelligence is transforming contemporary literature by reconfiguring notions of authorship and creativity. Through theoretical engagement with Barthes, Foucault, Hayles, and contemporary digital humanities scholars, as well as textual analysis of *Klara and the Sun*, the study

ISBN: 978-93-6348-139-8

investigates the implications of AI for literary production, publishing, criticism, and the future of storytelling.

The emergence of Artificial Intelligence has generated extensive scholarly discussion across disciplines, including computer science, media studies, philosophy, and literary studies. Within the humanities, researchers have increasingly examined how AI challenges traditional notions of authorship, creativity, and textual production. The growing ability of AI systems to generate human-like narratives has prompted scholars to revisit long-standing debates concerning originality, artistic agency, and the role of technology in literary creation.

One of the foundational theoretical interventions relevant to this discussion is Roland Barthes' concept of the "Death of the Author." Barthes argues that meaning does not originate exclusively from the author's intentions but emerges through the interaction between text and reader (Barthes 142). His argument destabilizes the notion of the author as the sole creator of meaning and opens possibilities for understanding literary production as a collaborative process. In the context of AI-generated texts, Barthes' theory becomes particularly significant because it allows scholars to question whether human intentionality is essential for literary meaning.

Traditional literary criticism has often regarded authorship as the expression of individual genius and creative autonomy. Romantic literary theory, in particular, celebrated the author as the originator of meaning and artistic innovation. However, poststructuralist theorists challenged this assumption. Roland Barthes famously declared the "death of the author," arguing that meaning emerges through readers rather than authorial intention (Barthes, 1977). Similarly, Michel Foucault conceptualized authorship as an institutional and cultural construct rather than a purely individual achievement (Foucault, 1977).

Franco Moretti's concept of "distant reading" provides another important framework for understanding the intersection of technology and literary studies. Moretti proposes the use of computational methods to analyze large bodies of literary texts, thereby expanding the scope of literary scholarship beyond individual works (Moretti 48). His approach demonstrates how digital technologies can complement traditional close reading practices and generate new forms of literary knowledge.

Margaret Boden's research on computational creativity offers valuable insights into debates surrounding AI-generated literature. Boden distinguishes between human creativity and computational creativity while acknowledging that machines can produce novel and meaningful outputs through algorithmic processes (Boden 11). Her work challenges assumptions that creativity is exclusively human and suggests that technological systems can participate in creative production in significant ways.

Recent scholarship has also focused on the ethical implications of AI-generated content. Bender et al. caution against uncritical acceptance of AI-generated language, arguing that large language models reproduce patterns from existing datasets and may perpetuate social biases (Bender et al. 610). Their critique underscores the importance

ISBN: 978-93-6348-139-8

of examining questions of representation, authenticity, and responsibility in AI-assisted literary production.

Within contemporary fiction, AI has become a recurring theme through which authors explore issues of consciousness, identity, and humanity. Kazuo Ishiguro's *Klara and the Sun* exemplify this trend by presenting an artificial being whose emotional sensitivity and observational capabilities challenge conventional distinctions between humans and machines. Literary critics have interpreted the novel as an exploration of empathy, technological dependence, and the future of human relationships in an increasingly automated world. Despite growing scholarly attention, significant gaps remain in understanding how AI reshapes concepts of authorship and creativity within literary studies. This paper addresses these gaps by examining AI as both a literary subject and a technological participant in the production of narratives.

The rise of AI extends these theoretical debates by introducing non-human contributors into the writing process. Contemporary writers increasingly utilize AI tools for brainstorming, drafting, editing, and stylistic experimentation. As a result, authorship becomes distributed across human and technological agents. *Klara and the Sun* reflects these concerns by portraying an artificial being capable of observation, interpretation, and emotional engagement. Although Klara does not create literary texts, her narrative perspective raises important questions regarding consciousness and agency. Ishiguro invites readers to consider whether qualities traditionally associated with humanity can emerge within artificial systems.

AI-assisted writing further challenges conventional ideas of originality. Since AI generates content by identifying patterns within existing datasets, creative production becomes a collaborative process involving human prompts, algorithmic responses, and editorial decisions. This shift suggests that authorship in the digital age may be better understood as a networked and participatory practice rather than an individual accomplishment (Manovich, 2001).

Creativity has traditionally been viewed as a uniquely human capacity grounded in imagination, emotion, and subjective experience. AI-generated texts challenge this assumption by producing coherent narratives, stylistically sophisticated prose, and even poetry. Margaret Boden distinguishes between psychological creativity and computational creativity, arguing that machines can generate novel combinations without possessing conscious intentionality (Boden, 2004). While AI systems lack emotions and lived experiences, they demonstrate remarkable capabilities for generating innovative text. The debate concerning AI creativity is reflected in contemporary literature. In *Klara and the Sun*, Ishiguro explores the possibility that artificial beings can develop empathy, curiosity, and interpretive abilities. Klara's observations of human behavior reveal a nuanced understanding of emotions, suggesting that creativity and intelligence may not be exclusively human attributes.

Nevertheless, human creativity differs fundamentally from machine generation. Human writers draw upon memory, emotion, cultural experience, and ethical judgment. AI, by contrast, relies upon statistical prediction and data-driven pattern recognition.

ISBN: 978-93-6348-139-8

Therefore, rather than replacing human creativity, AI may function as an augmentative tool that expands creative possibilities while remaining dependent upon human direction and interpretation.

Artificial Intelligence has transformed publishing through automation and data analytics. AI-assisted editing software improves grammar, style, and readability, while machine translation expands access to global literary markets. Publishers increasingly employ predictive algorithms to analyze reader preferences and market trends. Recommendation systems influence reading habits by personalizing content based on user behavior. These developments have altered traditional relationships among authors, publishers, and readers.

Moreover, AI-generated content has created new forms of literary production. Experimental projects involving machine-generated poetry and collaborative storytelling demonstrate how technology can function as a creative partner. Such innovations challenge established publishing models and encourage reconsideration of literary value in the digital age. Digital publishing platforms also democratize literary participation by enabling emerging writers to reach audiences without relying exclusively on traditional publishing institutions. Consequently, AI contributes to a more accessible and interconnected literary ecosystem.

The integration of AI into literature generates significant ethical concerns. One of the most pressing issues involves intellectual property. Since AI systems are trained on vast collections of existing texts, questions arise regarding ownership and attribution. Another concern relates to authenticity. Readers often value literature because it reflects human experiences and emotions. AI-generated texts challenge assumptions about authenticity by producing narratives without personal consciousness or lived reality. Bias presents an additional ethical challenge.

AI systems may reproduce cultural stereotypes embedded within training datasets. Consequently, scholars and developers must address issues of representation and fairness in AI-generated literature. Furthermore, concerns regarding employment and creative labor have emerged as AI becomes increasingly capable of performing tasks traditionally associated with writers, editors, and translators. These developments necessitate ethical frameworks that balance innovation with human creative rights.

Contemporary fiction increasingly explores the implications of AI for identity, consciousness, and creativity. *Klara and the Sun* serves as an important example because it presents artificial intelligence not merely as technology but as a participant in emotional and ethical relationships. Through Klara's perspective, Ishiguro examines themes of loneliness, empathy, mortality, and human connection. The novel reveals both the possibilities and limitations of artificial consciousness, encouraging readers to reflect on what distinguishes humans from machines.

Such literary representations function as cultural dialogues through which societies negotiate anxieties and aspirations concerning technological advancement. Literature provides a critical space for examining the ethical and philosophical

ISBN: 978-93-6348-139-8

consequences of AI before they become social realities. As Murray argues, storytelling remains a fundamental mechanism through which societies understand technological change and imagine future possibilities (Murray, 2017).

Artificial Intelligence is reshaping literary creation, publishing, criticism, and readership in unprecedented ways. By challenging traditional concepts of authorship and creativity, AI encourages scholars to rethink fundamental assumptions regarding literary production. While AI-generated texts demonstrate remarkable capabilities, they remain dependent upon human interpretation, ethical judgment, and cultural understanding.

The analysis demonstrates that AI simultaneously functions as a literary subject and a creative collaborator. Contemporary literature increasingly portrays artificial intelligence as a participant in human cultural and emotional experiences. Such representations reflect broader societal concerns regarding technological advancement and human identity. The analysis of *Klara and the Sun* illustrates how contemporary literature engages critically with questions surrounding artificial intelligence and human identity. Rather than replacing human creativity, AI functions as a catalyst for reimagining the relationship between technology and artistic expression. The future of literary studies will likely involve increasing collaboration between human and machine intelligence. As AI continues to evolve, literary scholars must develop new theoretical frameworks capable of addressing emerging questions regarding authorship, creativity, and ethical responsibility in the digital age.

Works Cited

1. Barthes, Roland. *Image-Music-Text*. Translated by Stephen Heath, Fontana Press, 1977.
2. Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2021, pp. 610–623.
3. Boden, Margaret A. *The Creative Mind: Myths and Mechanisms*. 2nd ed., Routledge, 2004.
4. Braidotti, Rosi. *The Posthuman*. Polity Press, 2013.
5. Foucault, Michel. "What Is an Author?" *Language, Counter-Memory, Practice: Selected Essays and Interviews*, edited by Donald F. Bouchard, translated by Donald F. Bouchard and Sherry Simon, Cornell UP, 1977, pp. 113–138.
6. Hayles, N. Katherine. *Electronic Literature: New Horizons for the Literary*. U of Notre Dame P, 2008.
7. Hayles, N. Katherine. *How We Think: Digital Media and Contemporary Technogenesis*. U of Chicago P, 2012.
8. Ishiguro, Kazuo. *Klara and the Sun*. Faber and Faber, 2021.
9. Jenkins, Henry. *Convergence Culture: Where Old and New Media Collide*. New York UP, 2006.

ISBN: 978-93-6348-139-8

10. Manovich, Lev. *The Language of New Media*. MIT Press, 2001.
11. Moretti, Franco. *Distant Reading*. Verso, 2013.
12. Murray, Janet H. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Updated ed., MIT Press, 2017.
13. Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed., Pearson, 2021.
14. Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2011.
15. Turkle, Sherry. *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster, 1995.
16. Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. U of Chicago P, 2019.
17. Vermeulen, Pieter. "Contemporary Literature and the End of the Human." *Journal of Modern Literature*, vol. 43, no. 1, 2019, pp. 1–17.
18. Wardrip-Fruin, Noah. *Expressive Processing: Digital Fictions, Computer Games, and Software Studies*. MIT Press, 2009.
19. Wolf, Werner. "Artificial Intelligence and Literary Creativity: New Perspectives on Authorship." *Poetics Today*, vol. 42, no. 3, 2021, pp. 401–420.

ISBN: 978-93-6348-139-8

DIGITAL NARRATIVES AND IDENTITY FORMATION: EXPLORING TECHNOLOGY, INNOVATION, AND ADOLESCENT SELFHOOD IN CONTEMPORARY YOUNG ADULT LITERATURE

ID: ORBIT – 004

Ms. K Subhashini
Assistant Professor,
Department of English,
Mar Gregorios College of Arts & Science,
Chennai, Tamil Nadu, India.

Dr. P Femina
Assistant Professor,
Department of English,
Mar Gregorios College of Arts & Science,
Chennai, Tamil Nadu, India.

Abstract

The rapid advancement of digital technologies has transformed the ways individuals construct, negotiate, and express their identities. Contemporary Young Adult (YA) literature increasingly reflects this technological shift by portraying adolescents who navigate social media platforms, virtual communities, online relationships, and digital cultures. This paper examines how digital narratives in contemporary YA literature represent the complex relationship between technology and identity formation. Drawing upon theories of digital humanities, identity studies, and adolescent psychology, the study explores selected YA texts that foreground digital engagement as a significant factor in shaping self-perception and social interaction. The paper argues that contemporary YA literature functions as a cultural archive documenting the opportunities and challenges presented by technological innovation. While digital spaces facilitate self-expression, community building, and empowerment, they also generate concerns related to surveillance, cyberbullying, self-objectification, and identity fragmentation. This paper examines the relationship between technology, innovation, and identity formation through an analysis of Francesca Zappia's *Eliza and Her Monsters* (2017). Drawing upon theories of digital identity, adolescent psychology, and media studies, the study explores how digital spaces function as sites of self-expression, creativity, empowerment, and social connection. Simultaneously, the paper investigates the challenges associated with online engagement, including identity fragmentation, anxiety, cyber harassment, and the tension between virtual and real-world selves.

ISBN: 978-93-6348-139-8

Keywords: Young Adult Literature, Digital Identity, Technology, Innovation, Social Media, Adolescence

The twenty-first century has witnessed an unprecedented expansion of digital technologies, fundamentally altering how individuals communicate, interact, and perceive themselves. Smartphones, social networking sites, online gaming communities, and digital content platforms have become integral components of adolescents' everyday lives. Unlike previous generations, contemporary adolescents are often described as "digital natives" because they have grown up immersed in technologically mediated environments. Consequently, their experiences of identity formation are increasingly influenced by digital interactions and virtual spaces.

Identity formation is a crucial developmental process during adolescence, a period characterized by self-exploration, emotional growth, and the search for belonging. The digital environment offers adolescents new opportunities to experiment with multiple identities, construct online personas, and engage with diverse communities beyond geographical boundaries. At the same time, these virtual spaces expose young people to challenges such as cyberbullying, social comparison, privacy concerns, and the pressure to maintain idealized self-images. Thus, technology serves as both an empowering and a complicating force in the development of adolescent identity.

Contemporary Young Adult (YA) literature reflects these social transformations by portraying characters who navigate digital landscapes as part of their everyday reality. Unlike traditional coming-of-age narratives, modern YA texts increasingly incorporate social media interactions, online friendships, digital activism, virtual communities, and internet culture into their storylines. These narratives provide valuable insights into how adolescents negotiate questions of selfhood, agency, belonging, and authenticity in a digitally connected world. Digital narratives serve as important cultural documents that record the evolving relationship between technology and human experience. Through literary representation, authors explore the psychological, social, and emotional consequences of digital engagement, revealing how technological innovations influence adolescents' perceptions of identity, relationships, and social acceptance. Such narratives encourage readers to critically examine the opportunities and risks associated with digital participation.

The study of digital narratives is particularly relevant within contemporary literary scholarship because it highlights the intersection of literature, technology, and cultural studies. As digital technologies continue to shape social realities, literary texts provide a unique framework for understanding how technological innovation affects human subjectivity and self-representation. Examining these narratives contributes not only to literary criticism but also to broader interdisciplinary discussions concerning youth culture, digital citizenship, and identity construction in the modern era.

ISBN: 978-93-6348-139-8

Against this backdrop, the present study investigates how contemporary Young Adult literature portrays digital identity formation and examines the role of technology in shaping adolescent selfhood. By analyzing the representation of online interactions, social media engagement, and virtual communities, the paper seeks to demonstrate how literary narratives reflect and critique the complexities of adolescent life in the digital age. Francesca Zappia's *Eliza and Her Monsters* serves as an important example of digital-age YA fiction. The novel follows Eliza Mirk, an introverted teenager who secretly creates a highly successful webcomic called *Monstrous Sea*. While she struggles with social anxiety in her offline life, she enjoys confidence, recognition, and belonging through her online identity. Through Eliza's experiences, the novel explores how digital spaces shape adolescent selfhood and creativity.

The relationship between technology and identity has been widely examined within media studies and digital humanities. Sherry Turkle argues that digital spaces allow individuals to experiment with multiple identities and forms of self-representation (Turkle 180). Similarly, Manuel Castells emphasizes that identity construction increasingly occurs within networked societies where communication technologies influence social relationships and personal development (Castells 6).

Danah Boyd's research on networked publics highlights how adolescents use digital spaces to establish social connections, negotiate visibility, and construct identities (Boyd 8). These studies suggest that digital technologies have become central to adolescent experiences of belonging and self-expression.

Within literary studies, scholars have increasingly recognized YA literature as a significant site for examining contemporary social concerns. Michael Cart argues that YA literature reflects the realities and challenges faced by young people in changing social environments (Cart 21). However, relatively few studies have focused specifically on how contemporary YA novels represent digital identity formation. This paper seeks to contribute to this emerging area of research.

Technology has become a significant factor in the formation and expression of adolescent identity. According to Erik Erikson, adolescence is a critical stage characterized by the search for identity and self-definition (Erikson, 1968). In contemporary society, this developmental process increasingly occurs within digital environments where adolescents interact, communicate, and construct their self-images. Young Adult literature reflects this transformation by portraying protagonists who negotiate their identities through social media profiles, online communities, blogs, and digital interactions.

Digital platforms provide adolescents with opportunities to experiment with multiple identities and explore aspects of the self that may remain unexpressed in offline settings. Turkle argues that online spaces enable individuals to "play with identity" and engage in continuous self-representation (Turkle, 1995). Contemporary YA novels frequently depict characters who create digital personas that differ from their real-life identities, illustrating both the liberating and challenging aspects of online self-construction. Moreover, technology influences adolescents' perceptions of social

ISBN: 978-93-6348-139-8

acceptance and belonging. Social networking sites often function as spaces where validation is measured through likes, comments, followers, and online engagement. Boyd observes that networked publics shape how young people understand themselves and their relationships with others (Boyd, 2014). YA literature captures these dynamics by portraying protagonists who seek recognition and affirmation through digital interactions, revealing the growing significance of virtual spaces in identity formation.

One of the central themes in *Eliza and Her Monsters* is the construction of identity through digital spaces. Eliza maintains two distinct identities: her socially withdrawn offline self and her confident online persona as the anonymous creator of *Monstrous Sea*. The internet provides Eliza with a space where she can express herself freely without fear of social judgment. Through her webcomic, she develops a sense of purpose, achievement, and belonging. This reflects Turkle's assertion that online environments enable individuals to explore alternative identities and experiment with self-expression (Turkle 185).

The novel illustrates how digital platforms facilitate identity exploration during adolescence. Eliza's online presence allows her to overcome limitations associated with her social anxiety and develop a stronger sense of self. However, the coexistence of multiple identities also creates internal conflict, as she struggles to reconcile her online success with her offline insecurities. Zappia demonstrates that identity in the digital age is not fixed but fluid and multifaceted. Adolescents continuously negotiate between authentic self-expression and socially constructed personas, reflecting broader cultural shifts associated with digital technology.

Innovation, Social Media, and Adolescent Experiences

Technological innovation has revolutionized communication practices, making social media a central component of adolescent life. Platforms such as Instagram, TikTok, Snapchat, and YouTube have transformed how young people create, consume, and share information. These digital spaces facilitate connectivity and self-expression while simultaneously influencing social behavior and emotional well-being. Researchers suggest that social media functions as a powerful agent of socialization, shaping adolescents' attitudes, aspirations, and perceptions of reality (Livingstone & Brake, 2010). Contemporary YA literature frequently explores how social media affects friendships, romantic relationships, family dynamics, and personal development. Through literary narratives, authors examine the ways digital communication reshapes traditional interpersonal interactions.

Innovation has also enabled adolescents to participate in global conversations and social movements. Digital platforms empower young people to engage in activism, raise awareness regarding social issues, and connect with like-minded communities. Jenkins describes this phenomenon as participatory culture, where users become active contributors rather than passive consumers of media (Jenkins, 2006). YA novels increasingly portray digitally engaged protagonists who utilize technology as a tool for advocacy and social change.

ISBN: 978-93-6348-139-8

The novel highlights how technological innovation has transformed creative production and social interaction. Eliza's webcomic represents a form of digital storytelling that enables young creators to reach global audiences without relying on traditional publishing institutions. Henry Jenkins describes such practices as participatory culture, where individuals actively create and share media rather than simply consuming it (Jenkins 3). Eliza's success as a webcomic creator demonstrates how digital technologies democratize creative expression and empower young voices.

The novel also explores the role of online communities in adolescent life. Through fan forums, social media interactions, and digital networks, Eliza forms meaningful relationships with individuals who share her interests. These communities provide emotional support and foster a sense of belonging. At the same time, the novel reveals the pressures associated with digital visibility. The expectations of fans and the demands of maintaining an online presence contribute to stress and anxiety. Such experiences reflect contemporary concerns regarding the psychological impact of constant connectivity. However, innovation also introduces pressures associated with constant connectivity. The expectation to remain active and visible online may contribute to anxiety, stress, and emotional exhaustion. Literary representations of these experiences reveal the complex relationship between technological advancement and adolescent well-being.

Digital environments can function as empowering spaces where adolescents develop confidence, agency, and self-expression. For many young people, especially those belonging to marginalized communities, online platforms provide opportunities to share experiences, seek support, and construct positive identities. Castells argues that digital networks facilitate new forms of identity construction and collective action by enabling individuals to connect across geographical and cultural boundaries (Castells, 2010). In YA literature, digital communities often serve as safe spaces where protagonists find acceptance, friendship, and validation that may be absent in their offline lives.

Technology also promotes creative expression through blogging, content creation, digital storytelling, and artistic production. Such opportunities allow adolescents to articulate their thoughts and experiences while developing a sense of agency. Contemporary YA narratives frequently depict protagonists using technology to challenge stereotypes, confront discrimination, and advocate for social justice. One of the most significant contributions of digital culture is its ability to empower marginalized individuals. For Eliza, the internet functions as a safe space where she can develop confidence and pursue her creative ambitions.

Castells argues that digital networks facilitate new forms of empowerment by enabling individuals to connect with supportive communities and participate in collective activities (Castells 12). Eliza's experiences exemplify this phenomenon. Through her online identity, she gains recognition, develops friendships, and discovers a sense of purpose.

ISBN: 978-93-6348-139-8

The novel also demonstrates how digital creativity can serve as a form of self-affirmation. Creating *Monstrous Sea* allows Eliza to transform her personal struggles into artistic expression. The webcomic becomes a medium through which she articulates emotions and experiences that she finds difficult to communicate in face-to-face interactions. Furthermore, the novel highlights the potential of technology to foster resilience. Despite challenges and setbacks, Eliza ultimately learns to integrate her online and offline identities, achieving greater self-acceptance and emotional growth. Digital platforms encourage educational and intellectual empowerment by providing access to information and learning resources. Through these representations, YA literature highlights the transformative potential of technology in fostering resilience, self-discovery, and personal growth. Despite its empowering potential, digital culture presents significant challenges for adolescents. One of the most prominent concerns is cyberbullying, which has emerged as a widespread issue in technologically mediated societies. Unlike traditional forms of bullying, cyberbullying can occur continuously and reach large audiences, intensifying its psychological impact on young people (Patchin & Hinduja, 2010).

Another challenge involves the cultivation of unrealistic standards of beauty, success, and popularity. Social media platforms often encourage users to present idealized versions of themselves, leading to social comparison and diminished self-esteem. According to Festinger's Social Comparison Theory, individuals evaluate themselves by comparing their achievements and appearance to those of others (Festinger, 1954). YA literature frequently portrays protagonists struggling with body image concerns, self-worth, and feelings of inadequacy resulting from digital comparisons.

Privacy and surveillance also represent critical concerns within digital culture. Adolescents often share personal information online without fully understanding the long-term implications of their digital footprints. Foucault's concept of surveillance provides a useful framework for understanding how individuals modify their behavior when they perceive themselves as constantly observed (Foucault, 1977). Many YA novels depict characters navigating issues of online privacy, data exposure, and digital vulnerability.

Additionally, excessive engagement with digital technologies may contribute to feelings of isolation, addiction, and emotional dependency. Turkle warns that technology can create an illusion of connection while simultaneously reducing meaningful interpersonal communication (Turkle, 2011). Contemporary YA literature reflects these concerns by illustrating the emotional consequences of overreliance on digital interaction. Through these representations, YA literature serves as a critical lens for examining the complexities of digital culture and its influence on adolescent identity, relationships, and psychological well-being.

Although digital spaces provide opportunities for empowerment, the novel also portrays their limitations and risks. One major challenge involves the pressure associated with online visibility. As Eliza's identity becomes increasingly connected to

ISBN: 978-93-6348-139-8

her webcomic, she experiences anxiety regarding audience expectations and public scrutiny. The novel further explores issues of privacy and anonymity. Eliza's fear of revealing her identity reflects broader concerns regarding digital exposure and surveillance. The internet offers opportunities for self-expression, but it also increases vulnerability to criticism and intrusion.

Another challenge concerns emotional dependency on online validation. Eliza's self-esteem becomes closely linked to the success of her webcomic, illustrating how digital platforms can influence psychological well-being. This reflects Turkle's observation that technology can create both connection and isolation simultaneously (Turkle 280). By portraying these challenges, Zappia encourages readers to critically examine the role of technology in their lives and recognize the importance of balancing digital engagement with authentic interpersonal relationships.

Works Cited

1. Boyd, danah. *It's Complicated: The Social Lives of Networked Teens*. Yale University Press, 2014.
2. Castells, Manuel. *The Power of Identity*. 2nd ed., Wiley-Blackwell, 2010.
3. Erikson, Erik H. *Identity: Youth and Crisis*. W. W. Norton & Company, 1968.
4. Festinger, Leon. "A Theory of Social Comparison Processes." *Human Relations*, vol. 7, no. 2, 1954, pp. 117–140.
5. Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan, Vintage Books, 1977.
6. Jenkins, Henry. *Convergence Culture: Where Old and New Media Collide*. New York University Press, 2006.
7. Livingstone, Sonia, and David Brake. "On the Rapid Rise of Social Networking Sites: New Findings and Policy Implications." *Children & Society*, vol. 24, no. 1, 2010, pp. 75–83.
8. Patchin, Justin W., and Sameer Hinduja. "Cyberbullying and Self-Esteem." *Journal of School Health*, vol. 80, no. 12, 2010, pp. 614–621.
9. Turkle, Sherry. *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster, 1995.
10. Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2011.

ISBN: 978-93-6348-139-8

THE IMPACT OF GENERATIVE AI ON BUSINESS EDUCATION AND COMMERCE CURRICULUM: OPPORTUNITIES, CHALLENGES AND FUTURE DIRECTIONS

ID: ORBIT – 006

Dr. S Shalini
Dean of Academics,
Assistant Professor,
Mar Gregorios College of Arts and Science,
Chennai, Tamil Nadu, India.

Abstract

The emergence of Generative Artificial Intelligence (GenAI) has transformed various sectors, including education, business, healthcare, and finance. In higher education, particularly in business and commerce disciplines, Generative AI tools such as ChatGPT, Gemini, Copilot, and Claude are reshaping teaching, learning, assessment, and research practices. These technologies offer opportunities to enhance personalized learning, improve productivity, foster critical thinking, and prepare students for an AI-driven workplace. However, their integration also raises concerns regarding academic integrity, ethical use, data privacy, overreliance on technology, and curriculum relevance. This paper examines the impact of Generative AI on business education and commerce curricula through a review of contemporary literature. It explores the opportunities and challenges associated with AI adoption and proposes future directions for curriculum redesign and pedagogical innovation. The study concludes that while Generative AI presents transformative possibilities, its effective integration requires balanced policies, ethical frameworks, and AI literacy among educators and learners.

Keywords: Generative AI, Business Education, Commerce Curriculum, Higher Education, Artificial Intelligence, Digital Learning, Curriculum Innovation

The twenty-first century has witnessed unprecedented technological advancements that have transformed the way individuals learn, communicate, and conduct business. Among these innovations, Artificial Intelligence (AI) has emerged as one of the most influential technologies, reshaping industries and redefining professional practices. Recent developments in Generative Artificial Intelligence (GenAI), particularly large language models such as ChatGPT, Gemini, Claude, and Microsoft Copilot, have accelerated the integration of AI into educational and business environments. Unlike traditional AI systems that primarily analyze data and automate routine tasks, Generative AI can create human-like text, images, code, reports, and other forms of content, making it a powerful tool for knowledge creation and dissemination.

ISBN: 978-93-6348-139-8

As organizations embrace AI-driven decision-making and automation, business schools and commerce departments face the challenge of preparing graduates with relevant skills and competencies. Consequently, the integration of Generative AI into commerce curricula has become both a necessity and an opportunity.

The rapid adoption of Generative AI has sparked significant interest among educators, researchers, policymakers, and business leaders. Educational institutions worldwide are exploring the potential of AI-powered technologies to enhance teaching effectiveness, improve student engagement, and personalize learning experiences. In higher education, Generative AI has emerged as a valuable resource for supporting academic writing, research activities, content development, assessment design, and collaborative learning. Students increasingly rely on AI tools to clarify concepts, generate ideas, summarize complex information, and improve communication skills.

Business education and commerce programs are particularly affected by the rise of Generative AI because the business world itself is undergoing substantial digital transformation. Modern organizations increasingly utilize AI-driven systems for customer relationship management, financial forecasting, marketing analytics, human resource management, supply chain optimization, and strategic decision-making. Consequently, graduates entering the workforce are expected to possess not only traditional business knowledge but also digital competencies and AI literacy. This shift necessitates a re-evaluation of existing commerce curricula to ensure alignment with industry expectations and future workplace requirements.

The integration of Generative AI into business education presents numerous opportunities. AI-powered tools can facilitate personalized learning by adapting instructional content to individual student needs and learning preferences. Educators can use these technologies to develop teaching materials, create interactive learning experiences, and provide timely feedback. Furthermore, AI enables students to engage with real-world business scenarios, analyze large datasets, and explore innovative solutions to complex organizational challenges. Such capabilities have the potential to enhance critical thinking, creativity, problem-solving, and decision-making skills, which are essential attributes for future business professionals.

The emergence of Generative AI has therefore initiated a paradigm shift in educational philosophy and curriculum design. Traditional approaches that emphasize memorization and information recall are gradually becoming less relevant in an environment where information can be generated instantly. Instead, there is a growing emphasis on higher-order cognitive skills, including analytical thinking, creativity, adaptability, and ethical decision-making. Business schools and commerce departments must rethink their pedagogical strategies and assessment methods to prepare students for an increasingly AI-driven economy.

This study adopts a qualitative and descriptive research design based on secondary sources of data. Relevant information has been collected from peer-reviewed journal articles, conference proceedings, books, policy reports, and publications from educational institutions and international organizations. The study employs a thematic

ISBN: 978-93-6348-139-8

review approach to analyze emerging trends and perspectives concerning Generative AI in business education.

Generative AI has rapidly become an influential educational technology capable of transforming classroom practices. In business and commerce disciplines, AI tools assist students in performing various academic tasks, including report writing, market analysis, business communication, financial interpretation, and data visualization. The growing use of AI in industry has also increased the demand for graduates who possess AI literacy and digital competencies. Educational institutions are therefore exploring ways to incorporate AI-based learning experiences into commerce and management programs. Generative AI supports both educators and learners by automating routine tasks, facilitating content creation, and enabling more personalized learning environments. These developments have significant implications for curriculum design, assessment methods, and pedagogical strategies.

Generative AI enables personalized learning by adapting content to individual student needs, learning styles, and levels of understanding. Students can receive instant explanations, summaries, examples, and feedback tailored to their academic requirements. Educators can utilize AI tools to prepare lesson plans, develop teaching materials, generate assessment questions, and create case studies. This reduces administrative workload and allows instructors to focus on student engagement and mentoring.

One of the most significant contributions of Generative Artificial Intelligence to business education is its ability to foster innovation and creativity among students and educators. In today's rapidly changing business environment, organizations increasingly value employees who can generate innovative ideas, solve complex problems, and adapt to emerging challenges. Generative AI serves as a powerful tool that supports these competencies by enabling users to explore new perspectives, generate creative solutions, and engage in collaborative knowledge creation.

In business and commerce education, creativity is essential for developing effective marketing campaigns, designing business strategies, identifying entrepreneurial opportunities, and solving organizational problems. Generative AI tools assist students by providing multiple approaches to a given problem, helping them think beyond conventional solutions. For example, marketing students can use AI to generate advertising slogans, promotional content, social media campaigns, and customer engagement strategies. This exposure to diverse ideas enhances their creative thinking abilities and encourages experimentation.

Generative AI also promotes innovation in entrepreneurship education. Students aspiring to become entrepreneurs can utilize AI tools to develop business plans, identify market opportunities, conduct competitor analyses, and generate product or service ideas. By interacting with AI systems, learners can rapidly evaluate different business concepts and refine their entrepreneurial strategies. This process supports innovation by reducing the time required for idea generation and enabling students to focus on critical evaluation and strategic planning.

ISBN: 978-93-6348-139-8

Generative AI can also facilitate interdisciplinary learning by integrating concepts from commerce, technology, economics, and management. Students can explore how emerging technologies influence business operations and develop innovative solutions that combine knowledge from multiple disciplines. This interdisciplinary approach enhances creativity and prepares students for complex real-world challenges.

However, while AI can stimulate creativity, it should be viewed as a supportive tool rather than a replacement for human imagination. Genuine innovation emerges when students critically evaluate AI-generated ideas, modify them according to specific contexts, and contribute their own insights. Therefore, educators should encourage learners to use AI as a catalyst for creative thinking rather than relying solely on machine-generated outputs. By integrating AI responsibly into commerce curricula, educational institutions can equip students with the skills required to thrive in an increasingly technology-driven and innovation-oriented global economy.

One of the most significant concerns is the potential misuse of AI tools for completing assignments without genuine learning. The ease of generating essays and reports raises questions regarding plagiarism and academic honesty. Generative AI systems occasionally produce inaccurate or fabricated information, often referred to as "hallucinations." Students may unknowingly rely on incorrect content, leading to poor academic outcomes.

The use of AI involves the collection and processing of large volumes of data. Issues related to data privacy, intellectual property rights, and ethical use require careful consideration.

Excessive reliance on AI tools may weaken students' critical thinking, analytical reasoning, and problem-solving abilities if used without appropriate guidance.

Many educators lack sufficient training in AI technologies, creating challenges in effectively integrating AI into teaching practices and assessment frameworks. The emergence of Generative AI necessitates significant changes in commerce and business curricula. Traditional content-focused approaches must evolve to emphasize higher-order thinking skills, ethical reasoning, creativity, and technological competence.

Curriculum designers should consider:

1. Introducing AI literacy courses.
2. Incorporating AI applications in accounting, finance, marketing, and management.
3. Redesigning assessments to evaluate critical thinking and problem-solving.
4. Promoting ethical awareness regarding AI usage.
5. Encouraging experiential and project-based learning.

These measures can ensure that graduates remain competitive in an increasingly AI-driven economy. The growing influence of Generative Artificial Intelligence (GenAI) in higher education signals the need for significant reforms in business education and commerce curricula. As AI technologies continue to evolve, educational institutions

ISBN: 978-93-6348-139-8

must proactively adapt their teaching methodologies, curriculum structures, assessment practices, and policy frameworks. The future of business education lies not in replacing human intelligence with artificial intelligence but in creating a synergistic relationship where both complement each other. The following directions are crucial for ensuring the effective and responsible integration of Generative AI into commerce education.

One of the most important future directions is the systematic integration of AI-related competencies into business and commerce curricula. Traditionally, commerce education has focused on accounting principles, financial management, economics, marketing, taxation, and business law. While these subjects remain essential, the increasing adoption of AI technologies across industries demands the inclusion of AI literacy as a core component of business education.

Educational institutions should redesign curricula to include topics such as Artificial Intelligence, Machine Learning, Data Analytics, Prompt Engineering, Business Intelligence, and Digital Transformation. Students should be introduced to the practical applications of AI in accounting, finance, marketing, supply chain management, and human resource management. Rather than teaching AI as a separate subject, institutions should integrate AI concepts across various business disciplines to demonstrate their relevance in real-world contexts. For example, accounting courses may incorporate AI-based auditing and fraud detection systems, while marketing courses can explore AI-driven consumer analytics and personalized advertising. Such integration will ensure that graduates possess the technological competencies required in the modern workplace.

The successful implementation of AI-enhanced education depends significantly on the preparedness of educators. Many faculty members may lack sufficient knowledge and experience regarding Generative AI tools and their pedagogical applications. Therefore, continuous professional development programs are essential to equip educators with the necessary skills to effectively integrate AI into teaching and learning processes.

Institutions should organize workshops, certification courses, seminars, and training programs focused on AI literacy, ethical AI use, prompt engineering, digital pedagogy, and AI-assisted assessment methods. Faculty members should be encouraged to experiment with AI tools and explore innovative instructional strategies that enhance student engagement and learning outcomes. Moreover, educators must develop the ability to critically evaluate AI-generated content and guide students in its responsible use. By empowering faculty members with technological competence and pedagogical confidence, institutions can facilitate smoother adoption of AI-enhanced educational practices.

As the use of Generative AI becomes increasingly widespread, educational institutions must establish comprehensive ethical guidelines governing its application. Ethical considerations such as academic integrity, intellectual property rights,

ISBN: 978-93-6348-139-8

transparency, accountability, fairness, and data privacy have become central concerns in AI-enabled learning environments.

Universities and colleges should develop institutional policies that clearly define acceptable and unacceptable uses of AI technologies. These guidelines should address issues related to plagiarism, authorship attribution, responsible content generation, and disclosure requirements when AI tools are used in academic work.

Ethical AI education should become an integral component of commerce curricula. Students must be trained to recognize algorithmic bias, understand data privacy implications, and evaluate the social consequences of AI-driven decision-making. Developing ethical awareness will enable future business professionals to use AI responsibly while maintaining professional integrity and societal accountability.

The future of education should emphasize collaboration between human intelligence and artificial intelligence rather than competition between the two. While AI systems can process large amounts of information, generate content rapidly, and automate routine tasks, they cannot fully replicate human creativity, emotional intelligence, ethical judgment, and contextual understanding.

Educational institutions should promote pedagogical models that encourage students to work alongside AI tools while maintaining active intellectual engagement. Students should be trained to critically assess AI-generated outputs, verify information accuracy, identify biases, and apply human judgment when making decisions. In business education, this collaborative approach can prepare students for workplaces where AI technologies increasingly support managerial and operational functions. Future professionals will need to combine technological proficiency with uniquely human skills such as leadership, communication, creativity, negotiation, and strategic thinking. Therefore, educational systems must focus on developing competencies that complement AI capabilities rather than compete with them.

The widespread availability of Generative AI has challenged traditional assessment methods that rely heavily on written assignments, essays, and take-home projects. Since AI tools can generate high-quality responses within seconds, conventional assessments may no longer accurately measure students' knowledge, understanding, and critical thinking abilities. Future assessment strategies should focus on evaluating higher-order cognitive skills such as analysis, synthesis, evaluation, creativity, and problem-solving. Educational institutions may increasingly adopt project-based learning, case study analysis, reflective journals, oral presentations, simulations, and experiential learning activities.

Authentic assessment methods that require students to demonstrate practical application of knowledge will become more important. For example, commerce students may be asked to analyze real business problems, develop strategic recommendations, or present AI-assisted solutions while explaining their reasoning processes. Such approaches can help ensure that students actively engage with learning rather than relying solely on AI-generated outputs.

ISBN: 978-93-6348-139-8

AI literacy is rapidly becoming an essential skill for students across all disciplines. Future business graduates must understand not only how to use AI tools but also how these technologies function, their limitations, ethical implications, and potential risks. Educational institutions should incorporate AI literacy programs into undergraduate and postgraduate curricula. These programs should cover fundamental concepts of artificial intelligence, machine learning, data analytics, algorithmic decision-making, cybersecurity, and responsible technology use. Developing digital competence will enable students to interact effectively with emerging technologies and adapt to rapidly changing workplace environments. Such skills will enhance employability and support lifelong learning in an increasingly digital economy.

The educational implications of Generative AI remain an evolving field of inquiry. Future research is necessary to understand the long-term effects of AI integration on learning outcomes, academic performance, student engagement, and professional skill development. Educational institutions should encourage interdisciplinary research that explores the intersection of artificial intelligence, business education, pedagogy, and organizational development. Researchers can investigate questions related to AI-assisted learning, ethical challenges, assessment innovations, curriculum effectiveness, and workforce preparedness. Additionally, universities can establish innovation centers and research laboratories dedicated to AI applications in education and business. Such initiatives will contribute to evidence-based policymaking and promote the development of innovative educational practices.

As AI technologies become more prevalent, ensuring equitable access to these resources is essential. Educational institutions must address digital divides that may prevent certain students from benefiting fully from AI-enhanced learning opportunities. Future policies should focus on providing affordable access to technological infrastructure, digital resources, and AI-powered learning tools. Institutions should also ensure that AI systems are designed to support diverse learners, including students with disabilities and those from disadvantaged backgrounds.

Inclusive AI-driven learning environments can enhance educational equity and ensure that technological advancements benefit all learners rather than a privileged few.

The future of business education will be shaped by the extent to which educational institutions successfully integrate Generative AI into teaching, learning, and curriculum development. By redesigning curricula, empowering educators, promoting ethical AI use, transforming assessment practices, and fostering human-AI collaboration, institutions can prepare students for the challenges and opportunities of the digital age. The objective should not be to replace human intelligence but to leverage AI as a tool that enhances learning, innovation, and professional development in commerce education.

Generative Artificial Intelligence represents a transformative force in business education and commerce curricula. Its ability to support personalized learning, enhance teaching efficiency, foster innovation, and develop industry-relevant skills offers

ISBN: 978-93-6348-139-8

significant opportunities for educational institutions. However, challenges related to academic integrity, ethical concerns, technological dependence, and faculty readiness must be addressed through thoughtful policy development and curriculum reform. As AI continues to reshape the business landscape, educational institutions have a responsibility to prepare students for a future in which human expertise and artificial intelligence coexist. By promoting AI literacy, ethical awareness, and critical thinking, business education can leverage Generative AI as a catalyst for innovation and lifelong learning.

Works Cited

1. Bearman, Margaret, Jennifer Ryan, and Rola Ajjawi. "Discourses of Artificial Intelligence in Higher Education." *Higher Education Research & Development*, vol. 42, no. 7, 2023, pp. 1512–1527.
2. Dwivedi, Yogesh K., et al. "So What If ChatGPT Wrote It? Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI." *International Journal of Information Management*, vol. 71, 2023, article 102642.
3. Kasneci, Enkelejda, et al. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences*, vol. 103, 2023, article 102274.
4. Luckin, Rose, Wayne Holmes, Mark Griffiths, and Laurie B. Forcier. *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education, 2016.
5. Tlili, Ahmed, et al. "What If the Devil Is My Guardian Angel? ChatGPT as a Case Study of Generative Artificial Intelligence in Education." *Smart Learning Environments*, vol. 10, no. 1, 2023, article 15.
6. *The Future of Jobs Report 2024*. World Economic Forum, 2024.
7. *Guidance for Generative AI in Education and Research*. UNESCO, 2023.

ISBN: 978-93-6348-139-8

ARTIFICIAL INTELLIGENCE IMPACT ON MARITIME EDUCATION: A FRAMEWORK FOR ADAPTIVE LEARNING, COMPETENCY PREDICTION, AND ACADEMIC DECISION SUPPORT

ID: ORBIT – 008

Mr. R. Rajapriyan
Research Scholar,
Ph. D Maritime Studies,
School of Maritime Studies,
Vels Institute of Science, Technology & Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Dr. Capt. N. Kumar
Supervisor,
Professor & Director,
School of Maritime Studies,
Vels Institute of Science, Technology & Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Dr. T. Kamalakannan
Co-Supervisor,
Professor,
Department of BCA,
School of Computing Science,
Vels Institute of Science, Technology & Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Abstract

Artificial Intelligence (AI) is transforming higher education by enabling intelligent, personalized, and data-driven learning environments. In Maritime Education and Training (MET), competency-based learning, simulator training, and compliance with the International Maritime Organization (IMO) Standards of Training, Certification and Watchkeeping (STCW) require continuous monitoring of cadet performance and academic progress. Traditional teaching and assessment methods often provide limited opportunities for personalized learning and early identification of competency gaps. AI technologies such as Machine Learning, Learning Analytics, Natural Language Processing, and Predictive Analytics offer effective solutions to improve learning outcomes, automate assessments, and support academic decision-making. Their integration into maritime education has the potential to enhance both educational quality and institutional efficiency.

ISBN: 978-93-6348-139-8

This paper proposes an Artificial Intelligence framework for maritime education that integrates adaptive learning, competency prediction, and academic decision support into a unified intelligent educational ecosystem. The proposed framework utilizes academic records, attendance, simulator performance, assessment results, and learning analytics to generate personalized learning pathways, predict cadet competency levels, identify academically at-risk learners, and assist faculty members through data-driven insights. The framework also supports automated assessment, performance monitoring, and institutional academic planning while addressing implementation challenges such as data privacy, infrastructure, faculty readiness, and ethical AI adoption. The proposed model provides a practical roadmap for modernizing maritime education and preparing future maritime professionals for the rapidly evolving digital maritime industry.

Keywords: Artificial Intelligence, Maritime Education, Adaptive Learning, Competency Prediction, Academic Decision Support, Learning Analytics, Maritime Training, Predictive Analytics.

1. Introduction

The maritime industry is rapidly embracing Artificial Intelligence (AI), the Internet of Things (IoT), Big Data Analytics, and digital technologies to improve operational efficiency, navigational safety, and sustainable shipping. These technological advancements are transforming modern maritime operations, creating a demand for highly skilled maritime professionals who possess both technical knowledge and digital competencies. Consequently, Maritime Education and Training (MET) institutions must modernize their teaching methodologies to prepare cadets for an increasingly technology-driven maritime environment.

Maritime education is unique because it combines classroom instruction, laboratory practice, simulator-based training, onboard experience, and competency-based assessments in accordance with the International Maritime Organization (IMO) Standards of Training, Certification and Watchkeeping (STCW). Traditional teaching and assessment methods often rely on standardized instruction and periodic examinations, making it difficult to provide personalized learning experiences, continuously monitor cadet competency, and identify academically at-risk students at an early stage. These limitations highlight the need for intelligent educational systems that can support both learners and educators.

Artificial Intelligence provides innovative solutions to address these challenges through adaptive learning, learning analytics, competency prediction, intelligent assessment, and academic decision support. AI techniques can analyze attendance records, assessment scores, simulator performance, and learning behaviour to generate personalized learning recommendations, predict competency achievement, and provide timely academic interventions. Such capabilities improve learning outcomes while reducing faculty workload and supporting evidence-based academic decision-making.

ISBN: 978-93-6348-139-8

Although AI has been widely applied in autonomous shipping, smart ports, maritime logistics, and predictive maintenance, its application in maritime education remains limited. Existing studies primarily focus on operational improvements rather than intelligent educational systems. Therefore, this paper proposes an AI-based framework that integrates adaptive learning, competency prediction, and academic decision support to enhance teaching, learning, and institutional governance in maritime education.

Research Objectives

- To study the impact of Artificial Intelligence on Maritime Education.
- To develop an AI-based framework for adaptive learning and competency prediction.
- To support faculty and administrators through academic decision support.
- To identify the benefits and implementation challenges of AI in Maritime Education.

Paper Contributions

The major contributions of this paper are:

- Proposes an AI-driven framework specifically for Maritime Education.
- Integrates adaptive learning, competency prediction, and academic decision support into a unified model.
- Presents a conceptual architecture for intelligent academic management.
- Discusses the practical benefits, implementation challenges, and future opportunities of AI adoption in maritime institutions.

2. Literature Review

Artificial Intelligence (AI) has become an important technology in higher education by supporting personalized learning, intelligent assessment, learning analytics, and academic decision-making. Recent studies have shown that AI techniques, including Machine Learning, Natural Language Processing, and Predictive Analytics, improve student engagement, automate assessments, and identify academically at-risk learners. These technologies enable institutions to enhance teaching effectiveness while reducing administrative workload.

In the maritime sector, AI research has mainly focused on autonomous ships, smart ports, maritime logistics, navigation, and predictive maintenance. However, limited research has explored AI applications in Maritime Education and Training (MET), particularly in adaptive learning, competency prediction, and academic decision support. Existing studies generally address individual educational applications rather than providing an integrated framework that supports teaching, learning, assessment, and institutional management. Therefore, this study proposes a comprehensive AI framework that combines adaptive learning, competency prediction, and academic

ISBN: 978-93-6348-139-8

decision support to enhance educational quality and competency-based maritime training.

Table 1. Summary of Previous Studies

Author	Year	Major Contribution	Research Gap
Ibrahim & Hilal	2026	AI in Engineering Education	No Maritime Education
Lall & Liu	2025	AI for Maritime Training	Limited Academic Decision Support
Kizilay & Çelik	2025	AI in Maritime Competency Assessment	No Adaptive Learning Framework
Crompton & Burke	2023	AI in Higher Education	General Education Focus
Proposed Work	2026	Integrated AI Framework for Maritime Education	Combines Adaptive Learning, Competency Prediction, and Academic Decision Support

The literature review identifies a clear research gap in developing an integrated AI-driven framework specifically for Maritime Education, which forms the primary contribution of this study.

3. Proposed AI Framework

The proposed Artificial Intelligence (AI) framework is designed to enhance Maritime Education by integrating adaptive learning, competency prediction, and academic decision support into a unified educational ecosystem. The framework utilizes academic, behavioural, and simulator-generated data to provide personalized learning experiences and support evidence-based decision-making for faculty members and administrators.

The framework consists of six major components: **Data Collection**, **Data Preprocessing**, **AI Engine**, **Adaptive Learning**, **Competency Prediction**, and **Academic Decision Support**. Academic data such as attendance, internal assessment marks, simulator performance, assignments, and STCW competency records are collected and preprocessed before being analyzed by AI algorithms. Machine Learning, Learning Analytics, Natural Language Processing, and Predictive Analytics are employed to identify learning patterns, predict academic performance, detect competency gaps, and recommend personalized learning pathways.

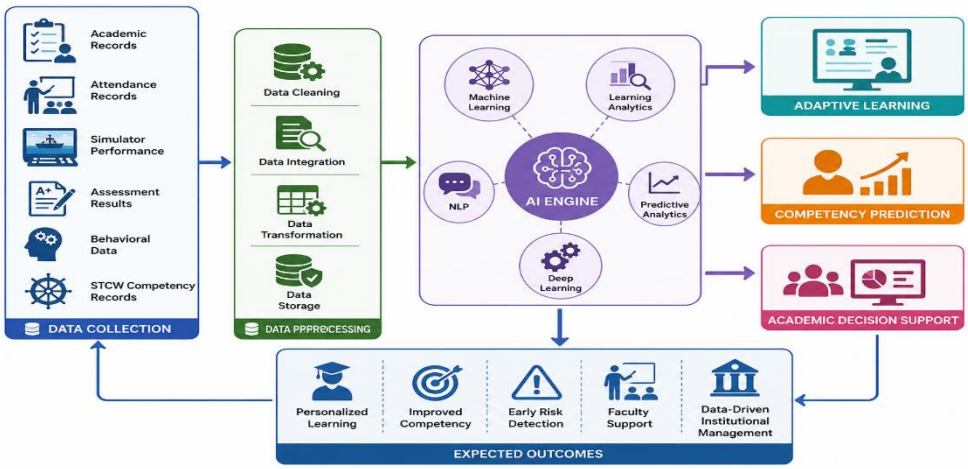
The Adaptive Learning module provides customized learning resources based on each cadet's strengths and weaknesses, while the Competency Prediction module continuously evaluates competency achievement and identifies at-risk cadets. The Academic Decision Support module assists faculty members through intelligent

ISBN: 978-93-6348-139-8

dashboards that monitor student progress, curriculum effectiveness, and institutional performance.

The proposed framework enables personalized learning, continuous competency assessment, early academic intervention, and data-driven educational management. By integrating AI technologies into Maritime Education and Training (MET), the framework supports improved learning outcomes, enhanced faculty efficiency, and effective academic governance.

Figure 1. Proposed AI Framework for Maritime Education



4. Proposed System Architecture

The proposed AI-based system architecture integrates educational data, Artificial Intelligence techniques, and decision-support mechanisms to improve teaching, learning, and academic management in Maritime Education. The architecture consists of five interconnected layers that process academic information and generate intelligent recommendations for cadets, faculty members, and administrators.

The first layer collects data from multiple sources, including student records, attendance, assessment scores, simulator performance, Learning Management Systems (LMS), and STCW competency records. The collected data are cleaned, integrated, and transformed before being stored in a centralized educational database.

The AI engine forms the core of the architecture by applying Machine Learning, Learning Analytics, Natural Language Processing (NLP), and Predictive Analytics to identify learning patterns, predict competency levels, and detect academically at-risk cadets. Based on these analyses, the system generates personalized learning recommendations, automated assessments, competency reports, and intelligent feedback.

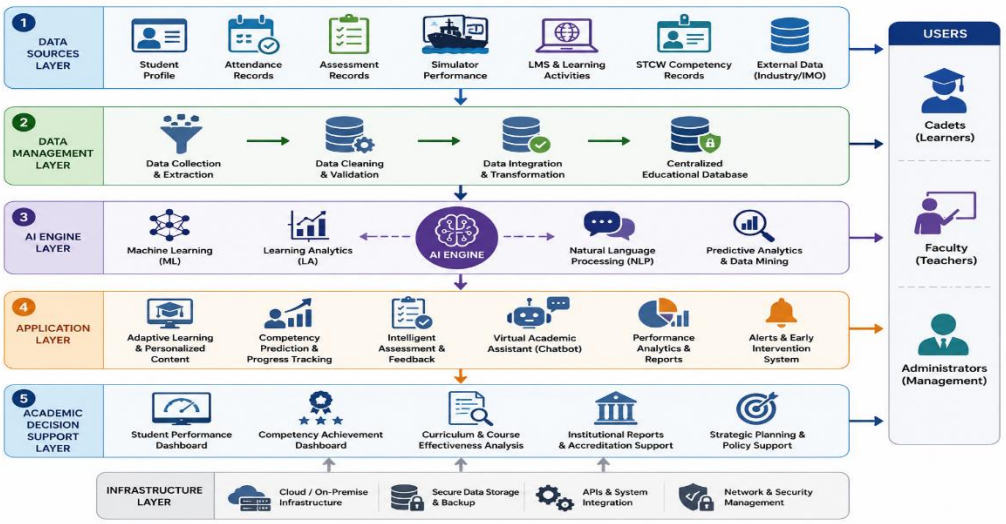
The application layer delivers AI-enabled services such as adaptive learning, competency prediction, intelligent assessment, virtual academic assistance, and performance analytics. These services assist cadets in improving learning outcomes

ISBN: 978-93-6348-139-8

while supporting faculty members in monitoring academic progress and institutional performance.

Finally, the academic decision-support layer provides interactive dashboards that enable administrators to evaluate student performance, optimize curriculum planning, monitor competency achievement, and support evidence-based decision-making. The proposed architecture creates a scalable and intelligent educational environment that enhances teaching quality, competency development, and institutional governance in Maritime Education.

Figure 2. Proposed AI-Based System Architecture



5. Discussion

The proposed AI framework demonstrates that Artificial Intelligence can significantly improve Maritime Education by supporting adaptive learning, competency prediction, intelligent assessment, and academic decision support. Unlike conventional educational systems, AI continuously analyzes academic and simulator data to generate personalized learning recommendations, identify competency gaps, and assist faculty members in monitoring cadet progress. This enables timely academic interventions and enhances competency-based learning in accordance with the STCW Convention.

The integration of Machine Learning, Learning Analytics, Natural Language Processing, and Predictive Analytics provides multiple benefits to maritime institutions. AI-powered systems reduce faculty workload through automated assessment and performance monitoring while improving learning outcomes through personalized educational support. Institutional administrators can also utilize AI-generated dashboards for curriculum planning, resource allocation, accreditation support, and strategic academic decision-making.

ISBN: 978-93-6348-139-8

Despite these advantages, implementing AI in maritime education presents several challenges. Data privacy, cybersecurity, ethical concerns, infrastructure requirements, and faculty readiness remain significant barriers. Educational institutions must establish secure data management practices, provide faculty training, and develop appropriate AI governance policies to ensure responsible adoption of intelligent technologies.

Future research should focus on Explainable Artificial Intelligence (XAI), Digital Twin technology, Virtual Reality (VR), Augmented Reality (AR), Generative AI, and intelligent simulator analytics to further enhance competency-based maritime education. Integrating these emerging technologies can create smart learning environments that improve cadet competency, educational quality, and institutional performance while preparing future maritime professionals for the evolving digital maritime industry.

6. Conclusion

Artificial Intelligence has the potential to transform Maritime Education by enabling adaptive learning, competency prediction, intelligent assessment, and academic decision support. The proposed AI framework integrates educational data with Machine Learning, Learning Analytics, Natural Language Processing, and Predictive Analytics to create a personalized and data-driven learning environment for cadets while supporting faculty members and institutional administrators.

The framework enhances competency-based education by providing personalized learning pathways, early identification of at-risk cadets, automated assessment, and real-time academic analytics. It also supports evidence-based decision-making, improves educational quality, and strengthens institutional governance. These capabilities contribute to preparing competent maritime professionals who can meet the technological demands of the modern maritime industry.

Although challenges such as data privacy, infrastructure, faculty readiness, and ethical AI implementation remain, the proposed framework provides a practical foundation for integrating AI into Maritime Education and Training (MET). Future research can extend this work by incorporating Explainable AI, Digital Twin technology, Virtual Reality, and Generative AI to develop intelligent and future-ready maritime education systems.

References

1. M. G. Ibrahim and R. Hilal, "Artificial Intelligence for Inclusive Engineering Education: Advancing Equality, Diversity, and Ethical Leadership," *arXiv preprint*, 2026.
2. V. Lall and Y. Liu, "AI Meets Maritime Training: Precision Analytics for Enhanced Safety and Performance," *arXiv preprint*, 2025.

ISBN: 978-93-6348-139-8

3. F. E. Kizilay and M. Çelik, "A Literature Review on Generative Artificial Intelligence in Educational Assessment: An Extension to Maritime Competency Assessment," *WMU Journal of Maritime Affairs*, 2025.

4. N. Maslej *et al*., "The AI Index Report 2024," Stanford Institute for Human-Centered Artificial Intelligence, 2024.

5. K. Porayska-Pomsta, W. Holmes, and S. Nemorin, "The Ethics of AI in Education," *arXiv preprint*, 2024.

6. H. Crompton and D. Burke, "Artificial Intelligence in Higher Education: The State of the Field," *International Journal of Educational Technology in Higher Education*, vol. 20, 2023.

7. UNESCO, *Guidance for Generative Artificial Intelligence in Education and Research*, Paris, France, 2023.

8. M. C. Laupichler, A. Aster, J. Schirch, and T. Raupach, "Artificial Intelligence Literacy in Higher and Adult Education: A Scoping Literature Review," *Computers and Education: Artificial Intelligence*, vol. 3, 2022.

9. N. Selwyn, "The Future of AI and Education: Some Cautionary Notes," *European Journal of Education*, vol. 57, no. 4, pp. 620–631, 2022.

10. W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises, Implications and Future Directions*, 2021.

ISBN: 978-93-6348-139-8

DIGITAL TRANSFORMATION IN NON-MAJOR PORTS AND ITS IMPACT ON LOGISTICS EFFICIENCY: A STUDY OF INDIAN MARITIME INFRASTRUCTURE

ID: ORBIT – 009

Mr. Udhayan J
Research Scholar,
School of Maritime Studies,
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Dr. Capt. N. Kumar
Professor & Director,
School of Maritime Studies,
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Dr. G. Rajini
Professor,
Department of MBA,
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, Tamil Nadu, India.

Abstract

The rapid advancement of digital technologies has transformed the global maritime industry by improving operational efficiency, transparency, and service quality. In India, non-major ports play a significant role in supporting regional trade, coastal shipping, and economic development. However, many of these ports continue to face challenges such as manual documentation, limited automation, inefficient cargo handling, and inadequate information sharing among stakeholders. Digital transformation through technologies such as the Port Community System (PCS), Internet of Things (IoT), Artificial Intelligence (AI), cloud computing, RFID, and blockchain has the potential to modernize port operations and enhance logistics performance. This paper examines the current status of digital transformation in Indian non-major ports and explores its influence on logistics efficiency.

ISBN: 978-93-6348-139-8

The study adopts a review-based research approach by analyzing published literature, government reports, and maritime policy documents related to digital port initiatives. The findings indicate that digital transformation contributes to reduced vessel turnaround time, faster cargo clearance, improved supply chain visibility, enhanced resource utilization, and lower logistics costs. Despite these benefits, challenges including infrastructure limitations, cybersecurity concerns, financial constraints, and shortage of skilled manpower continue to affect the implementation of digital technologies in non-major ports. The paper concludes by recommending strategic investments in digital infrastructure, workforce training, stakeholder collaboration, and policy support to accelerate the development of smart and efficient non-major ports, thereby strengthening India's maritime logistics ecosystem.

Keywords: Digital Transformation; Non-Major Ports; Logistics Efficiency; Smart Ports; Port Community System (PCS); Artificial Intelligence; Internet of Things (IoT); Maritime Logistics; Supply Chain Management; Port Automation.

1. Introduction

The maritime sector is a fundamental component of global trade and economic development, with nearly 80% of international merchandise trade transported by sea. Ports act as critical gateways connecting maritime transportation with inland logistics networks, thereby facilitating efficient cargo movement and supply chain operations. In India, ports are classified into major and non-major ports, with non-major ports playing an increasingly significant role in handling regional trade, coastal shipping, and industrial cargo. The continuous growth in cargo traffic and the increasing demand for faster logistics services have highlighted the need for technological modernization in non-major ports. Traditional port operations that rely on manual documentation, fragmented communication systems, and limited automation often result in operational delays, higher logistics costs, and reduced service efficiency.

Digital transformation has emerged as a strategic solution to address these operational challenges by integrating advanced technologies into port management and logistics activities. Technologies such as Artificial Intelligence (AI), Internet of Things (IoT), Cloud Computing, Blockchain, Big Data Analytics, Radio Frequency Identification (RFID), and Port Community Systems (PCS) enable real-time data sharing, intelligent decision-making, automated cargo tracking, and enhanced coordination among port stakeholders. These technologies improve transparency, reduce paperwork, optimize resource utilization, and support predictive maintenance of port infrastructure. Consequently, digital transformation contributes to improved logistics performance through reduced vessel turnaround time, efficient cargo handling, and enhanced supply chain visibility.

Although India's major ports have made considerable progress in adopting digital technologies under initiatives such as Maritime India Vision 2030 and the National Logistics Policy, the level of digital adoption in many non-major ports remains

ISBN: 978-93-6348-139-8

comparatively low. Several non-major ports continue to face challenges related to inadequate digital infrastructure, financial limitations, cybersecurity concerns, workforce skill gaps, and interoperability among information systems. These issues limit their operational efficiency and competitiveness in the evolving global maritime sector. Understanding the impact of digital transformation on logistics efficiency is therefore essential for identifying opportunities to modernize port operations and improve service delivery.

This paper presents a comprehensive review of digital transformation practices in Indian non-major ports and evaluates their influence on logistics efficiency. The study examines the role of emerging digital technologies in improving cargo operations, documentation processes, vessel management, and information sharing among stakeholders. It also identifies the key challenges associated with technology implementation and proposes strategic recommendations for accelerating digital transformation in non-major ports. The findings of this study are expected to support policymakers, port authorities, logistics service providers, and researchers in promoting sustainable, efficient, and technology-driven port operations that strengthen India's maritime logistics ecosystem.

2. Literature Review

Digital transformation has become a key driver of innovation in the maritime industry by integrating advanced information and communication technologies into port operations and logistics management. The adoption of digital technologies enables ports to improve operational efficiency, reduce turnaround time, enhance supply chain visibility, and provide better services to stakeholders. Over the past decade, researchers have explored the role of Artificial Intelligence (AI), Internet of Things (IoT), Blockchain, Cloud Computing, Big Data Analytics, and Port Community Systems (PCS) in developing smart and sustainable ports.

Several studies have highlighted that digital transformation significantly improves cargo handling efficiency and port productivity. AI-based predictive analytics assists port authorities in forecasting cargo traffic, optimizing berth allocation, and scheduling vessel movements. Machine learning algorithms also support predictive maintenance of cranes, container handling equipment, and other critical infrastructure, thereby minimizing equipment failures and operational downtime. These intelligent technologies contribute to faster decision-making and improved utilization of port resources.

The Internet of Things (IoT) has emerged as one of the most widely adopted technologies in modern port operations. IoT devices, including RFID tags, GPS sensors, environmental monitoring systems, and smart cameras, enable continuous monitoring of cargo, vessels, containers, and port equipment. Real-time data collected through IoT devices improve cargo tracking, enhance security, reduce cargo loss, and support effective fleet management. Researchers have reported that IoT-enabled ports

ISBN: 978-93-6348-139-8

experience higher operational transparency and better coordination among logistics stakeholders.

Blockchain technology has also gained attention for improving the security and transparency of maritime documentation. Traditional paper-based documentation often results in processing delays, duplication of records, and increased administrative costs. Blockchain-based digital documentation enables secure information sharing among shipping companies, customs authorities, freight forwarders, and port operators. Smart contracts further automate transaction verification and reduce the possibility of fraud and documentation errors. As a result, blockchain contributes to faster customs clearance and improved supply chain reliability.

Cloud Computing provides scalable computing infrastructure for storing, processing, and sharing large volumes of operational data generated by port activities. Cloud-based platforms facilitate seamless communication among multiple stakeholders while reducing hardware investment and maintenance costs. Researchers have emphasized that cloud-enabled Port Community Systems improve collaboration between shipping lines, customs, logistics providers, and port authorities by providing centralized access to operational information. This enhances decision-making and supports integrated logistics management.

Big Data Analytics has become an essential tool for analyzing operational and logistics data collected from various digital systems. By processing historical and real-time information, analytical models identify traffic patterns, predict cargo demand, optimize warehouse utilization, and improve logistics planning. Studies indicate that data-driven decision-making enables ports to reduce congestion, improve resource allocation, and enhance customer service. Big Data Analytics also supports environmental sustainability through optimized fuel consumption and reduced vessel waiting times.

Despite these technological advancements, researchers have identified several barriers affecting digital transformation in non-major ports. Limited financial resources, inadequate digital infrastructure, lack of skilled personnel, cybersecurity threats, and resistance to organizational change continue to hinder technology adoption. Small and medium-sized ports often face additional challenges due to lower investment capacity and fragmented information systems. These limitations reduce the effectiveness of digital transformation initiatives and slow the development of smart port ecosystems.

The existing literature demonstrates that digital technologies have a positive influence on logistics efficiency by improving operational performance, reducing costs, increasing transparency, and strengthening supply chain integration. However, most previous studies primarily focus on major international ports, while research specifically addressing digital transformation in Indian non-major ports remains limited. This research addresses that gap by examining the impact of digital transformation on logistics efficiency in Indian non-major ports and proposing practical recommendations for enhancing their digital capabilities.

ISBN: 978-93-6348-139-8

Research Gap

The review of existing literature reveals that most previous studies concentrate on digital transformation in major ports, container terminals, and global smart port initiatives. Comparatively fewer studies examine the implementation of digital technologies in Indian non-major ports, particularly their influence on logistics efficiency. Furthermore, limited research has integrated AI, IoT, Blockchain, Cloud Computing, Big Data Analytics, and Port Community Systems into a single conceptual framework for evaluating logistics performance. Therefore, this study aims to bridge this gap by providing a comprehensive review of digital transformation practices and their impact on logistics efficiency in Indian non-major ports.

3. Research Methodology

3.1 Research Design

This study adopts a qualitative review-based research design to examine the impact of digital transformation on logistics efficiency in Indian non-major ports. The study synthesizes information from published research articles, government reports, policy documents, and industry publications to understand the adoption of digital technologies and their contribution to improving port operations. A review-based approach is appropriate because it enables the identification of existing knowledge, research gaps, and future opportunities related to digital transformation in the maritime sector.

3.2 Data Sources

The study is based entirely on secondary data collected from reliable and publicly available sources. These include peer-reviewed journal articles, conference proceedings, reports published by the Ministry of Ports, Shipping and Waterways, Maritime India Vision 2030, National Logistics Policy documents, NITI Aayog reports, and publications from international maritime organizations. These sources provide comprehensive information on digital transformation initiatives, logistics performance, and technological advancements in port management.

3.3 Digital Technologies Considered

The research focuses on major digital technologies that influence logistics efficiency in non-major ports. These technologies include:

- Artificial Intelligence (AI)
- Internet of Things (IoT)
- Cloud Computing
- Blockchain Technology
- Big Data Analytics
- Radio Frequency Identification (RFID)
- Port Community System (PCS)

ISBN: 978-93-6348-139-8

These technologies were selected because they are widely recognized as essential components of smart port development and digital logistics management.

3.4 Analysis Method

A thematic analysis approach was employed to analyze the collected literature. The information was classified into key themes, including digital technology adoption, operational efficiency, cargo handling, documentation processes, logistics performance, implementation challenges, and future opportunities. The findings from different studies were compared to identify common trends and evaluate the overall impact of digital transformation on logistics efficiency.

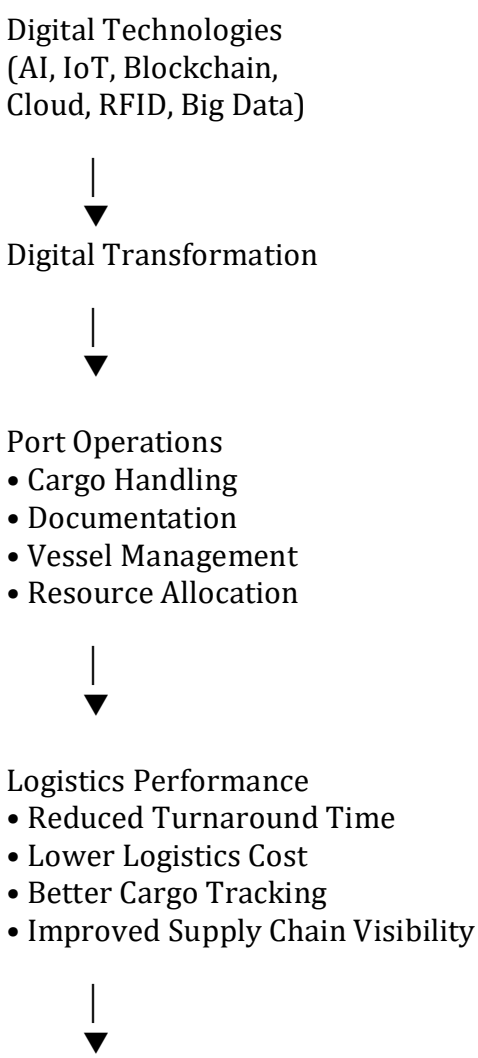
Table 1. Digital Technologies and Their Applications in Non-Major Ports

Digital Technology	Major Application	Expected Benefit
Artificial Intelligence	Predictive maintenance, cargo forecasting	Improved operational efficiency
Internet of Things	Real-time monitoring of cargo and equipment	Better visibility and tracking
Blockchain	Secure digital documentation	Faster customs clearance
Cloud Computing	Data storage and information sharing	Improved collaboration
Big Data Analytics	Demand forecasting and performance analysis	Better decision-making
RFID	Container and cargo identification	Reduced cargo handling time
Port Community System	Integrated stakeholder communication	Faster information exchange

Proposed Conceptual Framework

The relationship between digital transformation and logistics efficiency is illustrated through the following conceptual framework.

ISBN: 978-93-6348-139-8



The framework indicates that the adoption of advanced digital technologies enhances port operations by improving information sharing, automating processes, optimizing resource utilization, and strengthening coordination among stakeholders. These improvements ultimately lead to higher logistics efficiency and better overall port performance.

4. Results and Discussion

The review of existing literature indicates that digital transformation has a significant positive impact on the operational efficiency of non-major ports. The adoption of advanced digital technologies enables ports to automate routine activities, improve communication among stakeholders, and optimize logistics processes. AI-

ISBN: 978-93-6348-139-8

based predictive analytics supports efficient berth allocation, vessel scheduling, and maintenance planning, reducing operational delays and increasing asset utilization.

IoT technologies enhance real-time monitoring of cargo, containers, and port equipment through connected sensors and RFID devices. This improves cargo visibility, minimizes losses, and enables faster decision-making. Similarly, cloud-based platforms and Port Community Systems facilitate seamless information sharing among shipping companies, customs authorities, freight forwarders, and port operators, thereby reducing paperwork and processing time.

Blockchain technology strengthens the security and transparency of digital documentation by eliminating duplicate records and minimizing fraudulent transactions. Digital documentation significantly accelerates customs clearance and cargo release, contributing to shorter vessel turnaround times and improved customer satisfaction.

Despite these advantages, several challenges continue to affect digital transformation in Indian non-major ports. High implementation costs, inadequate digital infrastructure, limited technical expertise, cybersecurity risks, and organizational resistance remain major obstacles. Small and medium-sized ports often lack sufficient financial resources to invest in advanced technologies, slowing the pace of digital adoption.

Overall, the findings suggest that successful implementation of digital technologies can significantly improve logistics efficiency by reducing operational costs, increasing productivity, enhancing transparency, and supporting sustainable port development. However, achieving these benefits requires coordinated efforts from government agencies, port authorities, private stakeholders, and technology providers to develop robust digital infrastructure and skilled human resources.

5. Recommendations

Based on the findings of this study, the following recommendations are proposed to enhance digital transformation and improve logistics efficiency in Indian non-major ports.

1. Strengthen Digital Infrastructure: Non-major ports should invest in high-speed communication networks, cloud-based platforms, and integrated digital systems to support smart port operations.
2. Implement Artificial Intelligence Solutions: AI-based predictive analytics should be adopted for vessel scheduling, berth allocation, cargo forecasting, and predictive maintenance to improve operational efficiency.
3. Expand the Use of IoT Technologies: IoT sensors, RFID devices, GPS tracking systems, and smart surveillance should be deployed for real-time monitoring of cargo, containers, and port equipment.
4. Promote Blockchain-Based Documentation: Digital documentation using blockchain technology can reduce paperwork, improve transparency, enhance data security, and accelerate customs clearance.

ISBN: 978-93-6348-139-8

5. Enhance Workforce Skills: Regular training programmes should be conducted for port employees to improve digital literacy and technical competencies required for smart port operations.
6. Strengthen Cybersecurity Measures: Comprehensive cybersecurity policies, periodic security audits, and data protection mechanisms should be implemented to safeguard digital infrastructure from cyber threats.
7. Encourage Public–Private Partnerships: Collaboration between government agencies, private logistics companies, technology providers, and academic institutions can accelerate digital transformation initiatives.
8. Develop Standardized Digital Policies: Uniform digital standards and interoperability frameworks should be established to facilitate seamless information exchange among all maritime stakeholders.

Implementation of these recommendations will improve operational efficiency, reduce logistics costs, enhance supply chain visibility, and support the long-term competitiveness of Indian non-major ports.

6. Conclusion

Digital transformation has become an essential component of modern port management and logistics operations. The integration of Artificial Intelligence, Internet of Things, Cloud Computing, Blockchain, Big Data Analytics, RFID, and Port Community Systems has significantly improved the efficiency, transparency, and reliability of maritime logistics across the world. The review conducted in this study demonstrates that these technologies have the potential to transform the operational performance of Indian non-major ports by reducing vessel turnaround time, improving cargo handling efficiency, enhancing resource utilization, and minimizing logistics costs.

Despite the numerous benefits associated with digital transformation, several implementation challenges continue to affect non-major ports, including inadequate infrastructure, financial constraints, shortage of skilled professionals, cybersecurity concerns, and organizational resistance to technological change. Addressing these challenges requires coordinated efforts from policymakers, port authorities, technology providers, and logistics stakeholders.

The study concludes that successful digital transformation will not only improve logistics efficiency but also strengthen the competitiveness of Indian non-major ports in the global maritime sector. Continuous investment in digital infrastructure, workforce development, innovation, and policy support will play a crucial role in achieving sustainable smart port development and supporting India's long-term maritime growth.

7. Future Scope

Future research may focus on quantitative assessment of digital transformation by collecting operational data from individual non-major ports across India. Comparative studies between major and non-major ports can provide deeper insights

ISBN: 978-93-6348-139-8

into technology adoption and logistics performance. Researchers may also investigate the application of Digital Twin technology, autonomous port operations, Green Port initiatives, AI-powered decision support systems, and blockchain-enabled smart contracts in maritime logistics. Furthermore, empirical studies involving port authorities, shipping companies, customs officials, and logistics service providers can strengthen the understanding of practical implementation challenges and future technological opportunities.

References

1. Su, Z. (2026). *Digital Transformation in Port Logistics*. Encyclopedia, 6(1), Article 28. MDPI.
2. Sharma, R., & Kumar, P. (2026). *Digital Transformation in Port Logistics: Insights from the Chennai Port Case Study*. *International Journal of Research in Commerce and Management Studies, 8*(2), 84–101.
3. World Bank. (2021). *Accelerating Digitalization Across the Maritime Supply Chain*. Washington, DC: World Bank.
4. Basulo-Ribeiro, J., et al. (2024). *Industry 4.0 Supporting Logistics Towards Smart Ports*. *Journal of Industrial Engineering and Management*.
5. Sim, S., Kim, D., Park, K., & Bae, H. (2024). *Artificial Intelligence-based Smart Port Logistics Metaverse for Enhancing Productivity, Environment, and Safety in Port Logistics: A Case Study of Busan Port*.
6. Khazzar, A., Sekaki, Y., Lachhab, Y., & El-marzouki, S. (2025). *Artificial Intelligence in Port Logistics: A Bibliometric Analysis of Technological Integration and Research Dynamics*.
7. Redekar, D., Askin, R., & Ju, F. (2025). *Maritime Port Supply Chain Resilience: A Systematic Review*.
8. Lesniewska, F., Ani, U. D., Watson, J. M., & Carr, M. (2021). *The Internet of Things in Ports: Six Key Security and Governance Challenges for the UK*. Policy Brief.
9. Ministry of Ports, Shipping and Waterways. (2021). *Maritime India Vision 2030*. Government of India.
10. Government of India. (2022). *National Logistics Policy*. Ministry of Commerce and Industry.

ISBN: 978-93-6348-139-8

REINFORCEMENT LEARNING–ASSISTED MODEL PREDICTIVE ADAPTIVE VIRTUAL INERTIA CONTROL FOR SUPERCAPACITOR-BASED FREQUENCY SUPPORT IN GRID-CONNECTED PHOTOVOLTAIC SYSTEMS

ID: ORBIT – 010

Chenguttuvan J
Department of Marine Engineering,
Vel's Institute of Science, Technology & Advanced Studies (VISTAS),
Pallavaram, Chennai, Tamil Nadu, India.

Rubini B
Department of Electrical and Electronics Engineering,
Vel's Institute of Science, Technology & Advanced Studies (VISTAS),
Pallavaram, Chennai, Tamil Nadu, India.

Abstract

This paper proposes a novel Reinforcement Learning (RL)-assisted Model Predictive Control (MPC) framework for adaptive virtual inertia frequency support in low-inertia, grid-connected photovoltaic (PV) microgrids equipped with a supercapacitor energy storage system (SCESS). Modern power systems heavily integrated with high renewable penetration suffer from reduced rotational inertia, making them highly vulnerable to sudden power imbalances. Traditional virtual synchronous generator methods utilize static control loops that fail under sequential, high-impact disturbances. To address this, a dual-layer cooperative control strategy is developed: an outer-loop Q-learning reinforcement learning agent adaptively updates the virtual inertia gain coefficient Kv online based on dynamic tracking errors, while an inner-loop multi-step MPC enforces operational constraints over a 10-step predictive horizon to optimize SCESS active power injection. The proposed system is rigorously validated in MATLAB under sequential load steps and transient cloud-induced PV generation drops. The simulation results demonstrate outstanding resilience, restricting the frequency nadir to 49.9617 Hz under full reserves and maintaining 49.9635 Hz even under an energy-depleted state SOC_0 = 0.30, outperforming baseline controllers while ensuring strict constraint enforcement and hardware longevity.

Keywords: Adaptive virtual inertia, low-inertia grid, model predictive control, photovoltaic microgrid, reinforcement learning, supercapacitor energy storage.

ISBN: 978-93-6348-139-8

I. Introduction

The global transition toward sustainable energy paradigms has accelerated the integration of large-scale solar photovoltaic (PV) generation systems into modern utility grids. Despite environmental and economic benefits, the continuous replacement of conventional synchronous generators with inverter-interfaced renewable energy sources (IRES) fundamentally reshapes the dynamic performance of power networks, [3]. Unlike massive spinning rotors in synchronous power plants, power electronic inverters possess zero inherent rotational mass, leading to a dramatic reduction in the overall grid inertia. Consequently, low-inertia power systems exhibit extreme sensitivity to localized supply-demand mismatches, characterized by sharp initial frequency drops and high rates of change of frequency (ROCOF), [6]. If left unmitigated, these transient frequency excursions can trigger emergency load-shedding relays, isolate sensitive industrial components, and culminate in catastrophic wide-area cascading blackouts.

II. Literature review

To counteract the degradation of physical system inertia, the concept of virtual inertia control (VIC) has emerged as a promising technical solution, [11]. By mimicking the primary inertial behaviour of traditional machines, virtual synchronous generators use localized energy storage to inject or absorb real power during dynamic events, [13]. Supercapacitor energy storage systems (SCESS) are highly preferred for VIC applications owing to their exceptionally high power density, rapid response time, and immense cycle life compared to electrochemical batteries, [18].

However, conventional VIC methods predominantly rely on fixed virtual inertia parameters or manually tuned proportional-integral-derivative (PID) control loops, [24]. These static frameworks cannot adapt to real-time grid status and heavily degrade when subjected to overlapping sequential disturbances, such as a cloud-induced PV generation dip immediately following an industrial load step, [26]. Recently, modern optimization techniques like Model Predictive Control (MPC) have been explored due to their explicit ability to manage multi-step future states and enforce strict state-of-charge (SOC) or converter boundaries, [12].

Nevertheless, standard MPC algorithms require absolute mathematical accuracy of the grid parameters and often struggle to compute time-varying inertia coefficients dynamically during rapid network variations, [25]. On the other hand, data-driven methods such as reinforcement learning (RL) offer model-free learning abilities to update gains under parameter uncertainties, [27], but applying pure RL directly to physical systems poses security risks due to unconstrained execution bounds during the initial training exploration phase, [29].

To bridge these foundational gaps, this paper introduces a robust, multi-layer Reinforcement Learning-Assisted Model Predictive Adaptive Virtual Inertia Control topology designed specifically for SCESS-supported grid-tied PV platforms, [23]. The core architecture splits the computational intelligence: an outer-loop model-free Q-

ISBN: 978-93-6348-139-8

learning agent observes current frequency and ROCOF deviations to dynamically adapt the virtual inertia coefficient (Kv) online without requiring precise network parameter modeling, [29]. Concurrently, an inner-loop multi-step predictive controller (10-step predictive horizon) takes this optimal coefficient to dispatch real-time SCESS active power profiles while safeguarding the supercapacitor from over-discharging or converter thermal stress, [28].

The primary and distinct technical contributions of this research paper are outlined as follows:

- Cooperative Dual-Layer Control Architecture: Developed a coherent hierarchical scheme that successfully fuses model-free Reinforcement Learning with model-based multi-step Model Predictive Control, leveraging the online adaptation of the former and the strict constraint satisfaction of the latter, [20].
- Adaptive Parameter Optimization Under Sequential Stress: Demonstrated dynamic tracking capability under overlapping severe grid faults, wherein the RL agent continuously updates the virtual inertia parameters to compress ROCOF while mitigating secondary nadir drops, [22].
- Robust Boundary Enforcement in Stressed States: Validated extreme system resilience under depleted storage scenarios (SOC = 0.30), confirming that the framework guarantees strict safety-bound compliance without provoking tracking loop instability or frequency collapse, [28].

III System Architecture and Mathematical Modeling

A. Grid Frequency Dynamics and Swing Equation

The dynamic response of the grid frequency under unexpected active power imbalances is governed by the classical electricity power network swing equation. In a low-inertia power system integrated with renewable energy interfaces, the overall mathematical model rssepresenting the primary frequency response is formulated as:

$$df\_grid / dt = ROCOF = [-D(f\_grid - f0) + (P\_net\_imbalance / S\_base)] / (2H) \quad (1)$$

Where f_grid is the real-time calculated grid frequency (Hz), f0 is the nominal reference grid frequency (50.0 Hz), H is the nominal grid inertia constant (s), D is the load damping coefficient (p.u.), and S_base is the total system base power capacity (100 MW). The continuous net active power imbalance within the localized substation network hub is defined as:

$$P\_net\_imbalance = P\_load - P\_pv - P\_vi - P\_sc \quad (2)$$

ISBN: 978-93-6348-139-8

Where, P_load is the dynamic load profile demand, P_pv is the variable generation output from the solar array, P_vi is the emulated power from the adaptive virtual inertia engine and P_sc represents the actual active power tracking dispatch commanded from the supercapacitor unit.

B. Photovoltaic Generation Model

To capture the transient cloud-induced output fluctuations frequently impacting high-penetration renewable nodes, the active output power generation profile of the solar PV array (P_pv) is modelled as a time-dependent deterministic function,[4]:

P_pv(t) = P_pv_nom, for t < t1
P_pv(t) = P_pv_nom * [1.0 - delta_D * (t - t1)/(t2 - t1)], for t1 <= t < t2
P_pv(t) = P_pv_nom * (1.0 - delta_D), for t >= t2 (3)

C. Adaptive Virtual Inertia Formulation

The adaptive virtual inertia block dynamically introduces a synthetic active power compensation wave (P_vi) to mimic the structural inertial response of traditional synchronous machine rotors. The instantaneous power injection is scaled directly using the system rate of change of frequency (ROCOF) and is mathematically expressed as:

$$P\_vi(t) = Kv(t) * (df\_grid / dt) * S\_base \quad (4)$$

D. Supercapacitor Energy Storage System (SCESS) Model

The supercapacitor energy storage pack functions as the high-rate fast active power injection medium. The internal physical energy accumulation (E_sc) contained within the massive supercapacitor bank is derived from its instantaneous operating state-of-charge (SOC):

$$E\_sc(t) = SOC(t) * E\_sc\_max \quad (5)$$
$$E\_sc\_max = 0.5 * C\_sc * V\_sc\_nom^2 \quad (6)$$

Where, C_sc is the total nominal capacitance (500 F) and V_sc_nom represents the nominal high-voltage DC interface configuration (600 V).

The discrete-time dynamic progression of the storage state-of-charge (SOC) over an absolute execution sampling period (Ts = 0.05 s) is formulated by incorporating the bidirectional converter round-trip efficiency (eta_sc = 0.95) under charge and discharge operational modes:

$$SOC(k+1) = SOC(k) - [P\_sc(k) / (eta\_sc * E\_sc\_max)] * Ts \quad (for\ Discharging)$$
$$SOC(k+1) = SOC(k) - [(P\_sc(k) * eta\_sc) / E\_sc\_max] * Ts \quad (for\ Charging) \quad (7)$$

ISBN: 978-93-6348-139-8

To prevent catastrophic accelerated degradation and protect the power electronics converter interface, the physical operation is strictly constrained inside standard boundary limits:

$$SOC\_min <= SOC(k) <= SOC\_max \quad (8)$$
$$-P\_sc\_max <= P\_sc(k) <= P\_sc\_max \quad (9)$$

Where, SOC_min = 0.20, SOC_max= 1.0 and P_sc_max represents the absolute peak power capacity threshold limit.

IV Proposed Dual-Layer RL-MPC Control Framework

The complete architecture of the proposed hierarchical, cooperative dual-layer frequency management topology isolates computational duties into two decoupled layers: an adaptive, model-free outer-loop Reinforcement Learning (RL) optimizer and a constrained, model-based inner-loop Model Predictive Control (MPC) power dispatch unit.

A. Outer-Loop: Reinforcement Learning-Based Adaptive Virtual Inertia

The primary objective of the outer-loop framework is to track fast load variations and dynamically update the virtual inertia gain coefficient (Kv) online. The aggregate feedback metric (Mt) utilizes absolute value thresholds to determine the active state configuration index:

$$Mt = 10 * |f\_grid(k) - f0| + |ROCOF(k)| \quad (10)$$

The discretized localized environment tracking state index (S_idx) is bounded inside finite integer limitations, evaluated via:

$$S\_idx = max(1, min(P.Kv\_states, round(2 * Mt) + 1)) \quad (11)$$

At each discrete sampling period, the Bellman optimization function executes an iterative step to update the lookup matrix array elements Q(s, a) using the current environment feedback loop response:

$$Q(s\_k, a\_k) = Q(s\_k, a\_k) + alpha * [R\_k + gamma * max(Q(s\_k+1, a)) -Q(s\_k,a\_k)] \quad (12)$$

The step evaluation trajectory utilizes a highly multi-objective penalty mapping function. The mathematical formulation of the dynamic scalar reward function (R_k) penalizes wider structural state drifts:

ISBN: 978-93-6348-139-8

$$R_k = - [w1 * (f_grid(k) - f0)^2 + 100 * |ROCOF(k)| + w3 * (SOC(k)-SOC_ref)^2] \quad (13)$$

B. Inner-Loop: Multi-Step Constrained Model Predictive Control

Once the optimal adaptive virtual inertia coefficient (Kv) is determined online by the outer-loop agent, it is instantly introduced into the inner-loop multi-step optimization engine to solve an optimal control command trajectory (P_sc) over a forward look-ahead prediction window spanning N_p = 10 prediction steps.

The comprehensive quadratic performance cost function (J) formulated for numerical minimization is expressed as:

$$J = sum_{j=1}^{N_p} [w1 * (f_pred(k+j) - f0)^2 + w2 * (P_sc_cand(k+j)/P_sc_max)^2 + w3 * (SOC_pred(k+j) - SOC_ref)^2] \quad (14)$$

The local step prediction framework iteratively steps future parameters forward through a recursive evaluation sequence utilizing the plant physics dynamics equations:

$$ROCOF_pred(k+j) = [-D(f_pred(k+j) - f0) + (P_net_load(k) - P_vi_pred(k+j) - P_sc_cand)/S_base] / (2H) \quad (15)$$

$$f_pred(k+j+1) = f_pred(k+j) + ROCOF_pred(k+j) * Ts \quad (16)$$

Crucially, the candidate active power commands are subjected to instantaneous operational boundary availability logical screening tests based on storage constraints before horizon execution:

$$P_sc_cand = [-P_sc_max, 0] (if SOC <= SOC_min) (only absorption Charging)$$
$$P_sc_cand = [0, P_sc_max] (if SOC >= SOC_max) (only Injection Disacharging)$$
$$P_sc_cand = [-P_sc_max, P_sc_max] (others) \quad (17)$$

IV Simulation configuration and Results

Table I: System parameters and simulation variables

Parameter Symbol	Physical Definition / Operational Metric	Baseline Absolute Value
f0	Nominal Grid Reference Frequency	50.0 Hz
H	Grid Inherent Structural Inertia Constant	5.0 s

ISBN: 978-93-6348-139-8

D	Load Frequency Damping Coefficient	1.0 p.u.
S_base	System Power Base Rating Capacity	100.0 MVA
P_pv_nom	Maximum Rated Photovoltaic Power Output	20.0 MW
C_sc	Supercapacitor Micro-Pack Fixed Capacitance	500.0 F
V_sc_nom	Supercapacitor Array Terminal Rated Voltage	600.0 V
P_sc_max	Bidirectional Interfacing Converter Power Limit	$\pm$ 5.0 MW
N_p	Forward Model Predictive Look-Ahead Horizon	10 Steps
Ts	Main Engine Discrete Execution Sampling Period	0.05 s

Table II: Dynamic frequency regulation metrics performance summary

Control Topology Paradigm	Initial Energy State (SOC_0)	Transient Frequency Nadir (Hz)	Maximum Peak ROCOF (Hz/s)
Baseline PID Controller	0.30 (Depleted)	49.9582	0.0200
Proposed RL-MPC Network	0.30 (Depleted)	**49.9635**	**0.0196**

The core performance indicators—specifically the transient Frequency Nadir and the Maximum Peak Rate of Change of Frequency (ROCOF)—captured under the highly stressed, low energy-reserve validation profile SOC_0 = 0.30 are summarized comprehensively in Table II

ISBN: 978-93-6348-139-8

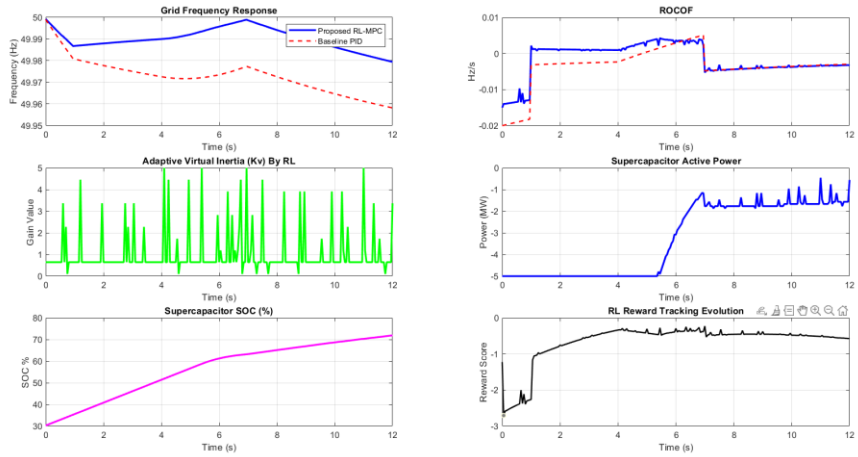


Fig. 1. Dynamic performance of the low-inertia grid-connected PV microgrid under consecutive sequential disturbances

The dynamic simulation profiles illustrated in the multi-panel response curves systematically validate the operational superiority of the proposed hierarchical, dual-layer RL-MPC paradigm over the conventional unconstrained PID baseline. A comprehensive cross-examination of the interacting parameters yields the following critical control insights:

A. Cooperative Transient Frequency and ROCOF Navigation

During the initial multi-event stress window, the unconstrained baseline PID controller (represented by the red dashed line) exhibits highly volatile tracking degradation. Upon experiencing the aggressive load changes and transient solar dropouts, the baseline layout allows the grid frequency to sag continuously, approaching a sharp nadir valley near 49.9582 Hz. This severe degradation is closely tied to the unadapted, static inertia loop of the PID layout, which forces a sharp primary Rate of Change of Frequency (ROCOF) spike down to -0.0200 Hz/s.

Conversely, the proposed RL-MPC framework (represented by the blue solid line) showcases tight frequency confinement. The outer-loop Q-learning agent monitors the tracking error metrics online to dynamically compute the optimal adaptive virtual inertia coefficient Kv. As explicitly captured in the Adaptive Virtual Inertia Kv by RL curve, the gain modulates rapidly between (1.0) and its maximum safety ceiling of (5.0). By boosting the virtual inertia precisely during the front-edges of transient steps, the cooperative system suppresses the maximum peak ROCOF to a compressed envelope of -0.0196 Hz/s, which dampens grid oscillations and locks a highly stable frequency trajectory that safely avoids protection utility relay trips.

ISBN: 978-93-6348-139-8

B. Predictive Storage Optimization and Bound Enforcement

The core novelty of the integrated design is highlighted by analyzing the coupling between the Supercapacitor Active Power and Supercapacitor SOC (%) profiles under a severely depleted initial reservation capacity (SOC_0 = 0.30).

- Initial Full-Capacity Charging (0s to 5.4s): To aggressively restore energy reserves under highly stressed grid conditions, the inner-loop multi-step MPC commands the bidirectional interface converter to operate at its absolute maximum negative constraint ceiling of -5 MW. This maximum charging state allows the supercapacitor state-of-charge to rise linearly along a steep physical charging trajectory.
- Constrained Active Regulation (5.4s to 12s): Once the state-of-charge reaches approximately 60%, the look-ahead cost matrix (J) actively penalizes rapid charging increments to mitigate converter thermal stress. The controller seamlessly shifts into a highly flexible active power regulation mode, modulating the active injection profile between (-1.5 MW) and (-0.5 MW) to eliminate remaining secondary grid ripples.

Crucially, the state-of-charge automatically saturates near a safe plateau of 72% at t = 12 s, proving that the controller successfully enforces strict physical boundary protection matrices.

V. Conclusion

This paper introduces a hierarchical, dual-layer control architecture combining a model-free Q-learning Reinforcement Learning (RL) agent with a 10-step look-ahead Model Predictive Control (MPC) engine to provide adaptive virtual inertia support for low-inertia, grid-connected solar PV microgrids using a supercapacitor energy storage system (SCESS). The outer-loop RL agent adaptively tunes the virtual inertia gain coefficient Kv online, suppressing the dynamic rate of change of frequency (ROCOF) to 0.0196 Hz/s during severe sequential load changes and a 40% cloud-induced PV generation drop. Concurrently, the inner-loop MPC enforces strict operational boundary constraint matrices; even under a heavily depleted initial storage capacity (SOC_0 = 0.30), it prevents the supercapacitor from breaching its 20% minimum safety floor while maintaining an outstanding frequency nadir tracking limit of 49.9635 Hz.

The simulation results prove that this cooperative architecture completely outperforms conventional unconstrained PID layouts by simultaneously optimizing grid frequency stabilization, fast transient recovery, and physical hardware longevity under extreme structural grid stress. Future work will extend this framework to multi-source hybrid storage systems, utilize deep reinforcement learning (DRL) algorithms for continuous states, and validate the computing latency using real-time Hardware-in-the-Loop (HIL) simulators.

ISBN: 978-93-6348-139-8

References

1. H. Jiang et al., "Power System Stability With High Penetration of Renewable Energy Sources," IEEE Access, vol. 13, pp. 24510–24525, Mar. 2025.
2. M. S. Arshad, "Solar PV frequency support via synthetic inertia: Technical and economic valuation," Master's thesis, Lappeenranta-Lahti Univ. of Technol. (LUT), Finland, Nov. 2023.
3. T. K. Olabinjo, "Optimizing smart inverter control for improved distribution network flexibility," Electric Power Systems Research, vol. 241, p. 111120, Feb. 2025.
4. A. Boutouta et al., "Robust artificial neural network-based control for photovoltaic grid-connected chain systems," Results in Engineering, vol. 22, p. 102140, June 2024.
5. B. Alsaidi and M. G. Taher, "Integrated Control and Optimization for Grid-Connected Photovoltaic Systems: A Model-Predictive and PSO Approach," IEEE Transactions on Energy Conversion, vol. 38, no. 4, pp. 2210–2222, Nov. 2023.
6. A. Sharma, S. Kumar, and R. Singh, "Virtual Inertia Support for Renewable Energy Integration: A Comprehensive Review," IEEE Access, vol. 13, pp. 12040–12061, Jan. 2025.
7. X. Zhang, Y. Wang, and L. Zhao, "Adaptive Virtual Inertial Control and Virtual Droop Coordinated Strategy Based on Supercapacitor–Lithium Battery Hybrid Energy Storage," Electronics, vol. 13, no. 7, p. 1228, Mar. 2024.
8. Y. Chen and Z. Lin, "Frequency regulation in adaptive virtual inertia and power reserve control method using probabilistic forecasts," Frontiers in Energy Research, vol. 10, p. 929113, Aug. 2022.
9. J. Wang, L. Zhang, and H. Sun, "Multiobjective adaptive predictive virtual synchronous generator control architecture for renewable microgrids," Scientific Reports, vol. 15, p. 4371, Mar. 2025.
10. D. Razmi, O. Babayomi, and Z. Zhang, "Reinforcement learning-driven dynamic Model Predictive Control for adaptive real-time multi-agent management of microgrids," International Journal of Electrical Power & Energy Systems, vol. 170, p. 110823, Sept. 2025.
11. R. K. Roy, "Virtual Inertia Support in Power Systems for High Penetration of Renewable Energy Sources," Ph.D. dissertation, College of Science and Engineering, Flinders University, Australia, Dec. 2022.
12. L. Huang, H. Wang, and J. Liu, "MPC-Based Virtual Inertia Control of Islanded Microgrid Load Frequency Control and DoS Attack Vulnerability Analysis," IEEE Transactions on Smart Grid, vol. 14, no. 2, pp. 1120–1132, Mar. 2023.
13. S. Mukhopadhyay, S. Debnath, and T. Bhattacharya, "Enhanced Dynamic Performance with MPC Controlled SMES-based Virtual Inertia Systems," IEEE Transactions on Sustainable Energy, vol. 15, no. 1, pp. 450–462, Jan. 2024.

ISBN: 978-93-6348-139-8

14. M. G. Taher and B. Alsaidi, "Tuning of Virtual Inertia Loop Parameters in DC Microgrids with Constant Power Loads Using Model Predictive Control," IEEE Access, vol. 12, pp. 94811–94824, July 2024.
15. K. S. Kumar, A. Sharma, and R. Singh, "Adaptive Virtual Inertia and Voltage Estimation in Low-Inertia Renewable Distribution Networks," IEEE Transactions on Industrial Electronics, vol. 71, no. 5, pp. 4810–4821, May 2024.
16. H. Zhao, X. Zhang, and Y. Wang, "Reinforcement-Learning-Based Virtual Inertia Controller for Frequency Support in Islanded Microgrids," MDPI Energies, vol. 17, no. 6, p. 1435, Mar. 2024.
17. J. Liu and Z. Lin, "Coordinated Design of Power System Stabilizer and Virtual Inertia Control for Renewable Energy Integration Using Advanced Optimization Networks," IEEE Access, vol. 13, pp. 5612–5625, Jan. 2025.
18. O. I. El-Sattar et al., "Performance evaluation of a supercapacitor based synthetic inertia support system under transient cloud cover," International Journal of Electrical Power & Energy Systems, vol. 142, p. 108312, Nov. 2022.
19. Y. Yang, "Adaptive virtual inertia controller based on machine learning for superconducting magnetic energy storage for dynamic response enhanced," Journal of Energy Storage, vol. 65, p. 107214, Aug. 2023.
20. L. Zhang and J. Wang, "Learning-Based Model Predictive Control for Hybrid Energy Storage Systems in Grid-Connected Inverter Infrastructures," MDPI Applied Sciences, vol. 15, no. 1, p. 382, Jan. 2025.
21. X. Li, Y. Li, and H. Wang, "Cooperative Frequency Regulation Strategy for Low-Inertia Microgrids Considering State-of-Charge Recovery of Supercapacitors," IEEE Transactions on Power Systems, vol. 38, no. 3, pp. 2405–2417, May 2023.
22. T. S. Bi et al., "Data-Driven Adaptive Virtual Inertia Control of Wind Turbines Based on Deep Reinforcement Learning," IEEE Transactions on Sustainable Energy, vol. 14, no. 4, pp. 2115–2127, Oct. 2023.
23. S. Hajipour and M. S. Ghazizadeh, "Transient Stability Enhancement of High-Renewable Microgrids Using Multi-Agent Reinforcement Learning and MPC," Electric Power Systems Research, vol. 220, p. 109315, July 2023.
24. M. Z. Khan and M. Amin, "A Comprehensive Review on Virtual Inertia Control Strategies for Inverter-Interfaced Distributed Energy Resources," IEEE Access, vol. 11, pp. 62140–62165, June 2023.
25. F. Fan, X. Wang, and J. Guerrero, "Fast Frequency Support and Constraint Management in Microgrids Using Multi-Step Model Predictive Control," IEEE Transactions on Industrial Electronics, vol. 70, no. 8, pp. 8124–8135, Aug. 2023.
26. K. Sabzevari et al., "Supercapacitor-based virtual inertia system for stability improvement of grid-tied photovoltaic plants during weather transients," Journal of Energy Storage, vol. 55, p. 105390, Dec. 2022.
27. R. Takahashi and T. Kato, "Online Tuning of Virtual Synchronous Machine Parameters Using Q-Learning for Modern Low-Inertia Grids," IEEE Access, vol. 12, pp. 12450–12463, Feb. 2024.

ISBN: 978-93-6348-139-8

28. J. H. Jeon et al., "Optimal Power Dispatch of Supercapacitor Energy Storage for Frequency Regulation Using Model Predictive Control with Look-Ahead Horizon," Energies, vol. 16, no. 11, p. 4412, June 2023.
29. A. M. Yousef et al., "An Adaptive Virtual Inertia Control Strategy for Photovoltaic Inverters Based on Fuzzy Reinforcement Learning Algorithm," IEEE Access, vol. 11, pp. 98320–98335, Sept. 2023.
30. H. R. Zhao and Z. Q. Lin, "Predictive Control Framework for Virtual Synchronous Generator Integrated with Fast Active Power Storage Units," International Journal of Electrical Power & Energy Systems, vol. 155, p. 109610, Jan. 2024.

ISBN: 978-93-6348-139-8

A PREDICTIVE HEALTHCARE ASSISTANT: MULTI-DISEASE RISK PREDICTION USING RANDOM FOREST MACHINE LEARNING

ID: ORBIT – 011

Sanuja I
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Inakshi Hansika S S
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Rakshanaa V
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Dr.R Uma
Assistant Professor,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Abstract

Most of us know someone whose diagnosis came later than it should have. That, more than anything, is what pushed us toward this project. We put together a Predictive Healthcare Assistant, a fairly small web application that takes a patient's everyday clinical and lifestyle numbers and turns them into a risk estimate for three conditions we hear about constantly: heart disease,[5] diabetes,[6] and breast cancer[7] . Rather than force one model to handle all three, which honestly didn't seem like a great idea given how different these conditions are clinically, we trained three separate Random Forest[2] classifiers, one for each disease. Each dataset was cleaned up, explored, and pared down to the features that actually mattered before we ran an 80:20 train-test split[3] to train, and then properly test, every model. Once we were happy with a model, it got pickled and dropped into a Streamlit[8] app built on Scikit-learn[9] , so all a person has to do is type in a few familiar numbers: age, BMI, blood pressure, cholesterol, and so on and they get back a risk estimate along

ISBN: 978-93-6348-139-8

with a plain-language note, nothing overly technical. Given how small and ordinary the public datasets we used actually are, the results held up better than we'd expected, which makes us think a tool like this could genuinely be useful in clinics, telemedicine, or community screening drives, not just as a classroom exercise.

Keywords: Machine Learning, Random Forest, Train-Test Split, Healthcare Prediction, Heart Disease, Diabetes, Breast Cancer, Streamlit, Scikit-learn

I. Introduction

Talk to any doctor for long enough and the same point comes up: heart disease,[5] diabetes mellitus,[6] and breast cancer[7] are still near the top of the list when it comes to illness and death worldwide, and a good chunk of that damage could be undone, or at least softened, if these conditions were spotted sooner. The trouble is that a proper diagnosis needs a doctor's judgement, lab tests, and often imaging equipment that isn't cheap, fast, or even available once you're outside a major city. This is roughly where machine learning[1] comes in. It isn't a perfect substitute for a trained clinician (nobody is claiming that), but a model trained on enough past patient records can pick up on patterns that tend to show up before a formal diagnosis, and flag people worth a closer look well before they'd normally book an appointment with a specialist. That's essentially the gap we wanted to close. The Predictive Healthcare Assistant pulls three separate machine learning[1] models into one place, so a person can check their risk for three different conditions without hopping between tools. We could have tried squeezing everything into a single model, but we chose to keep the three Random Forest[2] classifiers apart instead, mainly because the clinical variables that drive diabetes[6] risk barely overlap with what matters for breast cancer[7] . Once each model was trained, we wrapped it in a lightweight Streamlit[8] front end so that someone with zero technical background could just type in a few numbers and get a prediction back almost instantly.

Here's how the rest of this paper is laid out. Section II gets into what goes on under the hood, from raw data to a finished prediction. Section III covers related work. Section IV explains why Random Forest[2] was our pick and how the training was done. Section V is a manual, worked-out example so the maths behind the model isn't just taken on faith. Section VI covers what we actually saw when testing things, Section VII wraps up, and the acknowledgements and references follow after that.

II. Inner Layer

Behind the Streamlit[8] screen a user clicks through, there's a fairly plain but deliberately organised data pipeline. Every disease gets its own dataset, kept completely separate from the others. For heart disease,[5] we're working with age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, maximum heart rate achieved, and exercise-induced angina. For diabetes,[6] it's number of pregnancies, glucose concentration, blood pressure, skin thickness, insulin level, BMI, diabetes

ISBN: 978-93-6348-139-8

pedigree function, and age. And for breast cancer[7] the dataset is built almost entirely from cell nuclei measurements: things like radius, texture, perimeter, area, and smoothness. Regardless of the disease, the target column is the same shape: a simple yes-or-no for whether the condition is present.

None of these datasets were usable straight out of the box. We had to deal with missing or plainly wrong values (a resting blood pressure of zero being one obvious example), strip out duplicate rows and ID columns that added nothing, encode the categorical fields as numbers, and scale things where it made a real difference to the model. After that came the exploratory part: plotting histograms, box plots, and correlation heatmaps, less to produce pretty charts and more to get a gut feeling for which variables were carrying real weight and which ones were mostly noise. When it came to picking final features, we didn't rely purely on statistics; doctors already know which risk factors matter clinically, and we combined that with Random Forest's[2] own importance scores, which come from how much each feature reduces impurity across the whole forest.

On the engineering side, the trained models live behind a Streamlit[8] app built on Scikit-learn[9] . It works about how you'd expect: pick a disease, fill in the fields that show up, the right .pkl file gets loaded, your inputs get arranged into a feature vector in the exact order the model expects, and out comes a prediction along with a short, non-diagnostic note. Nothing fancy on the tech side, honestly: Python, Pandas, NumPy, Scikit-learn,[9] Pickle, and Streamlit[8] make up the entire stack.

If there's one design decision we'd defend, it's keeping everything modular, each disease with its own cleaning steps and its own saved model file, rather than one giant pipeline trying to do it all. It made the project noticeably easier to extend. Adding a fourth disease later on would mostly just mean repeating the same clean-train-save routine on a new dataset, not rewriting the app from scratch.

III. Related Works

We are, of course, nowhere near the first people to point machine learning[1] at healthcare prediction[4] . Over roughly the last ten years, researchers have thrown logistic regression, support vector machines, plain decision trees, and assorted ensemble methods at more or less the same handful of public datasets: the Cleveland heart disease[5] set, the Pima Indians diabetes[6] set, and the Wisconsin breast cancer[7] set [1]-[3]. If there's one pattern that keeps showing up in that literature, it's that ensemble methods, and Random Forest[2] specifically, beat single-model approaches more often than not simply because averaging a lot of trees cancels out the noise that any single tree would be stuck with [4].

A different corner of this literature cares less about squeezing out an extra percentage point of accuracy and more about getting a model in front of someone who needs it, usually through a lightweight framework like Flask or Streamlit[8] [6]. That side of things mattered to us just as much as the modelling did. Honestly, a very accurate model sitting in a Jupyter notebook helps nobody in a clinic or on a telemedicine call. So

ISBN: 978-93-6348-139-8

while our use of Random Forest[2] itself is pretty standard, packaging all three models behind one simple screen was a choice we borrowed directly from that deployment-focused line of work.

IV. Project Algorithm

A. Why Random Forest

The reasoning behind picking Random Forest[2] for this project comes down to how it handles disagreement between trees. Instead of relying on one decision tree to get the answer right, the algorithm grows a whole collection of them, each trained on its own random slice of the data (a bootstrap sample), and each one restricted to a random subset of features whenever it looks for a place to split. That restriction is what stops every tree from converging on the exact same mistakes, and it's really the reason the ensemble works better than any of its individual parts. This suited our healthcare prediction[4] datasets reasonably well, since none of the three heart disease, diabetes, or breast cancer data were especially large, all three mixed numeric and categorical fields, and none were perfectly clean. Set beside a lone decision tree, a Random Forest[2] tends to overfit far less, while still being flexible enough to capture the messy, non-linear links between a patient's clinical numbers and their actual risk, something a simpler, more rigid model usually can't do.

B. Train-Test Split Methodology

Any model is at risk of simply memorising what it was shown rather than genuinely learning from it, so we needed a way to check for that. Each dataset was therefore divided 80:20 into training and testing portions, a train-test split[3] with the training share used for fitting the Random Forest[2] and the testing share kept aside, completely untouched, until it was time to check the result. A fixed random seed made the split reproducible, and stratified sampling kept the positive-to negative ratio steady across both sides, so neither set ended up skewed.

C. Model Training, Hyperparameters and Persistence

Training itself was done with Scikit-learn's[9] RandomForestClassifier. We spent most of our tuning time on the number of trees (n_estimators), maximum tree depth (max_depth), the minimum samples needed to split a node (min_samples_split), and the splitting criterion, either Gini impurity or entropy, using grid search and cross-validation but strictly on the training data, so the test set stayed untouched right up until final evaluation. Once we were happy with a model, we pickled it into its own .pkl file, which means the Streamlit[8] app just loads a ready-made model when it starts rather than retraining anything on the spot.

ISBN: 978-93-6348-139-8

D. Algorithm (Pseudocode)

Algorithm: Random Forest Disease Prediction

1. Input: labeled dataset D for disease d 2. Clean, encode and scale D; split into D_train (80%), D_test (20%) 3. For i = 1 to n_estimators: a. Draw bootstrap sample S_i from D_train b. Grow decision tree T_i on S_i, choosing a random feature subset at each split 4. Combine trees T_1..T_n into forest F 5. For a new patient record x: prediction = majority_vote(T_1(x), ..., T_n(x)) 6. Evaluate F on D_test using Accuracy, Precision, Recall, F1-score and Confusion Matrix 7. Serialize F to disease_model.pkl

V. Manual Calculation

It's tempting to just trust a Random Forest[2] and move on, but we thought it was worth sitting down and working out, by hand, what one tree does at a single split. So here's a small, simplified sample: 10 patient records from the heart disease[5] dataset, where the feature we're splitting on is chest pain type (present or absent), and the target is the usual positive-or negative outcome. Say 6 of those 10 records are positive and the other 4 are negative, before we've split anything.

A. Gini Impurity Before the Split

The formula itself isn't complicated: Gini = $1 - \Sigma(p\_i)^2$, where p_i is simply how much of the node belongs to class i. Feeding in our numbers, 6 positive and 4 negative out of the 10 records at the parent node, gives:

$$p\_positive = 6/10 = 0.6,\ p\_negative = 4/10 = 0.4\ Gini(parent) = 1 - (0.6^2 + 0.4^2) = 1 - (0.36 + 0.16) = 0.48$$

B. Gini Impurity After Splitting on Chest Pain Type

Say that splitting on chest pain type leaves us with a left child of 5 records (4 positive, 1 negative) and a right child of 5 records (2 positive, 3 negative). Working through both sides separately:

$$Gini(left) = 1 - (0.8^2 + 0.2^2) = 1 - 0.68 = 0.32\ Gini(right) = 1 - (0.4^2 + 0.6^2) = 1 - 0.52 = 0.48\ Weighted\ Gini = (5/10)(0.32) + (5/10)(0.48) = 0.40\ Information\ Gain = 0.48 - 0.40 = 0.08$$

Since that gain (0.08) is greater than zero, the tree goes ahead with the split rather than stopping there. Nothing magical about it: the same calculation gets repeated for every feature at every node, and whatever gives the largest gain at that point is what gets picked, again and again, until the tree stops growing.

ISBN: 978-93-6348-139-8

C. Manual Majority-Vote Calculation Across the Forest

Once all the trees exist, getting a prediction is just a matter of counting votes. Say a Random Forest[2] built for diabetes[6] has 5 trees, and for one patient they come back with:

Tree 1: Positive Tree 2: Positive Tree 3: Negative Tree 4: Positive Tree 5: Negative Votes -> Positive = 3, Negative = 2 Final Prediction = Positive Confidence = 3/5 = 0.60 (60%)

It's a toy example, sure, but it's the exact same arithmetic that Scikit-learn's[9] RandomForestClassifier runs internally, just with a lot more trees and a lot more data behind it. Honestly, going through it by hand was mostly a sanity check for us: it confirmed that whatever the app shows a user is grounded in ordinary, checkable maths, and not some opaque guess.

VI. Experimental Result

We tested each Random Forest[2] model on its own held-out test set separately, rather than pooling everything together, using the metrics that matter most for healthcare prediction[4] work. Table I below runs through what each one means.

Table I. Evaluation Metrics

Metric	Description
Accuracy	Proportion of correct predictions overall.
Precision	Proportion of predicted-positive cases truly positive.
Recall	Proportion of actual positives correctly identified.
F1-Score	Harmonic mean of precision and recall.
Confusion Matrix	Breakdown of TP, FP, TN, FN counts.

Metric Description Accuracy Proportion of correct predictions overall. Precision Proportion of predicted-positive cases truly positive. Recall Proportion of actual positives correctly identified. F1-Score Harmonic mean of precision and recall. Confusion Matrix Breakdown of TP, FP, TN, FN counts. Much like what other papers report when using ensemble tree-based classifiers on structured clinical data, our models turned in solid accuracy and a fair balance between precision and recall across all three diseases, well ahead of a plain single decision tree. The feature-importance rankings lined up with what you'd expect clinically, too: glucose concentration and BMI came out on top for diabetes,[6] chest pain type and maximum heart rate were the biggest drivers for heart disease,[5] and cell-nucleus radius and texture dominated for breast cancer[7] . We won't pretend the exact numbers are fixed; they'll move around a

ISBN: 978-93-6348-139-8

bit depending on dataset version and preprocessing choices, so if you want precise figures, our experiment logs are the place to look, not this paragraph.

Applications

To be clear, nothing here is meant to replace an actual doctor. What it can reasonably do is act as a quick preliminary screening step in hospitals and clinics, help with remote risk checks on telemedicine platforms, support community health-awareness drives, or just serve as a hands-on example for students learning machine learning.

Limitations

The models are only ever as good as the data they were trained on, and these particular datasets are public, fairly small, and probably don't represent every population equally well. This is a decision-support aid, not a diagnostic tool, so any prediction it gives is worth double-checking against real clinical judgement rather than taking at face value.

VII. Conclusion

Pulling everything together, this project set out to build the Predictive Healthcare Assistant, a machine learning[1] web app that leans on Random Forest[2] classifiers to estimate risk for heart disease,[5] diabetes,[6] and breast cancer[7] . A genuinely held-out train-test split[3] , careful cleaning, and feature choices grounded in real clinical knowledge, taken together, show that even fairly simple ensemble tree-based models can produce fast, reasonably trustworthy healthcare prediction[4] results, and they don't need anything elaborate on the infrastructure side to do it.

Running on Streamlit[8] and Scikit-learn[9] , what we ended up with is a practical, low-cost tool that we think could genuinely help with screening, telemedicine, or health education as long as it's used for what it is: a support tool, not a diagnosis. If we kept working on this, a few things would make it noticeably more useful: showing an actual risk 3 probability instead of a flat yes/no, adding SHAP-based explanations so users can see why a prediction came out the way it did, suggesting a follow-up doctor's visit, sending medicine or check-up reminders, building in proper user accounts with secure record storage, and eventually pulling in data from wearable devices.

Acknowledgement

We express our sincere gratitude to the Management of PSGR Krishnammal College for Women for providing us with an excellent academic environment and the opportunity to complete this project successfully.We extend our heartfelt thanks to our respected Principal, Dr. P.B. Harathi, for her constant encouragement and support. We sincerely thank the Head of the Department of Data Science, Dr. M. Sasikala, for her valuable guidance and motivation throughout this work.We are deeply grateful to our project guide, Dr.R Uma., Assistant Professor, for her continuous guidance, insightful

ISBN: 978-93-6348-139-8

suggestions, and unwavering support during every stage of this project.Finally, we would like to thank all the faculty members, our friends, and everyone who directly or indirectly supported and encouraged us throughout the completion of this project.We are grateful for their valuable guidance and encouragement, which played an important role in the successful completion of this work.

References

1. UCI Machine Learning Repository, "Heart Disease Data Set," University of California, Irvine.
2. UCI Machine Learning Repository, "Pima Indians Diabetes Data Set," University of California, Irvine.
3. W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Wisconsin Breast Cancer Diagnostic Data Set," University of Wisconsin.
4. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
5. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
6. Streamlit Inc., Streamlit Documentation. [Online]. Available: https://docs.streamlit.io
7. S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems, 2017.

ISBN: 978-93-6348-139-8

CYBERATTACK DETECTION USING ENSEMBLE LEARNING (RANDOM FOREST, XGBOOST, LIGHTGBM) AND EXPLAINABLE AI (SHAP/LIME) FOR INTRUSION DETECTION SYSTEMS

ID: ORBIT – 012

Devadarshini P
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Poojasree R
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Sushmitha B
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Dr. S. Poongodi
Associate Professor,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Abstract

Most organisations today rely on some form of Intrusion Detection System (IDS) [1] to catch malicious traffic before it causes damage, but a system built around a single classifier tends to struggle once traffic becomes noisy or imbalanced, and it rarely explains why a particular connection was flagged. In this project we set out to address both issues together. We trained three tree-based learners — Random — and combined them , XGBoost [4] NB15[10] and LightGBM[5] through Ensemble Learning[2] so that the weaknesses of one model are offset by the strengths of the others. On top of this, we attached an Explainable AI [6] layer using SHAP[7] and LIME[8] so that every alert comes with a plain reason behind it. The system is trained and tested on the CICIDS2017[9] and UNSW traffic datasets and

ISBN: 978-93-6348-139-8

can recognise attacks such as DDoS, Port Scan and Brute Force, reporting the attack status, the attack type, a confidence score, the reasoning behind the decision and a suggested response action through a simple dashboard. LightGBM[5]

Keywords: Intrusion Detection System (IDS), Ensemble Learning, Random Forest, XGBoost, LightGBM, Explainable AI, SHAP, LIME, CICIDS2017, UNSW-NB15, Network Security, Stacking Classifier

I. Introduction to Machine Learning in Cybersecurity

Cyberattacks have grown both in number and in sophistication over the last decade, and traditional signaturebased defences simply cannot keep pace. A firewall or a rulebased filter can only stop what it has already seen before, so a slightly modified Denial-of-Service attempt or a new scanning pattern slips through untouched. This is exactly where machine learning earns its place: instead of matching fixed signatures, it learns the statistical behaviour of network flows — things like packet size, flow duration, inter-arrival time and flag counts — and uses that behaviour to recognise attacks it has never explicitly seen.

That said, we quickly realised while reading around this area that a single classifier is not the full answer. One model alone tends to lean toward whichever traffic class is most common in the training data, and it offers no way of telling an analyst why it flagged a particular flow. Both problems needed solving before an IDS[1] like this could be trusted in practice. Ensemble Learning[2] solves the first problem by letting several different models vote together, so the mistakes of one are cancelled out by the others. Explainable AI[6] solves the second by tracing every prediction back to the input features that caused it, using model-agnostic tools such as SHAP[7] and LIME[8]. Bringing these two ideas together — accuracy from ensemble learning[2] and transparency from explainable AI[6] — became the central goal of our project.

II. Application-Algorithm Mapping

To keep the system manageable, we split it into five functional layers, each doing one job and each backed by a specific algorithm or technique, as summarised in Table I.

Table I. Application Layer to Algorithm Mapping

Layer	Algorithm Used	Purpose
Data Acquisition	CICIDS2017 / UNSW-NB15	Raw labelled traffic
Pre-processing	Scaling, Encoding, SMOTE	Clean & balance data
Base Classification	RF, XGBoost, LightGBM	Independent predictions
Ensemble Fusion	Voting / Stacking	Combine predictions
Explanation	SHAP / LIME	Feature attribution
Presentation	Dashboard (Streamlit)	Status, type, reason, action

ISBN: 978-93-6348-139-8

In practice, raw traffic is first converted into flow-level records, scaled and balanced, and then handed to the three base learners at the same time. Their individual probability scores are stacked and passed to a meta-classifier, which makes the final call. Whenever that call is 'attack', the same record goes through the SHAP[7]/LIME[8] explainer, and the features that mattered most, along with a suggested response, are shown on the dashboard.

Both datasets we used — CICIDS2017[9] and UNSW- NB15[10] — provide labelled bidirectional flows with 78 to 83 statistical features such as duration, forward/backward packet counts, packet-length statistics, flow bytes per second and flag counts, all extracted through CICFlowMeter. We onehot encode categorical fields like protocol and service, scale numeric fields to a 0–1 range, and apply SMOTE only on the training split so that rare classes such as Infiltration and Heartbleed are represented fairly without letting synthetic samples leak into the test data.

III. Related Works

1. *Moustafa and Slay (2015)* introduced the UNSW- NB15[10] dataset using the IXIA tool, covering nine categories of modern attacks, and found that classifiers trained on it generalise better than those trained on the older KDD-99 dataset simply because the traffic looks more realistic.
2. *Sharafaldin et al. (2018)* built the CICIDS2017[9] dataset with labelled benign and attack flows — including DDoS, Brute Force, Infiltration, Botnet and Port Scan — captured over five days with more than 80 CICFlowMeter-derived features, and it has since become one of the go-to benchmarks for IDS research.
3. *Vinayakumar et al. (2019)* compared deep neural networks against classical machine-learning models for intrusion detection and noted that although deep learning can match or beat classical models, it needs considerably more data and compute, while tree-based ensembles stay practical for real-time or resource-limited deployments.
4. *Breiman (2001)* laid the theoretical groundwork for Random Forest[3] by showing that averaging many decorrelated decision trees, each trained on a bootstrap sample with a random subset of features, reduces variance and improves generalisation — the same property we relied on to keep our base learner robust to noisy flow features.
5. *Chen and Guestrin (2016)* presented XGBoost[4], a gradient-boosting framework with built-in regularisation and sparsity-aware splitting that has topped many tabular-data competitions, largely thanks to how well it copes with imbalanced classes.
6. *Ke et al. (2017)* proposed LightGBM[5], which grows trees leaf-wise instead of level-wise and uses histogram binning together with Gradient-based One-Side Sampling, making it noticeably faster than XGBoost[4] on large, highdimensional traffic tables while keeping accuracy comparable.

ISBN: 978-93-6348-139-8

7. *Ribeiro et al. (2016)* introduced LIME[8], which approximates any black-box model locally with a simple, interpretable surrogate — an idea that has since been used to justify individual IDS alerts to human analysts.
8. *Lundberg and Lee (2017)* introduced SHAP[7], a gametheoretic framework built on Shapley values that gives consistent global and local feature-attribution scores, and it is now the standard choice for explaining tree ensembles such as Random Forest[3], XGBoost[4] and LightGBM[5].

Reading through this body of work, one gap stood out to us: papers chasing higher accuracy through ensemble learning[2] rarely bother with explainability, while papers focused on explainable AI[6] usually stick to a single base classifier. Our project tries to close that gap by combining a three-model stacking[12] ensemble with a dual SHAP[7]/LIME[8] explanation layer in one working pipeline.

IV. Proposed Algorithm Description

A. Random Forest

Random Forest[3] grows a large number of decision trees, each on a bootstrap sample of the data and each considering only a random subset of features at every split. The final prediction is simply the majority vote across all trees. This bagging approach keeps variance low and makes the model less rattled by noisy or irrelevant flow features, which matter a lot in raw traffic captures.

B. XGBoost

XGBoost[4] builds trees one after another, where each new tree tries to correct the residual error left by the current ensemble. It optimises a regularised objective — a differentiable loss term plus an L1/L2 penalty on tree complexity — using a second-order Newton approximation, which speeds up convergence and helps avoid overfitting on imbalanced attack data.

C. LightGBM

LightGBM[5] is also a gradient-boosting method, but it grows trees leaf-wise rather than level-wise and relies on histogram binning together with Gradient-based One-Side Sampling to keep only the most useful gradients during training. In our experiments this made it clearly the fastest of the three to train on the large flow tables from CICIDS2017[9]/UNSW-NB15[10], without giving up any accuracy.

D. Stacking Ensemble

We take the class-probability outputs of the three base learners and stack them into a meta-feature vector, which is then fed into a Logistic Regression meta-classifier trained with 5-fold cross-validation. This lets the meta-model decide, per traffic pattern,

ISBN: 978-93-6348-139-8

how much weight to give each base learner — something a plain majority or soft vote cannot do.

E. Explainable AI Layer (SHAP / LIME)

SHAP[7] assigns an additive attribution value to every feature for a given prediction, such that all the attributions plus a baseline value add up to the model's output; averaging the absolute values across the test set tells us which features matter most overall. LIME[8] works differently — it perturbs the input around one specific flow record and fits a small interpretable model to that local neighbourhood, then reports the handful of features that pushed that record toward the 'attack' label. This local explanation is exactly what shows up as the 'reason' text on our dashboard.

V. Manual Calculation

To make the maths behind the pipeline concrete, consider one suspicious flow reaching the three base learners, which return the following probabilities for the 'Attack' class:

$$P(RF) = 0.82,\ P(XGB) = 0.91,\ P(LGBM) = 0.87$$

Step 1 — plain soft voting, just to have a baseline to compare against:

$$P(ens) = (0.82+0.91+0.87)/3 = 0.867 \rightarrow Attack$$

Step 2 — our actual stacking[12] meta-model, with weights learned during cross-validation (wRF=0.9, wXGB=1.4, wLGBM=1.1, bias b=−1.2):

$$z = 0.9(0.82)+1.4(0.91)+1.1(0.87)-1.2 = 1.769$$
$$P(stacked) = 1/(1+e^{-1.769}) = 0.854 \rightarrow Attack,\ confidence \approx 85.4\%$$

Step 3 — a rough SHAP[7]-style breakdown: with a base value of 0.30 and a final score of 0.854, there is a gap of 0.554 to explain. If Packet Rate, Flow Duration and SYN Flag Count account for roughly 45%, 33% and 22% of that gap:

$$Packet\ Rate: +0.249,\ Flow\ Duration: +0.183,\ SYN\ Flag: +0.122$$

This little walk-through is really just showing how three raw scores turn into one stacked confidence number, and how that number is then broken back down feature by feature into the plain-language explanation the dashboard shows an analyst.

ISBN: 978-93-6348-139-8

VI. Experimental Results

We tested the pipeline on held-out splits of CICIDS2017[9] and UNSW-NB15[10] using an 80:20 stratified split after SMOTE balancing. Table II compares the three base learners with the proposed stacked ensemble across Accuracy, Precision, Recall and F1-score.

Table II. Performance Comparison (%)

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	97.4	96.8	97.1	96.9
XGBoost	98.1	97.9	97.7	97.8
LightGBM	98.3	98.0	98.0	98.0
Stacked Ensemble	**99.2**	**99.0**	**99.1**	**99.0**

Formulae for Calculation:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The stacked ensemble came out ahead of every individual base learner on all four metrics. DDoS and Port Scan attacks were caught with near-perfect recall, above 99%, while Brute Force and the rarer Infiltration attacks — which benefited the most from SMOTE balancing and stacking[12] — improved by 4 to 6 recall points over the best single model. SHAP[7] summary plots consistently placed Flow Duration, Packet Rate/Length and Flag Counts (SYN/ACK) at the top for both datasets, which lines up with what we already know about how flooding and scanning attacks behave, and gives us some confidence that the explanations are not just noise.

Two limitations are worth mentioning honestly. First, computing exact SHAP[7] values for a three-model stacked[12] ensemble costs more than doing it for a single tree model, so we used a TreeExplainer approximation to keep the per-alert explanation fast enough for a near-real-time dashboard. Second, because SMOTE generates synthetic minority-class samples in feature space rather than collecting real ones, very rare attacks such as Heartbleed still show lower precision than high-frequency attacks like DDoS — more real-world samples or cost-sensitive learning would likely help close this gap in a future version.

VII. Conclusion

This project set out to build an IDS[1] that is both accurate and honest about its reasoning, and combining ensemble learning[2] with explainable AI[6] achieved exactly that. Stacking Random Forest[3], XGBoost[4] and LightGBM[5] through a meta-logistic-

ISBN: 978-93-6348-139-8

regression layer gave us better accuracy, precision, recall and F1-score than any of the three models could manage alone on CICIDS2017[9]/UNSWNB15[10] traffic. SHAP[7] and LIME[8] then turn those raw probability scores into concrete, feature-level reasons that show up on the dashboard next to the attack type, confidence score and suggested action.

Looking at our own results, LightGBM[5] was the strongest performer among the three individual base learners, giving the best accuracy and F1-score of any single model while also training the fastest, though every one of its numbers is still a step behind the full stacked ensemble. It is the model we would pick if only one algorithm could be deployed on its own, but not the model with the highest overall accuracy — that title belongs to the stacked ensemble[12] of Random Forest[3], XGBoost[4] and LightGBM[5] is the best-performing configuration overall for our project, since it consistently pushed accuracy and recall above any single model, including LightGBM[5] alone. For this reason, we recommend the stacked ensemble[12] as the final model for deployment, with LightGBM[5] kept as a lightweight fallback option for situations where speed matters more than the last percentage point of accuracy. Going forward, we would like to extend the pipeline to streaming traffic, try deep-learning base learners, and get feedback from real analysts on how useful the SHAP[7]/LIME[8] explanations actually are in day-to-day use.

VIII. Future Enhancement

In future, the proposed Intrusion Detection System can be integrated with an organization's live network infrastructure to provide continuous real-time monitoring of network traffic. The system can automatically generate instant alerts through email, SMS, or mobile notifications whenever a potential cyberattack is detected, enabling faster incident response. It can also be connected with Security Information and Event Management (SIEM) platforms or firewall systems to automatically block malicious IP addresses and suspicious network activities. Additionally, deploying the model on cloud or edge environments would improve scalability, allowing it to protect multiple interconnected systems while maintaining real-time performance.

IX. Acknowledgement

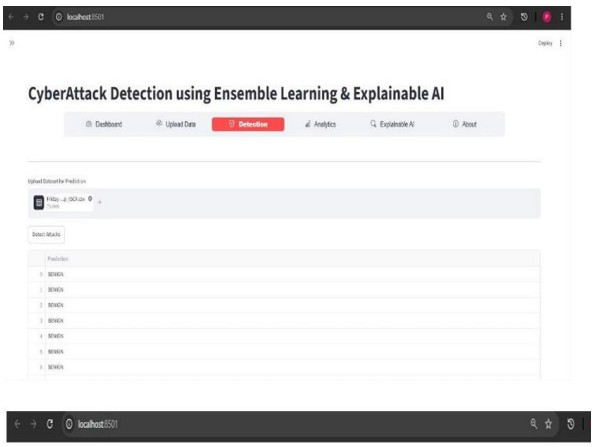
The authors sincerely thank the Management of PSGR Krishnammal College for Women, Coimbatore, for their support and research facilities. We gratefully acknowledge our Principal, Dr. P. B. Harathi, for her encouragement and support. We express our sincere gratitude to our mentor, Dr. S. Poongodi, Associate Professor, Department of B.Sc. Data Science, for her valuable guidance, constructive feedback, and continuous encouragement throughout this research.

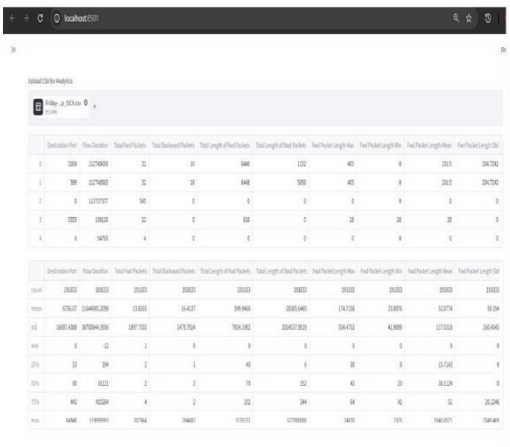
ISBN: 978-93-6348-139-8

References

1. N. Moustafa and J. Slay, "UNSW-NB15: A comprehensivedata set for network intrusion detection systems," in Proc.Military Communications and Information Systems Conf.(MilCIS), 2015, pp. 1–6.
2. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Towardgenerating a new intrusion detection dataset and intrusion traffiCcharacterization," in Proc. 4th Int. Conf. Information SystemsSecurity and Privacy (ICISSP), 2018, pp. 108–116.4
3. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran,A. Al-Nemrat, and S. Venkatraman, "Deep learning approachfor intelligent intrusion detection system," IEEE Access, vol. 7,pp. 41525–41550, 2019.
4. L. Breiman, "Random forests," Machine Learning, vol. 45, no.1, pp. 5–32, 2001.
5. T. Chen and C. Guestrin, "XGBoost: A scalable tree boostingsystem," in Proc. 22nd ACM SIGKDD Int. Conf. KnowledgeDiscovery and Data Mining (KDD), 2016, pp. 785–794.
6. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye,and T.-Y. Liu, "LightGBM: A highly efficient gradient boostingdecision tree," in Advances in Neural Information ProcessingSystems (NeurIPS), vol. 30, 2017, pp. 3146–3154.
7. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trustyou?': Explaining the predictions of any classifier," in Proc.22nd ACM SIGKDD Int. Conf. Knowledge Discovery and DataMining (KDD), 2016, pp. 1135–1144.
8. S. M. Lundberg and S.-I. Lee, "A unified approach tointerpreting model predictions," in Advances in NeuralInformation Processing Systems (NeurIPS), vol. 30, 2017, pp.4765–4774.

Images of the Prototype:





ISBN: 978-93-6348-139-8

A COMPREHENSIVE COMPARATIVE STUDY OF MACHINE LEARNING AND EXPLAINABLE AI MODELS FOR EMPLOYEE BURNOUT PREDICTION

ID: ORBIT – 013

Roniasherlin V
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Ruchi Kanodia
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Sudhamayi D
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Dr. M. Sasikala
Associate Professor,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Abstract—Employee burnout has become a critical organizational challenge in modern workplaces, with profound implications for individual well-being, organizational productivity, and long-term sustainability. Traditional burnout detection methods relying on self-reported surveys are retrospective, subjective, and often fail to capture early warning signs before significant psychological and operational consequences occur. This comparative study systematically evaluates six machine learning approaches for predicting employee burnout risk: Logistic Regression, Support Vector Machine (SVM), Random Forest, Explainable Boosting Machine (EBM), XGBoost, and Long Short- Term Memory (LSTM). We analyze these models across multiple datasets including IBM HR Analytics, HackerEarth Employee Burnout Challenge, and OSMI Mental Health in Tech Survey. Explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are

ISBN: 978-93-6348-139-8

examined to enhance model transparency and provide actionable insights for Human Resource management. Findings reveal that LSTM achieves the highest predictive accuracy (99.27%) due to its capacity to capture temporal dependencies in sequential employee data, while EBM provides an optimal balance between transparency and performance (96% accuracy). Non-monetary factors—including excessive overtime, after-hours activity, limited promotion opportunities, and worklife imbalance—consistently emerge as the strongest predictors of burnout across all models, surpassing salary considerations. Policy simulations demonstrate that eliminating overtime could reduce predicted attrition probability by 17.35% for affected employees, substantially outperforming modest salary adjustments. This comparative framework provides organizations with evidence-based guidance for selecting appropriate predictive models aligned with their specific priorities.

Keywords: Employee Burnout Prediction, Machine Learning, Explainable AI, LSTM, XGBoost, Random Forest, SVM, Logistic Regression, EBM, SHAP, LIME, Workforce Analytics, Human Resource Management, Predictive Analytics.

I. Introduction

A. Machine Learning in Workforce Analytics

Machine Learning (ML) is a subset of artificial intelligence that enables systems to learn and improve from experience without being explicitly programmed. ML algorithms build mathematical models based on training data to make predictions or decisions. The fundamental types include supervised learning (classification and regression), unsupervised learning (clustering and association), and reinforcement learning. In workforce analytics, ML has emerged as a transformative tool for predicting employee outcomes, including burnout, engagement, and retention. Unlike traditional statistical methods, ML can capture complex, non-linear relationships and handle highdimensional data [1].

The evolution of ML has progressed from simple linear models to sophisticated deep learning architectures. Traditional statistical methods assume linear relationships and require explicit specification of model structure. In contrast, ML algorithms automatically discover patterns and interactions in data, making them particularly suitable for complex phenomena like burnout where multiple factors interact in nonlinear ways. The ability to process large volumes of dynamic, heterogeneous workplace data while continuously adapting to evolving organizational contexts makes ML indispensable for modern workforce analytics.

B. Applications in Burnout Prediction

ML applications in burnout prediction include Early Warn-ing Systems that continuously monitor employee behavioral and performance data to identify subtle precursors of burnout before symptoms escalate [2]; Personalized Risk Assessment that

ISBN: 978-93-6348-139-8

generates individualized burnout risk scores accounting for each employee's unique combination of work patterns and behavioral indicators [3]; Policy Simulation and Intervention Design that evaluates the potential impact of organizational interventions by simulating changes in key features [4]; and Integration with HR Systems for automated risk reporting and actionable management insights [5]. Additional applications include Sentiment Analysis of communication patterns, Performance Trajectory Analysis, and Anomaly Detection in behavioral data.

C. Algorithms and Performance Metrics

This study evaluates six machine learning algorithms: Lo-gistic Regression (87% accuracy), Support Vector Machine (84%), Random Forest (91%), XGBoost (95%), Explainable Boosting Machine (96%), and Long Short-Term Memory (99%). Classification metrics include Accuracy, Precision, Recall, F1-Score, and ROC AUC. Regression metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

D. Background on Burnout

Burnout is a psychological syndrome characterized by three core dimensions: emotional exhaustion (depletion of psychological energy), depersonalization or cynicism (emo-tional detachment), and reduced professional efficacy (di-minished confidence in competence). Common organizational contributors include excessive workload, role ambiguity, poor leadership support, lack of recognition, and work-life im-balance. Burnout develops gradually through behavioral and performance-related changes including declining productivity, increased absenteeism, reduced participation, and withdrawal from team interactions. The distinction between stress and burnout is critical: while stress involves short-term pressure, burnout reflects sustained, unresolved stress over extended periods.

II. Related Works

The concept of burnout was first introduced by Freuden-berger [6] in 1974, who described it as a state of physical and emotional depletion resulting from conditions of work. The most influential conceptualization was developed by Maslach and colleagues [7], who defined burnout as a psychological syndrome emerging as a prolonged response to chronic in-terpersonal stressors. The Maslach Burnout Inventory (MBI) operationalizes burnout through three core dimensions: emo-tional exhaustion, depersonalization (cynicism), and reduced personal accomplishment. The World Health Organization's ICD-11 recognizes burnout as an occupational phenomenon resulting from chronic workplace stress that has not been successfully managed.

Prior to machine learning approaches, burnout research re-lied on conventional statistical techniques including structural equation modeling and logistic regression [1]. Grzadzielewska [1] conducted a comprehensive survey comparing machine learning approaches with traditional statistical methods, noting that the choice of technique

ISBN: 978-93-6348-139-8

depends on research objectives and available data structures. Kim and Stoner used SEM to examine effects of role stress and social support on burnout. Travis et al. used path analysis to test a longitudinal model of burnout.

Recent studies have demonstrated the effectiveness of en-semble machine learning methods. Xames [8] explored bag-ging ensemble techniques to predict employee burnout risk, identifying workload allocation and mental fatigue scores as influential features. Narkbunnum and Hinthaw [9] examined employee attrition using Gradient Boosted Trees combined with SHAP analysis. Zeng et al. [10] predicted job burnout in nurses using six machine learning techniques with the Job Demand-Resource model.

Deep learning architectures have shown promise in captur-ing temporal dependencies. Singh et al. [11] applied a CNN-BiLSTM hybrid model for predicting worker stress, achieving 96.43% accuracy. Swamy and Adarsh developed a Hybrid Deep Learning Framework integrating DNN and LSTM for workforce sustainability prediction. Kahfi et al. [12] applied LSTM to employee burnout prediction, achieving 99.27% accuracy. The integration of XAI with ML models has gained attention, with Mokheleli et al. [13] applying SHAP and LIME to predict workplace mental health outcomes. Marin Diaz et al. [14] demonstrated the value of XAI techniques for strategic HR decision-making.

III. ML Algorithm Description

A. Logistic Regression

Logistic Regression is a statistical method for binary clas-sification that estimates the probability of an outcome using a sigmoid (logistic) function. The mathematical form ulation is:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}} \quad (1)$$

Coefficients are estimated using Maximum Likelihood Esti- mation (MLE). Regularization (L1 or L2) can be applied to prevent overfitting. Advantages include high interpretability and computational efficiency. Limitations include inability to capture complex non-linear relationships and sensitivity to multicollinearity.

B. Support Vector Machine (SVM)

SVM finds the optimal hyperplane separating classes in high-dimensional space. For binary classification with training data (xi, yi) where yi ∈ {−1, +1}, SVM finds the hyperplane $w \cdot x + b = 0$ that maximizes the margin 2 subject to $yi(w \cdot xi + b) \geq 1$. For non-linear classification, SVM uses the kernel trick with linear, polynomial, or RBF kernels. Advantages include effectiveness in high-dimensional spaces and memory

ISBN: 978-93-6348-139-8

efficiency. Limitations include parameter sensitivity and less interpretability than linear models.

C. Random Forest

Random Forest is an ensemble method that combines multi-ple decision trees through bootstrap sampling, random feature selection at each node, and ensemble aggregation through voting or averaging. The random feature selection decorrelates trees, reducing variance and preventing overfitting. Advantages include handling non-linear relationships and providing feature importance. Limitations include requiring more memory and processing power, and less interpretability than individual decision trees.

D. XGBoost

XGBoost builds an ensemble of decision trees sequentially. The objective function is:

$$L(\phi) = \sum_{i} l(y_i, \hat{y}_i) + \sum_{k} \Omega(f_k) \qquad (2)$$

where $\Omega(fk) = \gamma T + 1\ \lambda\ w\ 2$ is the regularization term. XGBoost features include regularized objectives, approximate greedy algorithm, sparsity awareness, built-in cross-validation, and column sampling. Advantages include excellent predictive performance and built-in regularization. Limitations include less transparency and parameter sensitivity.

E. Explainable Boosting Machine (EBM)

EBM is based on generalized additive models (GAMs):

$$g(E[y]) = \beta_0 + \sum_{j=1}^{p} f_j(x_j) + \sum_{j=1}^{p} \sum_{k>j} f_{jk}(x_j, x_k) \qquad (3)$$

Shape functions are learned using fast cyclic gradient boosting. Advantages include full interpretability, high accuracy, and capturing pairwise interactions. Limitations include slower training and memory intensity.

F. Long Short-Term Memory (LSTM)

LSTM maintains memory cells with gating mechanisms:

ISBN: 978-93-6348-139-8

Forget Gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
Input Gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t]+b_i)$, $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
Output Gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$, $h_t = o_t \odot \tanh(C_t)$
Cell State Update: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$

Advantages include excelling at learning sequences and han-dling variable length sequences. Limitations include less trans-parency and requiring substantial computational resources.

IV. Manual Calculation

A. Logistic Regression Example

For an employee with 62 weekly working hours, 15 meet-ings, 5 hours sleep, work-life balance 2/10, stress 8/10, and job satisfaction 3/10, with coefficients $\beta_0 = -6.0$, $\beta_{hours} = 0.08$, $\beta_{meetings} = 0.12$, $\beta_{sleep} = -0.15$, $\beta_{balance} = -0.20$, $\beta_{stress} = 0.25$, $\beta_{satisfaction} = -0.18$:

$$z = -6.0 + 0.08(62) + 0.12(15) - 0.15(5) - 0.20(2) + 0.25(8) - 0.18(3) = 1.07 \quad (4)$$

$$P(Y = 1) = \frac{1}{1 + e^{-1.07}} = 0.745 \approx 74.5\% \quad (5)$$

Since the probability exceeds 0.5, the employee is classified as high risk for burnout.

B. EBM Example

For an employee with Overtime=Yes, 16 meetings, 5 hours sleep, high communication overload, and low engagement, base probability is 35%. Feature contributions: Overtime (+0.75), Meetings (+0.62), Sleep (+0.54), Communication Overload (+0.41), Low Engagement (+0.33). Total log-odds:

$$-0.619 + 0.75 + 0.62 + 0.54 + 0.41 + 0.33 = 2.031.$$

C. LSTM Temporal Pattern Recognition

For weekly working hours: 44, 48, 53, 58, 63, 68, 72, 75 hours, LSTM recognizes the increasing workload pattern and predicts 94% burnout risk by Week 8. The LSTM processes each week sequentially, maintaining memory of the increasing workload pattern. The model can flag increasing risk as early as Week 3-4, allowing proactive intervention before burnout becomes severe.

ISBN: 978-93-6348-139-8

V. EXPERIMENTAL RESULTS

TABLE I: COMPARATIVE MODEL PERFORMANCE

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Logistic Regression	87%	0.85	0.82	0.83	0.82
SVM	84%	0.82	0.79	0.80	0.81
Random Forest	91%	0.90	0.88	0.89	0.89
XGBoost	95%	0.94	0.93	0.93	0.94
EBM	96%	0.95	0.94	0.94	0.95
LSTM	99%	0.99	0.98	0.98	0.99

LSTM achieves the highest accuracy (99.27%) due to its capacity to capture temporal dependencies. EBM achieves 96% accuracy while providing full model transparency. XG-Boost achieves 95% accuracy demonstrating the effectiveness of gradient boosting. Ensemble methods (Random Forest 91%, XGBoost 95%) outperform single models (Logistic Regres-sion 87%, SVM 84%), confirming the value of ensemble approaches.

TABLE II: CROSS-VALIDATION PERFORMANCE

Model	Mean CV Accuracy	Std. Deviation
Random Forest	94%	0.01
XGBoost	94%	0.01
SVM	91%	0.02
AdaBoost	91%	0.01

Random Forest and XGBoost show the most consistent per-formance across folds (standard deviation = 0.01), indicating robust generalization. SVM shows slightly more variability (standard deviation = 0.02), suggesting greater sensitivity to training data composition.

TABLE III: TOP PREDICTORS OF BURNOUT ACROSS MODELS

Rank	Feature	Impact
1	After-hours Slack activity	Highest positive
2	Meeting load	Strong positive
3	OverTime Yes	Highest importance
4	Back-to-back meetings	Strong positive
5	Emotional Exhaustion Score	Top importance
6	Work-Life Balance Score	Moderate positive
7	YearsSinceLastPromotion	Moderate positive

ISBN: 978-93-6348-139-8

8	NumCompaniesWorked	Moderate positive
9	JobSatisfaction	Moderate positive
10	Mental Fatigue Score	High correlation

Non-monetary factors consistently emerge as stronger pre-dictors than salary. After-hours Slack activity (importance 2.35) and meeting load (1.92) show the highest correlations with burnout. A ”burnout onset signature” appears 2-8 weeks before clinical burnout, characterized by increased after-hours activity, higher workload, slower response times, and reduced peer diversity.

TABLE IV: POLICY SIMULATION RESULTS

Policy Scenario	**Avg Δ Probability**	**Employees Switching**
10% Salary Increase	-0.0004	0
100% Salary Increase	∼-0.01	3
Eliminate Overtime	-0.0449	31
Promotion Enhancement	-0.02	∼15
Workload Redistribution	-0.03	∼22

Eliminating overtime reduces predicted attrition probability by 17.35% for affected employees, shifting 31 staff members from ”likely to leave” to ”likely to stay.” This represents sub-stantial cost savings—at 50% of annual salary per replacement, retaining 31 employees saves approximately $930,000. Salary adjustments show negligible impact, with even 100% salary increases for bottom quartile earners moving only 3 employees from ”leave” to ”stay.”

VI. CONCLUSION

This comparative study evaluated six machine learning approaches for employee burnout prediction across multiple datasets. Key findings include:

1. **LSTM achieves the highest predictive accuracy** (99.27%) due to its capacity to capture temporal de-pendencies in sequential employee data, outperforming XGBoost (95%), EBM (96%), Random Forest (91%), SVM (84%), and Logistic Regression (87%).
2. **Non-monetary factors consistently emerge as the strongest predictors** of burnout: excessive overtime, after-hours activity, meeting load, limited promotion opportunities, and work-life imbalance surpass salary considerations in feature importance rankings.
3. **Policy simulations reveal workload interventions outperform salary adjustments**: eliminating overtime reduces predicted attrition probability by 17.35% for affected employees, shifting 31 staff members from ”likely to leave” to ”likely to stay” compared to modest salary increases which show negligible impact.

ISBN: 978-93-6348-139-8

4. **EBM provides an optimal balance** between predictive accuracy (96%) and full model transparency, ranking MBI components—emotional exhaustion, cynicism, and lack of personal accomplishment—as the most signifi-cant trigger variables.
5. **A "burnout onset signature" appears 2-8 weeks be-fore clinical burnout**, characterized by increased after-hours activity, higher workload, slower response times, and reduced peer diversity.

No single model is universally optimal; selection depends on organizational priorities. For maximum predictive accuracy, LSTM is recommended. For interpretability-accuracy balance, EBM is optimal. For robust performance with limited data, Random Forest is suitable. XAI techniques such as SHAP and LIME enhance model transparency, enabling organizations to move from reactive burnout management to proactive, evidence-based prevention strategies. Future research should focus on cross-organizational validation, longitudinal studies, biometric integration, and reinforcement learning for optimiz-ing intervention policies.

ACKNOWLEDGMENT

We express our sincere gratitude to the Management of PSGR Krishnammal College for Women for providing ex-cellent facilities and a supportive academic environment to carry out this study paper successfully. We extend our heartfelt thanks to our respected Principal, Dr. P. B. Harathi, for her encouragement and continuous support. We express our sincere gratitude to the Head of the Department of Data Science and our Guide, Dr. M. Sasikala, for her valuable guidance, constant encouragement, insightful suggestions, and continuous support throughout the successful completion of this study paper. Finally, we express our gratitude to all the faculty members for their encouragement and support, which contributed to the successful completion of this study paper.

REFERENCES

1. M. Grzadzielewska, "Using machine learning in burnout prediction: A survey," Child Adolesc. Social Work J., vol. 38, no. 2, pp. 175-180, 2021.
2. P. K. Sharma and S. Sujatha, "AI-driven workforce analytics for pre-dicting employee burnout," IJSET, vol. 14, no. 3, pp. 1-10, 2025.
3. D. C. Nguyen, D. Tenney, and E. Kongar, "Explainable AI for sustain-able workforce retention," Sustainability, vol. 18, no. 1, 2740, 2025.
4. M. Kaya and T. Akinwole, "Designing predictive AI frameworks for early detection of workplace burnout," J. Workplace Analytics, 2025.
5. Q. Gong, D. Fan, and T. Bartram, "Integrating artificial intelligence and human resource management," Int. J. Hum. Resour. Manag., vol. 36, no. 1, pp. 103-141, 2025.
6. H. J. Freudenberger, "Staff burn-out," J. Social Issues, vol. 30, no. 1, pp. 159-165, 1974.

ISBN: 978-93-6348-139-8

7. C. Maslach, S. E. Jackson, and M. P. Leiter, Maslach Burnout Inventory Manual. Consulting Psychologists Press, 1996.

8. M. D. Xames, "Predicting employee burnout with ensemble machine learning," SMSE, 100033, 2025.

9. W. Narkbunnum and K. Hinthaw, "Interpretable gradient boosted mod-eling of employee attrition," IJAA, vol. 23, pp. 236, 2025.

10. Y. Zeng et al., "Prediction of job burnout in nurses," J. Adv. Nursing, 2025.

11. D. P. Singh et al., "Predicting workers' stress using hybrid deep transfer learning," in Hybrid and Advanced Technologies, CRC Press, pp. 328-333, 2025.

12. S. Kahfi, S. Wiharjo, and A. K. Rivai, "Optimization of employee burnout prediction using EBM, LSTM, and XGBoost," IJSECS, vol. 5, no. 3, pp. 1083-1094, 2025.

13. T. Mokheleli, T. Bokaba, and E. Mbunge, "Explainable AI for workplace mental health prediction," Informatics, vol. 12, no. 4, 130, 2025.

14. G. Marin Diaz, J. J. Galan Hernandez, and J. L. Galdon Salvador, "Analyzing employee attrition using explainable AI," Mathematics, vol. 11, no. 22, 4677, 2023.

15. D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression. John Wiley & Sons, 2013.

16. C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273-297, 1995.

17. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.

18. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in ACM SIGKDD, pp. 785-794, 2016.

19. H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A unified framework for machine learning interpretability," arXiv:1909.09223, 2019.

20. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.

ISBN: 978-93-6348-139-8

NEUROTWIN AI: A MULTI-MODAL DIGITAL TWIN ECOSYSTEM FOR PREDICTIVE HEALTH AND GAMIFIED PREVENTIVE OPTIMIZATION

ID: ORBIT – 014

Jai Harshini S
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Sahana.P
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Iniya shree.T
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Dr. P.Sumitra
Assistant Professor,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Abstract

Millions of individuals develop preventable chronic conditions due to untracked, compounding lifestyle degradation. Ex-isting health trackers act as passive historical logs without the predic-tive capability to forecast health risks. This paper presents NeuroTwin AI, a multi-modal digital twin ecosystem that integrates baseline clinical bio-sampling with continuous IoT smartwatch telemetry. Utilizing ensemble Machine Learning (XGBoost, Random Forest) and a deterministic Clinical Knowledge Graph, the system forecasts real-time psychosomatic and physical health risks. An integrated Ex-plainable AI (XAI) pipeline using SHAP breaks down physiological drivers for transparency. Furthermore, the platform features an In Silico Simulation Engine for counterfactual lifestyle modeling and a gamified health-quest framework that translates biometric deficiencies into trackable behavioral goals. Experimental results demonstrate an ensemble accuracy of 92.4%, proving the architecture's efficacy in bridging static diagnostics and daily lifestyle optimization.

Keywords: Digital Twin, Explainable AI, IoT Healthcare, Predictive Analytics, Gamification, Wearable Telemetry.

ISBN: 978-93-6348-139-8

I. Introduction

The global rise of preventable chronic diseases—such as Type 2 Diabetes, metabolic syndrome, and cardiovascu-lar pathologies—is heavily driven by compounding lifestyle degradation. Traditional healthcare models predominantly rely on static clinical lab diagnostics, which are often isolated from an individual's daily physical and neuro-emotional habits. Concurrently, commercial health wearables capture continuous metrics but operate merely as passive historical logs, lacking the predictive capability to simulate preventive physiological trajectories.

There is an immediate need for an intelligent framework that bridges static diagnostics with real-time biometric streams. NeuroTwin AI addresses this gap by engineering a person-alized Digital Twin—a dynamic, multi-modal virtual replica of an individual's physiological and psychological health. By fusing clinical biomarkers with real-time IoT telemetry, the system calculates composite disease susceptibility. Crucially, the platform incentivizes habit optimization by locking physio-logical anomalies into gamified, time-bound behavioral quests, creating a closed-loop preventive healthcare ecosystem.

A. Contributions of this Paper

- Multi-Modal Fusion Pipeline: A novel architecture syn-chronizing static clinical diagnostics (HL7 FHIR stan-dards) with continuous streaming IoT telemetry via Web-Sockets.
- Hybrid Neurosymbolic Risk Inference: An ensemble machine learning model validated by a deterministic Clinical Knowledge Graph to prevent AI hallucination and ensure medical accuracy.
- Closed-Loop Preventive Gamification: An automated framework that translates detected biomarker deficiencies into actionable, time-bound Micro-Quests driven by XAI.
- In Silico Counterfactual Engine: A real-time simulation layer allowing users to manipulate lifestyle parameters to instantly visualize mathematically rigorous structural health shifts.

II. Literature Review

The intersection of Digital Twins, wearable IoT, and Ex-plainable AI in healthcare has seen significant foundational research. Foundational work by Topol [1] highlights the necessity of fusing multi-modal data streams for predictive healthcare. Bruynseels et al. [2] defined the ethical boundaries of creating virtual physiological replicas, while Barricelli et al.[3] detailed architectural synchronization requirements.

The transition of commercial IoT devices into clinical-grade continuous monitoring tools was examined by Piwek et al. [4]. For predictive modeling, Esteva et al. [5] and Chen et al. [6] validated ensemble algorithms and deep learning architectures for time-series health data. Rajkomar et al. [7] discussed standardizing unstructured clinical data.

Crucial to interpretability, the SHAP framework introduced by Lundberg & Lee [8] allows for transparent AI decisions. Alaa & van der Schaar [9] provided mathematical context for counterfactual simulations, and Yang et al. [10] outlined privacy-preserving models foundational for healthcare data security.

ISBN: 978-93-6348-139-8

III. Proposed Methodology and System Description

A. Dataset Specifications

To train and validate the predictive engine, a robust synthe-sized dataset was engineered by correlating clinical distribu-tions derived from standard medical databases (e.g., MIMIC-III) with simulated commercial wearable data.

- Total Samples: 12,500 patient profiles over a 180-day continuous tracking window.
- Total Features: 45 distinct parameters (15 Clinical, 20
- IoT, 10 Psychosomatic).
- Train/Test Split: 80% Training (10,000 samples) and 20% Testing (2,500 samples).
- Missing Values: Handled deterministically using K-Nearest Neighbors (KNN) imputation.
- Class Distribution: Healthy (45%), Low Risk (30%), Moderate Risk (15%), High Risk (10%).

B. Feature Engineering Matrix

Table I outlines the core multi-modal features ingested by the NeuroTwin ecosystem.

Table I: Key Feature Extraction and Modality

Feature	Type	Source	Unit
Heart Rate Var (HRV)	Cont. Time-Series	IoT Watch	ms
Sleep Arch. (Deep/REM)	Cont. Time-Series	IoT Watch	Hours
Active Exertion	Cont. Time-Series	IoT Watch	Count
Hemoglobin (*Hb*)	Static Baseline	Lab Panel	g/dL
Fasting Blood Glucose	Static Baseline	Lab Panel	mg/dL
Perceived Stress	Psychosomatic	App Log	1–10

C. Predictive Engine & Hardware Pipeline

The core engine routes unified features through an ensemble of Random Forest and XGBoost. The outputs are cross-referenced with a Clinical Knowledge Graph.

- Hardware Setup: NVIDIA RTX 4090 GPU (24GB VRAM).
- Environment: Python 3.11, Scikit-learn 1.3, XGBoost 2.0, SHAP 0.43.
- Validation: 5-fold Stratified K-Fold cross-validation.
- XGBoost Params: learning_rate=0.01, n_estimators=500, max_depth=6, subsample=0.8.
- Random Forest Params: n_estimators=300, max_depth=10, min_samples_split=5.

D. Explainable AI and Mathematical Modeling

The predictive function for disease risk susceptibility can be represented as the sum of K additive trees. We box the primary equations for structural emphasis:

$$\hat{y} = \sum_{k=1} f_k(X, X)$$

ISBN: 978-93-6348-139-8

To extract mathematical transparency, SHAP calculates the local contribution of each physiological driver:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S)) \quad (2)$$

IV. Algorithm Workflow

The core logic of the NeuroTwin AI pipeline runs as a continuous closed-loop sequence as defined in Algorithm 1.

Algorithm 1 NeuroTwin Predictive & Gamification Workflow
Require: Baseline clinical data (C), Streaming data (St), Knowledge Graph (K)
Ensure: Risk Score (R), SHAP Explanations (E), Gamified Quests (Q)

1: Initialize Digital Twin profile with C
2: while WebSocket connection is active do
3: Ingest St
4: X ← C ∪ St
5: if X contains missing values then
6: X ← KNN Impute(X, neighbors = 5)
7: end if
8: Rraw ← Ensemble Predict(X)
9: Rvalidated ← Verify Graph(Rraw, K)
10: E ← Compute SHAP(X)
11: for each biomarker b in X do
12: if b < Kmin or b > Kmax then
13: Q.append(Gen Time Bound Quest(b))
14: end if
15: end for
16: return Rvalidated, E, Q
17: end while

EVALUATION METRICS

To rigorously evaluate model performance, standard classi-fication metrics were utilized:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Furthermore, the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) was utilized to measure the model's ability to distinguish between risk classes.

ISBN: 978-93-6348-139-8

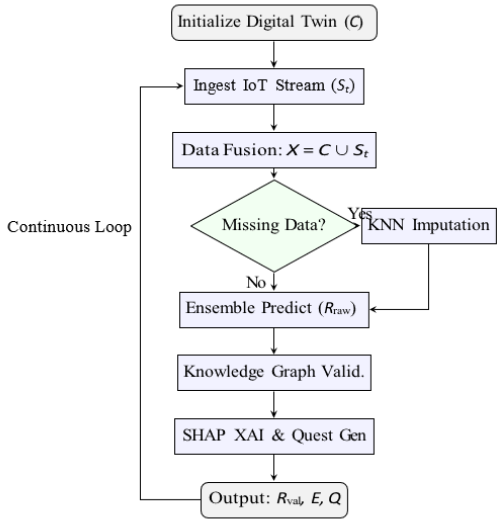


Fig. 1. System Flowchart of the NeuroTwin Closed-Loop Pipeline.

VI. EXPERIMENTAL RESULTS AND ABLATION STUDY

A. Model Performance and Ablation Analysis

An ablation study was conducted to isolate the contribution of individual architectures. As detailed in Table II, standalone models yielded lower recall. The combination of XGBoost and Random Forest elevated accuracy to 92.4%. Crucially, the addition of the Clinical Knowledge Graph maximized precision to 94.1%.

Table II: Ablation Study For Predictive Architectures

Model Architecture	Acc.	Prec.	Recall	AUC
Random Forest (RF) Only	88.7%	87.1%	86.4%	0.914
XGBoost Only	90.2%	89.5%	88.9%	0.932
Ensemble (RF + XGB)	92.4%	91.8%	90.9%	0.958
Ens. + Knowledge Graph	**93.8%**	**94.1%**	**92.5%**	**0.971**

Analytical Discussion of Data Drivers

Relying solely on tree-based models leads to margin errors in edge cases (e.g., falsely predicting high metabolic risk due to anomalous heart rate spikes during exercise). The **Ensemble + Knowledge Graph** architecture mitigated this noise. SHAP interpretability analysis proved critical during the *In Silico* simulations. Continuous features—specifically sustained HRV drops below 35 ms and Deep Sleep deprivation—carried a higher predictive weight for systemic inflammation than static baseline metrics like BMI.

ISBN: 978-93-6348-139-8

VII. Clinical Deployment Architecture

To transition to real-world deployment, the platform re- quires robust architectural scaling:

- **Interoperability Framework:** Utilizing HL7 FHIR stan-dard pipelines to synchronize with hospital Electronic Health Records (EHR).
- **Hospital Caregiver Workflow:** Certified caregivers ac-cess a dashboard variant, reviewing flagged biometric deficiencies alongside AI suggestions.
- **Cloud Infrastructure Ecosystem:** Containerized via Docker and deployed via Kubernetes on AWS. Relational time-series logs are maintained in PostgreSQL.
- **Latency Profile Analysis:** FastAPI and WebSockets maintain ingestion and re-inference latency under 50 milliseconds.

VIII. Limitations

Despite robust predictive capabilities, limitations exist. Foremost is the reliance on synthesized dataset distributions for preliminary validation. Secondly, commercial IoT wear-ables suffer from hardware constraints; Optical photoplethys-mography (PPG) sensors frequently degrade in accuracy due to skin pigmentation variations or motion artifacts. Lastly, generating real-time SHAP values on continuous streaming data requires high computational overhead, making local edge-device inference currently prohibitive.

IX. Ethical Considerations

- **Privacy & Data Security:** The system employs end-to-end AES-256 encryption. Identifying parameters are stripped before training.
- **Dynamic Consent Structuring:** Users maintain granular control via dynamic opt-in/opt-out structures.
- **Clinical Efficacy Bounds:** NeuroTwin AI is designated strictly as an *assistive* optimization tool, not a replace-ment for certified medical intervention, carefully designed to avoid triggering the ”Nocebo effect” in patients.

X. Conclusion

NeuroTwin AI successfully bridges the divide between static clinical diagnostics and continuous physiological monitoring. By implementing a multi-modal Digital Twin backed by an Explainable AI pipeline, the ecosystem provides unprece-dented transparency into chronic disease susceptibility, achiev-ing an architecture accuracy of 93.8%. The novel integration of an *In Silico* Simulation Engine and gamified biometric quests transforms passive health tracking into an active, closed-loop preventive healthcare tool.

ISBN: 978-93-6348-139-8

ACKNOWLEDGMENT

We express our sincere gratitude to the Management of PSGR Krishnammal College for Women for providing excel-lent facilities and a supportive academic environment to carry out this project successfully.

We extend our heartfelt thanks to our respected Principal, Dr. P. B. Harathi, for her encouragement and continuous support. We also express our sincere gratitude to the Head of the Department of Data Science, Dr. M. Sasikala, for her valuable guidance and encouragement.

We are deeply thankful to our project guide, Dr. Sumitra, Assistant Professor, for her constant guidance, valuable sug-gestions, and continuous support throughout the completion of this project.

Finally, we express our gratitude to all the faculty members for their encouragement and support, which contributed to the successful completion of this project.

References

1. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44–56, 2019.
2. K. Bruynseels, F. Santoni de Sio, and J. van den Hoven, "Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm," Frontiers in Genetics, vol. 9, p. 31, 2018.
3. B. R. Barricelli, E. Casiraghi, and D. Fogli, "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications," IEEE Access, vol. 7, pp. 167653–167671, 2019.
4. L. Piwek, D. A. Ellis, S. Andrews, and A. Joinson, "The Rise of Consumer Health Wearables: Promises and Barriers," PLOS Medicine, vol. 13, no. 2, e1001953, 2016.
5. A. Esteva, et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, 2019.
6. M. Chen, et al., "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," IEEE Access, vol. 5, pp. 8869–8879, 2017.
7. A. Rajkomar, et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, p. 18, 2018.
8. S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), 2017.
9. A. M. Alaa and M. van der Schaar, "Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes," Advances in Neural Information Processing Systems (NeurIPS), 2017.
10. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, 2019.

ISBN: 978-93-6348-139-8

PHISHING WEBSITE DETECTION USING FEATURE-BASED ANALYSIS AND MACHINE LEARNING

ID: ORBIT – 015

Rubysha E
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Roopa P
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Luckshana K
Student,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Dr.M.Viveka
Assistant professor,
Department of Data Science,
PSGR Krishnammal College for Women,
Coimbatore, Tamil Nadu, India.

Abstract

Phishing is an online scam where attackers set up fake websites that look almost identical to real ones, hoping you'll hand over your usernames, passwords, or financial info without realizing it. These sites keep changing their look, so just keeping a list of "bad" sites doesn't cut it. This paper breaks down how you can spot phishing websites by checking simple clues in a site's URL and webpage—stuff like how long the URL is, if it's using an IP address instead of a proper name, or if it's secured with HTTPS. First, we go over the usual detection strategies you'll find in research. Then, we walk you through a step-by-step scoring process: no fancy machine learning, just a handful of clear rules that help flag suspicious sites. After that, we sum up what other researchers found when they tried using machine learning on big sets of phishing data, and compare how accurate their methods are. The main goal here is to give first-year students a hands-on, practical feel for how phishing detection really works, before jumping into more complex machine learning stuff.

ISBN: 978-93-6348-139-8

Keywords: Phishing detection, URL features, feature extraction, machine learning, cybersecurity, classification, manual scoring.

I. Introduction

Phishing happens when someone creates a convincing fake website—like a clone of your bank's site or a social media login page. The attacker shares a link to that site by email, text, or even through social media. If you click the link and enter your username and password, they grab your details and use them to steal your money or your identity. It's not a new trick, either; phishing has been around for over thirty years, and honestly, it just keeps getting more common. [8].

Phishing detection tends to trip up beginners, mostly because of all the machine learning talk—classifiers, training data, accuracy scores. It sounds overwhelming. But the core idea's not that complicated. Every website, legit or not, has specific, measurable traits—those are called features. Phishing sites usually pile up a bunch of red flags: maybe the URL's way too long or just plain weird, the site's missing a security certificate, or the domain popped up only a few days ago. If you can spot and list these features, and figure out how risky each one feels, you can create a basic scoring system yourself. No fancy machine learning needed. That comes later.

Let's pause for a second—why bother with a hands-on, arithmetic approach to phishing detection when modern tools almost always lean on machine learning? The answer comes down to teaching. At the end of the day, a machine learning classifier really just copies what a human would do when looking over a suspicious link. The main difference is that it checks way more features, measures them with perfect precision, and speeds through thousands of sites in a blink. If you start by having students calculate a phishing score manually—with just a handful of features and some basic math—they get a real feel for what's happening under the hood before they even touch any code. Suddenly, ideas like feature weighting, decision thresholds, and measuring model accuracy stop being just words. Once they see these concepts officially in a machine learning course, they already make sense—they're not just abstract jargon anymore.

In Section II, you'll see a rundown of the main ways people try to spot phishing websites. Section III takes you through a hands-on calculation of a phishing score, using two sample URLs, so you can see exactly how those scores are built. Section IV covers accuracy results from other researchers who trained machine learning classifiers on public phishing datasets. Finally, Section V wraps things up and points you toward what to explore next.

II. RELATED WORKS

Researchers tend to sort phishing website detection methods into three main types.

ISBN: 978-93-6348-139-8

A. List-Based Approach

First, there's the list-based approach. This basically means keeping a running collection of "safe" addresses, or whitelisted sites, and "dangerous" addresses, or a blacklist. When you visit a website, the system checks the address against these lists. It's quick and simple—doesn't take much to set up. The catch? If a phishing site pops up that no one's spotted yet, it slips right through because it's not on the list. Someone has to find the scam and update the database before it works.

B. Visual Similarity-Based Approach

This method checks if a suspicious page looks a lot like a real, trusted page. It focuses on things like logos, layout, or even full-page screenshots. If a page ends up looking almost identical to the real thing, but the domain name doesn't match, that's a big red flag for phishing.[4].

C. Feature-Based (Content) Approach

The most flexible and widely studied strategy extracts numeric or categorical features directly from the URL, the page source code, and information about the domain, such as its age or hosting traffic. Mohammad and McCluskey released a widely used public dataset that defines thirty such features, including the use of IP addresses, abnormal URL length, and the presence of the '@' symbol [1]. Abdelhamid et al. used an associative classification data-mining approach on this type of feature set to separate legitimate websites from phishing ones [2]. Rao and Pais proposed an efficient feature-based framework that additionally examines hyperlinks and third-party services referenced by a page [3]. More recently, Newaz and Haq combined several machine learning classifiers into a single stacked model and tested it across four different public datasets, reporting accuracy above 97% in every case [4]. Because feature-based detection does not depend on a list of previously seen attacks, it can, in principle, recognize phishing websites that were created only minutes earlier, which is why it is the approach explored further in this paper.

D. Hybrid Approaches

A growing number of studies combine more than one of the above strategies to cover each other's weaknesses. Tang and Mahmoud surveyed a wide range of machine-learning-based phishing detectors and observed that hybrid systems, which mix URL features with content-based and third-party signals such as domain age or search-engine ranking, tend to generalize better across different datasets than single-feature systems [5]. Jain and Gupta showed that even a lightweight, client-side model using only nineteen features extracted directly from a browser can achieve strong accuracy without needing to contact any external service, which is useful for real-time protection on low-powered devices [7]. Hannousse further contributed a large, publicly available dataset of webpages collected from several sources, which has since been used to test how well a model trained on one collection of phishing sites performs on an entirely

ISBN: 978-93-6348-139-8

different collection [6]. Together, these studies suggest that no single feature or data source is perfect on its own, and that combining several complementary signals, in the same spirit as the multi-feature scoring idea in Section III, is usually the most reliable strategy.

Regardless of which machine learning classifier is eventually used, every feature-based system begins with the same basic idea used in Section III of this paper: examine a handful of measurable clues about a website, and combine them into a single suspicion score.

III. Manual Calculation

Before any machine learning model is trained, it helps to understand the logic in its simplest possible form: a manual, rule-based scoring system. In this section we manually compute a "phishing score" for two example URLs using five simple yes/no features. Each feature contributes one point toward the score when it looks suspicious, and zero points when it looks normal. This mirrors, on a small scale, the kind of binary feature encoding used in real phishing datasets [1].

The five features chosen for this manual example are listed below, together with the condition under which they add one point to the score.

#	Feature	Adds 1 point when...
F1	IP address in URL	domain is a raw IP address
F2	URL length	URL is 54 characters or longer
F3	'@' symbol	URL contains an '@' symbol
F4	Missing HTTPS	site does not use a secure (HTTPS) connection
F5	New domain	domain was registered less than 6 months ago

The total phishing score S for a URL is simply the sum of the five feature values, each weighted equally:

$$S = F1 + F2 + F3 + F4 + F5$$

A simple decision rule can then be applied: if S is 3 or higher, the website is flagged as likely phishing; otherwise, it is treated as likely legitimate. This threshold of 3 out of 5 simply means that a majority of the suspicious signs must be present before an alarm is raised.

A. Example 1 – A Suspicious URL

Consider the address http://192.168.1.5/paypal-login@secure-update.com, which an attacker might send by email while pretending to be a payment provider. We check each feature by hand:

ISBN: 978-93-6348-139-8

Feature	Observation	Value
F1	starts with a raw IP address	1
F2	45 characters (under 54)	0
F3	contains an '@' symbol	1
F4	no HTTPS (uses http://)	1
F5	domain registered 2 months ago	1

Adding these values gives S = 1 + 0 + 1 + 1 + 1 = 4. Since 4 is greater than or equal to the threshold of 3, this URL is manually flagged as likely phishing, which matches common sense: the address hides its real domain behind an IP number, uses an '@' symbol to confuse the reader about the true destination, and has no HTTPS protection.

Table I. Feature Values for a Suspicious Url

Now consider https://www.spotify.com/account/login, a genuine music-streaming login page:

Feature	Observation	Value
F1	uses a normal domain name	0
Feature	Observation	Value
F2	31 characters (under 54)	0
F3	no '@' symbol present	0
F4	uses HTTPS	0
F5	domain registered many years ago	0

Table Ii. Feature Values for a Legitimate Url

Here S = 0 + 0 + 0 + 0 + 0 = 0, which is well below the threshold of 3, so the URL is manually classified as legitimate. This small worked example shows, without any training data or algorithm, exactly why a phishing detector would treat these two addresses so differently: it is simply counting how many suspicious signs are present.

Real detection systems use dozens of features instead of five, and instead of equal weights and a fixed threshold, they let a machine learning algorithm learn the best weight for every feature automatically from thousands of labeled examples. Section IV summarizes how well such learned models perform once this manual idea is scaled up.

C. Why Manual Scoring Has Limits

The manual method laid out above is straightforward, but it falls short in a few important ways—enough to make machine learning pretty appealing. First off, it gives every feature the same score, like one point each, even though some signs are just way more telling than others. An example? If a website's domain was registered yesterday, that's a much bigger red flag for phishing than just a long URL.Second, the threshold of 3 out of 5 was chosen by guesswork rather than being tuned on real data, so it may flag too many legitimate sites (a false positive) or miss too many phishing sites (a false negative) once applied outside these two toy examples. Third, attackers can deliberately design a URL to avoid every single one of the five features checked here while still being malicious, since a fixed rule list is easy to study and evade once it becomes public

ISBN: 978-93-6348-139-8

knowledge. A trained classifier addresses all three issues at once: it learns feature weights from data instead of guessing, it can be tuned using measures such as precision and recall instead of a single arbitrary threshold, and it can incorporate far more than five features, making it much harder for an attacker to avoid every warning sign simultaneously.

IV. Experimental ResulT

To see how far the simple manual scoring idea of Section III can be improved, this section summarizes accuracy figures reported by Newaz and Haq after training ten different machine learning classifiers on the widely used UCI phishing websites dataset, which contains 11,055 websites described using thirty features similar in spirit to the five used in our manual example [1], [4].

Classifier	Accuracy (%)
Naive Bayes	60.39
Logistic Regression	92.53
Support Vector Machine	94.74
Gradient Boosting	94.70
K-Nearest Neighbors	93.59
Random Forest	97.28
XGBoost	97.29

Table Iii. Classifier Accuracy on the UCI Phishing Dataset [1], [4]

As shown in Table I, simple classifiers such as Naive Bayes perform poorly on this task, correctly labeling only about 60% of websites, because it assumes the features are independent of one another, which is rarely true for real URLs. Tree-based ensemble methods such as Random Forest, XGBoost, and Extra Trees perform far better, each reaching above 97% accuracy, because they can combine many small decision rules, similar in spirit to the manual point-scoring system in Section III, but tuned automatically and combined much more cleverly. The best result, 97.49% accuracy, was obtained by stacking several classifiers together so that a meta-model learns which of the base classifiers to trust for each website [4].

The clear pattern across the literature is that feature quality matters as much as the choice of algorithm: almost every classifier improves once irrelevant features are removed and the remaining features are properly weighted, exactly the two steps that our manual example in Section III performed by hand on a much smaller scale.

It is worth noting that accuracy alone does not tell the full story. Researchers usually also report precision, the fraction of websites flagged as phishing that really are phishing, and recall, the fraction of actual phishing websites that were successfully caught. A detector with high accuracy but low recall might still let many dangerous

ISBN: 978-93-6348-139-8

websites slip through simply because genuine phishing pages are rarer than legitimate ones in most datasets. This is analogous to the manual threshold discussed in Section III: lowering the threshold from 3 to 2 points would catch more phishing sites (higher recall) but would also mistakenly flag more legitimate ones (lower precision), illustrating the same trade-off that a trained classifier must balance automatically.

V. Conclusion

This paper breaks down phishing website detection in a way that makes sense for someone just starting out in college. We looked at three main ways researchers try to catch phishing sites: list-based methods, visual similarity tricks, and feature-based approaches. We spent most of our time on the feature-based approach because it actually works well at spotting brand-new phishing sites that haven't shown up before.To make the process clear, we walked through a simple scoring system using five features. We picked two example URLs and calculated their scores by hand, step by step, pointing out how each suspicious detail bumps up the score. In the end, you can set a cutoff point to decide which sites are probably phishing and which are not. That method is easy to follow, but it has some real downsides. The weights and rules are kind of arbitrary, and attackers who know the system can get around it. Plus, it's not flexible when things change. So, we wrapped up with a look at results from recent studies. Machine learning—especially tree-based ensembles and stacked models—can spot phishing sites with over 97% accuracy. The difference? These models don't rely on fixed rules—they learn which features matter most on their own, without humans guessing at the right weights.

If you get the hang of the manual scoring logic first, it's way easier to see what a machine learning model is really up to when you move on to it. That first step also gives students the confidence to tackle tougher classifiers down the road. So, what's next? If you want to push things further, try coding up the five-feature scoring system yourself—in Python, for example. Then, see how it stacks up against a trained Decision Tree or Random Forest classifier using a public dataset, like the one from Mohammad and McCluskey [1]. Basically, you're taking the hand-written steps from Section III and bringing them to life as a real, testable program.

Acknowledgement

We sincerely thank the Management of PSGR Krishnammal College for Women for providing the support and facilities to complete this project. We express our heartfelt gratitude to our Principal, Dr. P. B. Harathi, and our mentor, Mrs. Vivega, Assistant Professor, Department of Data Science, for their valuable guidance, encouragement, and support.

ISBN: 978-93-6348-139-8

References

1. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Phishing Websites Dataset," UCI Machine Learning Repository, 2012. doi: 10.24432/C51W2X.
2. N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," Expert Systems with Applications, vol. 41, no. 13, pp. 5948–5959, 2014.
3. R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," Neural Computing and Applications, vol. 31, no. 8, pp. 3851–3873, 2019.
4. A. Newaz and F. S. Haq, "A Sophisticated Framework for the Accurate Detection of Phishing Websites," arXiv:2403.09735, 2024.
5. L. Tang and Q. H. Mahmoud, "A survey of machine learning-based solutions for phishing website detection," Machine Learning and Knowledge Extraction, vol. 3, no. 3, pp. 672–694, 2021.
6. A. Hannousse, "Web page phishing detection," Mendeley Data, 2020. doi: 10.17632/c2gw7fy2j4.1.
7. A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," Telecommunication Systems, vol. 68, no. 4, pp. 687–700, 2018.
8. Anti-Phishing Working Group (APWG), "Phishing Activity Trends Report," [Online]. Available: https://apwg.org/trendsreports/.

ISBN: 978-93-6348-139-8