Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

Role of Artificial Intelligence in Enhancing Security and Fraud Detection in Digital Online Payment Systems: A Perspective

Ch. Keerthi[1], Dr. B. Nandini[2]

[1]MCA, TS-SET, UGC-NET, Lecturer in Computer Science, Girraj Government Degree College (Autonomous), Nizamabad, Telangana, India
[2]Associate Professor, Department of CSE, Telangana University, Dichpally, Nizamabad, Telangana, India

Abstract:
Digital payment systems have become the backbone of global commerce, but their rapid expansion has been paralleled by a sharp rise in payment fraud, identity theft, and cyber-enabled financial crime. This paper examines the role of Artificial Intelligence (AI) in enhancing the security and fraud-detection capability of digital online payment systems, drawing on recent industry reports, regulatory data, and empirical machine learning studies. The study adopts a descriptive-analytical approach, synthesising secondary data from central bank publications, market-research reports, and peer-reviewed comparative studies of algorithms such as Random Forest, Artificial Neural Networks, Support Vector Machines, and Gradient Boosting Results show that AI-powered fraud detection tools, such as the ensemble method Random Forest, consistently outperform their human counterparts, with accuracy between 92 and 100 percent in experimental and production environments; real-time behavioural analytics, biometric authentication and natural language processing also take fraud protection beyond transaction-level screening to include phishing, social engineering and mule-account detection. The scale of the challenge, as well as the regulatory response to the menace of digital payment fraud, is evident from the Indian Unified Payments Interface (UPI) ecosystem, where the value of digital payment fraud fluctuated despite an over 40 per cent increase in the number of transactions year-on-year, and the Reserve Bank of India's (RBI) MuleHunter.AI initiative. However, the paper still points to certain issues that have not been overcome, such as class imbalance, manipulation by a growing swarm of AI-savvy fraudsters, a lack of explanation, data-privacy restrictions, and disparity in adoption by institutions of varying sizes. The paper concludes that AI plays an essential role in the current payment-security architecture, but it must be complemented with a multi-layered approach that includes technological solutions, regulations, and consumer-awareness initiatives to effectively withstand the ever-changing threat landscape.

Keywords: Artificial Intelligence, fraud detection, digital payments, machine learning, cybersecurity, and UPI.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

1. Introduction

Change towards a cashless economy has changed the way people, businesses and governments exchange value globally. From mobile wallets to card networks, real-time interbank transfer rails to unified payment interfaces, digital online payment systems now handle billions of transactions annually – and this number is expected to keep climbing (MarketsandMarkets, 2025). In emerging economies like India, the Unified Payments Interface (UPI) has seen significant growth in transactions, reaching 185.9 billion transactions in the 2024-25 fiscal year compared to 137.9 billion transactions in 2023-24, a rise of 41.7 percent (Medianama, 2025). The advantages of digital payments—convenience, speed, and financial-inclusion benefits—have been well established, to be sure, but there's been a parallel and equally fast rise in payment-related fraud, cybercrime, and financial deception, from card-not-present fraud and phishing to the latest AI-driven social-engineering campaigns.

Although traditional rule-based fraud-detection systems rely on static thresholds and manually set red flags, they have been unable to match the speed and complexity of today's digital transactions (Preciado Martínez & Náñez Alonso, 2025). The use of automated bots, synthetic identities, dark-web marketplaces, and more recently, AI-generated deepfakes to evade traditional controls is growing in popularity among fraudsters (CoinLaw, 2026). In response, financial institutions, payment processors, and regulators have resorted to Artificial Intelligence (AI) and machine learning (ML) as their main technological defense. Compared to static rule-based systems, AI-based systems can learn from previous transaction patterns, detect new fraud patterns, and catch anomalies on the fly, potentially before a transaction is approved (U.S. Bank, 2026).

The need for this transformation is evidenced by recent survey data: In 2024, 79 percent of companies reported attempted and/or actual payment fraud, compared to 65 percent two years prior (U.S. Bank, 2026); and 76 percent of businesses indicated they were enhancing investment in AI, machine learning, and analytics specifically for fraud prevention (SEON, 2025). Meanwhile, adoption is not uniform: Even though business-email-compromise and AI-powered social-engineering attacks are surging (Association for Financial Professionals, 2026), many organisations are still using legacy controls due to cost factors or the belief that AI technology is not ready. This paper examines the role of AI in enhancing the security and fraud-detection capability of digital online payment systems. Specifically, it aims to: (a) review the principal AI and machine-learning techniques currently applied to payment security, including supervised classification, unsupervised anomaly detection, deep learning, natural language processing, and behavioural biometrics; (b) synthesise comparative evidence on the detection accuracy of these techniques drawn from recent empirical studies; (c) examine the scale and trajectory of digital payment fraud and AI-based countermeasures in the Indian context, using UPI as an illustrative case; and (d) critically assess the limitations, risks, and future directions associated with AI-driven fraud detection. The paper adopts a descriptive-analytical methodology based on secondary data drawn from central-bank publications, market-research reports, and peer-reviewed comparative studies, and is organised into eight sections: introduction, literature review, AI techniques for fraud detection, the Indian digital-payments landscape, market growth and industry adoption, challenges and limitations, recommendations, and conclusion.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

2. Literature Review

A substantial and growing body of literature has examined the comparative performance of machine-learning algorithms in detecting fraudulent financial transactions. Afriyie et al. (2023) tested the performance of logistic regression, decision trees and random forest classifiers on a credit-card transaction dataset and found that random forest out-performed the others with a predictive accuracy of 96 percent and an area under the curve (AUC) of 98.9 percent; they also observed that fraudulent transactions were concentrated among customers aged above 60 years and during late-night periods, highlighting the importance of temporal and behavioural features in model design. Both Preciado Martínez and Náñez Alonso (2025) analyzed bank transactions from over 565,000 real-world transactions, using random forest, artificial neural network, and naïve Bayes models, with Preciado Martínez reporting that all legitimate transactions were correctly classified, while Náñez Alonso reported that random forest classified 95.79 percent. Both of these models reported that the random forest model was the most robust model, and they argue that the results validate the viability of the random forest model for real-time banking deployment when the data set is extremely class imbalanced, like those common in a fraud data set.

Other work has been done to compare the algorithms with other algorithms, such as k-nearest neighbours, support vector machines, naïve Bayes, and gradient boosting. In most of the studies, these are found to be more powerful than simpler linear models, and deep-learning-based models like artificial neural networks are especially effective in capturing nonlinear, high-dimensional relationships in large-scale transaction data, but can be more computationally expensive and less interpretable. Industry-level analyses support these academic results: Industry-wide analysis suggests that AI-based fraud-detection tools can be as accurate as 92 percent when used under production conditions (ZIGRAM, 2026), which is generally in line with the 92-100 percent range of the academic studies reviewed above.

In addition to the classification at the transaction level, papers have started to focus on multi-layered AI architectures combining behavioural biometrics, network/graph analysis and natural-language processing. The logic behind the Reserve Bank of India's MuleHunter is the same as that of a graph-based approach, which looks at account-to-account, device-to-device, and beneficiary-to-beneficiary relationships to identify coordinated fraud rings and mule-account networks that would bypass transaction-by-transaction screening. The AI/ML-based system was introduced in December 2024 by the AI tool in two public-sector banks to detect accounts where money from illegal transactions is laundered (Business Standard, 2025). Behavioural-biometric, on the other hand, continuously validates the person using typing frequency, touchscreen pressure, and how the device is handled, adding an authentication layer that is continuous and not so obvious to the end user. A parallel strand of literature documents the scale of the threat that AI-based defences must address. Industry surveys report that payment fraud affected a majority of organisations in 2024 and 2025, with business-email compromise rising sharply year-on-year (Association for Financial Professionals, 2026), while market-intelligence sources describe a surge exceeding 1,000 percent in AI-enabled fraud, including deepfake and synthetic-identity fraud, within a single year, far outpacing the growth of traditional fraud techniques (BNY, 2026). This asymmetry, in which fraudsters themselves increasingly deploy generative AI to craft convincing phishing content and synthetic identities, has produced a continual escalation between attackers and defenders, in which the sophistication of

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

defensive machine-learning systems must keep pace with adversarial innovation. Collectively, this literature establishes both the demonstrated technical effectiveness of AI-based fraud-detection models and the escalating threat environment that necessitates their continued refinement, the two themes that structure the remainder of this paper.

3. AI Techniques for Fraud Detection and Security in Digital Payments

AI-based fraud-detection architecture in modern payment systems is rarely built on a single algorithm. Instead, financial institutions typically deploy a layered combination of techniques, each addressing a different aspect of the fraud problem, from real-time transaction scoring to post-transaction network analysis. Table 1 summarises the principal categories of techniques currently applied to digital-payment security, before the following subsections examine each in turn.

Table 1- AI Techniques Applied in Digital Payment Fraud Detection and Security

Technique	Core Function	Representative Application
Supervised Machine Learning (Random Forest, Logistic Regression, Gradient Boosting)	Classifies transactions as fraudulent or legitimate using labelled historical data	Real-time transaction scoring and authorisation decisions
Unsupervised Learning / Anomaly Detection	Flags deviations from normal behaviour without labelled fraud examples	Detecting novel fraud typologies absent from training data
Deep Learning / Artificial Neural Networks	Models complex, nonlinear relationships across large feature sets	Sequential spending-pattern analysis, high-dimensional risk scoring
Natural Language Processing (NLP)	Analyses message and website content for deception indicators	Phishing/smishing detection, scam-message alerts
Behavioural Biometrics	Continuously authenticates users via interaction patterns	Detecting account takeover after login
Graph / Network Analytics	Maps relationships among accounts, devices, and beneficiaries	Mule-account and fraud-ring detection (e.g., RBI's MuleHunter.AI)

Note. Compiled by the authors based on Afriyie et al. (2023), Preciado Martínez and Náñez Alonso (2025), Business Standard (2025), and CoinLaw (2026).

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

3.1 Supervised Machine Learning

Historical transaction datasets can be labelled as either fraudulent or legitimate, and supervised learning algorithms can learn the decision boundaries that they can then apply to new transactions that are not in their training set. Popular algorithms used include logistic regression, decision trees, random forest, gradient boosting, and support vector machines. Of these, ensemble methods, especially random forest, have consistently proven to be the most effective in various independent studies, with the accuracy rate ranging from 92 percent to 100 percent (Afriyie et al., 2023; Preciado Martínez & Náñez Alonso, 2025; ZIGRAM, 2026). Table 2 summarises comparative accuracy findings drawn from recent empirical studies alongside an industry-reported benchmark.

Table 2- Comparative Accuracy of Machine Learning Algorithms in Fraud-Detection Studies

Study / Source	Algorithm(s) Evaluated	Best-Reported Performance
Afriyie et al. (2023)	Logistic Regression, Decision Tree, Random Forest	Random Forest: 96% accuracy; AUC 98.9%
Preciado Martínez & Náñez Alonso (2025)	Random Forest, Neural Network, Naïve Bayes	Random Forest: 100% (legitimate) / 95.79% (fraud) accuracy
ZIGRAM (2026) (industry-aggregated data)	AI-powered detection tools (general)	Up to 92% accuracy identifying suspicious activity

Note. Compiled by the authors from the sources cited in the first column.

3.2 Unsupervised Learning and Anomaly Detection

Because fraudulent transactions are rare relative to legitimate ones and fraud typologies evolve continuously, purely supervised approaches risk failing to detect novel fraud patterns absent from historical training data. Supervised methods overcome this by learning the statistical distribution of normal transaction patterns and identifying anomalies, but they need labeled examples of fraud. Unsupervised methods, such as clustering algorithms and autoencoder-based anomaly detection, do not have this requirement. For production systems, hybrid models such as supervised classification and unsupervised anomaly scoring are gaining popularity, with 62 percent of organisations surveyed saying that they implemented real-time transaction monitoring, where the models would be sensitive to known fraud patterns and still learn from emerging fraud patterns.

3.3 Deep Learning and Neural Networks

Deep-learning architectures, including multilayer artificial neural networks and, in more recent implementations, convolutional and recurrent networks, are capable of modelling complex, nonlinear interactions among large numbers of transaction and behavioural features. While computationally more expensive and less interpretable than tree-based ensembles, neural-network approaches have shown

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

competitive or superior performance on complex, high-dimensional datasets and are particularly well suited to sequential data, such as a customer's historical spending trajectory, where the order and timing of transactions carry predictive information beyond any single transaction's static attributes.

3.4 Natural Language Processing for Phishing and Social-Engineering Detection

A growing share of payment fraud originates not at the point of transaction but earlier, through phishing emails, fraudulent SMS messages, and deceptive websites designed to harvest credentials or induce victims to authorise fraudulent transfers voluntarily. Natural language processing (NLP) models scan message content, sender metadata, and embedded URLs for linguistic and structural indicators of deception, enabling banking applications to warn users before a compromise occurs. As over 70 percent of phishing emails are now believed to incorporate AI elements, the ability has become even more significant, as fraudsters increasingly use generative AI to create more convincing and personalised scam emails (CoinLaw, 2026).

3.5 Behavioural Biometrics and Continuous Authentication

Behavioural-biometric systems use a user profile, which includes information about how users type or touch the screen, how they swipe or move their fingers, how they navigate the device and more, and monitor and compare live sessions to the profile. The step-up authentication and/or transaction holds can be triggered even with a correct password and/or one-time password, helping in preventing account takeover fraud that bypasses password- and one-time-password-based authentication systems, which is a growing concern as half of the Indian UPI fraud victims had reported experiencing an unauthorised transaction after their credentials were compromised or their app settings were tampered with (Business Standard, 2025).

3.6 Graph and Network Analytics

Fraud has become a coordinated effort using mule accounts, shell identities, and interconnected devices, instead of being single transactions. Graph-based AI models refer to accounts, devices, and beneficiaries as nodes, and transactions as edges, using network-analysis techniques to identify clusters and pathways that are indicative of money laundering or mule-account activity. The MuleHunter by the Reserve Bank of India. AI, which was piloted in two public-sector banks in December 2024, is an example of this at the national level, alongside exclusive internet domains like .bank.in and .fin.in that are being introduced to curb phishing and instill trust in digital banking (Business Standard, 2025).

4. Indian Digital Payments Landscape: Trends and Fraud Patterns

The case of India is an edifying example of the challenges of fast digital payments growth and growing fraud risks, as the country's Unified Payments Interface (UPI) ecosystem is so large and the Reserve Bank of India (RBI) has such fine-grained reports. The volume of transactions on UPI increased by 41.7 percent to 185.9 billion in 2024-25 as compared to 2023-24, while the value of transactions rose by 30.3 percent to ₹260.6 lakh crore; by the end of FY2024-25, UPI transactions accounted for 84 percent of all

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

transactions on India's retail payment systems (Medianama, 2025). The widespread adoption is attributed to the affordability of the internet, increasing financial inclusion, and interoperability among banks and non-bank payment apps like Google Pay, PhonePe, and Paytm (BusinessToday, 2024).

This growth has been accompanied by an increase in fraud, but not in direct proportion to transaction value. Table 3 shows the fraud value that has been reported for each UPI throughout the three fiscal years, as presented by the RBI and National Payments Corporation of India (NPCI) disclosures, from which PwC India (2025) has compiled the values.

Table 3- UPI Fraud Value as a Share of Transaction Value in India, FY2022-23 to FY2024-25

Fiscal Year	UPI Fraud Value (₹ crore)	Fraud as % of Transaction Value
2022-23	573	0.4%
2023-24	1,087	0.5%
2024-25 (partial; Apr-Sep, with full-year trend projection)	485 (~970 projected full year)	0.4%

Note. Adapted from PwC India (2025), based on RBI and NPCI data.

While the percentage of transaction value lost to fraud has remained comparatively small, between 0.4 and 0.5 percent, the absolute rupee value of losses has grown substantially in line with transaction volumes, and survey evidence suggests the visible, reported figures understate the true scale of the problem: a 2025 LocalCircles survey of more than 32,000 UPI users across 365 districts found that one in five families with a UPI user had experienced fraud at least once in the preceding three years, and that 51 percent of victims did not report the incident to any authority (Business Standard, 2025).

Regulatory reporting through the RBI's broader fraud-monitoring framework, which captures fraud across all banking channels rather than UPI alone, shows a more complex pattern. Table 4 summarises digital-payment (card/internet) fraud cases and values reported for the 2023-24 and 2024-25 fiscal years, alongside corresponding UPI transaction volumes for context.

Table 4- Digital Payment (Card/Internet) Fraud Reported under RBI's Banking Fraud Framework

Metric	FY2023-24	FY2024-25
Number of digital payment fraud cases	29,082	13,516
Value of digital payment fraud (₹ crore)	1,457	520
Digital payments' share of total reported banking fraud cases	—	56.5%

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

Metric	FY2023-24	FY2024-25
Corresponding UPI transaction volume (billion)	131.1	185.9

Note. Compiled by the authors from Reserve Bank of India Annual Report data, as reported in BusinessToday (2024) and Medianama (2025).

The number of digital-payment fraud cases fell from 29,082 in FY2023-24 to 13,516 in FY2024-25, even as the value defrauded fell correspondingly from ₹1,457 crore to ₹520 crore; however, digital payments' share of all reported banking-fraud cases rose to 56.5 percent in FY2024-25, reflecting both the declining prevalence of other fraud categories, such as physical cheque fraud, and the continued concentration of fraud within digital and card-based channels (Medianama, 2025). It should be noted that reported fraud totals across RBI disclosures have also been affected by a Supreme Court judgment mandating the reclassification and fresh reporting of previously unclassified cases, which caused the overall defrauded amount across all banking channels to rise sharply, from ₹12,230 crore to ₹36,014 crore, independent of any underlying change in fraud incidence (Medianama, 2025).

Meanwhile, Indian regulators and payment-system operators have had to come up with a multi-layered defence mechanism using AI. NPCI has introduced a centralised solution for fraud monitoring to banks, which leverages AI/ML-based models to alert and decline suspicious transactions, as well as device-binding and two-factor authentication, and sets daily transaction limits (Press Information Bureau, 2025). The RBI's proposed Digital Payments Intelligence Platform, headed by former managing director (MD) of NPCI, A. P. Hota, will be used for advanced analytics to detect and handle frauds in real-time across banks, and MuleHunter.Specifically, AI is being used to identify and block mule-account networks that are used to facilitate the laundering of illegally acquired funds (Business Standard, 2025). As of March 2025, the Central Payments Fraud Information Registry has also been expanded to include 49 scheduled urban cooperative banks and hundreds of regional rural and district cooperative banks, and is now a more comprehensive effort to move away from institution-level, siloed fraud controls to system-wide, AI-driven fraud intelligence sharing (Medianama, 2025).

5. Market Growth and Industry Adoption of AI-Based Fraud Detection

The scale of investment in AI-based fraud-detection technology reflects both the severity of the threat and growing confidence in AI's effectiveness as a countermeasure. Multiple independent market-research firms project substantial growth in the global fraud detection and prevention (FDP) market over the remainder of the decade, though estimates vary depending on scope and methodology. Table 5 summarises three recent projections.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

Table 5-Global Fraud Detection and Prevention (FDP) Market Size Projections

Source	Base-Year Estimate	Projected Value	Target Year / CAGR
Markets and Markets (2025)	$32.00 billion (2025)	$65.68 billion	2030 / 15.5%
Grand View Research (2026)	$35.3 billion (2025)	$129.4 billion	2033 / 18.1%
CoinLaw (2026)	$73.62 billion (current)	Growing steadily	21.2% annual growth

Note. Compiled by the authors from the sources cited in the first column.

Despite variation in absolute figures, all major forecasts agree on double-digit compound annual growth, driven by continued migration of enterprise workloads to cloud environments, the expansion of real-time payment rails, and the increasing sophistication of AI-powered anomaly detection (MarketsandMarkets, 2025). Behavioural-analytics platforms, a subset of the broader market, are projected to grow from approximately $2.06 billion to $7.63 billion by 2034, while fraud-analytics software specifically is projected to nearly double from $4.7 billion in 2025 to $9.8 billion by 2033 (CoinLaw, 2026), reflecting particularly strong demand for the real-time monitoring capabilities discussed in Section 3.

Adoption is not, however, evenly distributed. A 2025 industry survey found that while 86 percent of surveyed companies now allocate more than 3 percent of revenue to anti-fraud measures and 76 percent report intensifying investment specifically in AI, machine learning, and analytics, adoption of real-time transaction monitoring, arguably the capability most dependent on AI, had reached only 62 percent of organisations surveyed (SEON, 2025). Small and medium-sized enterprises present a particular adoption gap: despite facing disproportionate exposure to fraud, with average per-incident losses exceeding $141,000 according to Association of Certified Fraud Examiners data cited by MarketsandMarkets (2025), such enterprises often lack the capital and technical capacity to deploy sophisticated AI-based systems, relying instead on more affordable cloud-based and API-led fraud-analytics services now emerging to serve this segment.

This is a good indicator of the tangible benefits that can be delivered through AI-powered fraud detection—The U.S. Department of the Treasury reported that improved fraud-detection processes incorporating machine learning allowed them to prevent and recover more than $4 billion in fraud and improper payments in fiscal year 2024, up from $652.7 million the previous fiscal year, with risk-based screening alone resulting in $500 million in prevented losses and $2.5 billion more in high-risk transaction prioritised review. While total online-payment fraud losses are expected to exceed $362 billion globally through 2028 (U.S. Department of the Treasury, 2024), these figures offer a tangible example of the ROI that effective AI fraud detection systems can bring, on a massive scale.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

6. Obstacles and restrictions of AI for fraud detection

While it has proven to be effective in detecting payment fraud, there are some major challenges associated with the deployment of AI in this area. Firstly, class imbalance is still a technical challenge: Fraudulent transactions are a small percentage of all transactions, as shown in the 95 fraudulent transactions in a 141,174-transaction sample set studied by Preciado Martínez and Náñez Alonso (2025). Typically, this involves resampling techniques, cost-sensitive learning, or synthetic data generation methods, with each approach having a set of its own trade-offs between precision and recall.

Secondly, there is the risk of adversarial adaptation, which is increasing. Fraud-detection models are built using past data; therefore, they are reactive to fraud patterns that have already appeared in the past, and fraudsters continually evolve their fraud methods to avoid the detection signatures that have been previously discovered. But as fraudsters now leverage generative AI in their scams, reported growth of AI-based social engineering attacks has reached at least 61 percent in the last three years, deepfake fraud attempts are increasing by 2,100 percent, and losses from deepfake fraud are estimated to total around $1.5 billion from 2024 to 2026 (CoinLaw, 2026). This imbalance forces institutions to retrain and redeploy models more often than in the past, when software updates occur.

Thirdly, there are tensions between explainability and regulatory compliance. These complex models, like deep neural networks, can be powerful but blackbox systems that are hard for institutions to explain to regulators as well as auditors or affected customers why a specific transaction was flagged or rejected. This challenge is especially pronounced in jurisdictions where consumer-protection and anti-discrimination laws are stringent, as financial institutions will need to be able to prove that their automated decisions are not systematically discriminatory against specific demographic groups, geographic regions or transaction types.

Fourth, there are still issues of data privacy and cross-institutional data sharing. However, data sharing between institutions is a very effective tool for fraud detection, especially graph-based and network-analytic methods, as fraud rings often span several banks and payment systems. But privacy laws, competitive issues and the nature of financial information make it difficult to share transaction information, bringing the benefits of a network to AI-based fraud detection detection into a degree of compromise.

Fifth, Costs and capabilities limit adoption, especially for smaller institutions and SMEs as discussed in Section 5. The advanced AI fraud-detection tools demand both the computational power and expertise in data science to develop and maintain them, as well as integrating with existing transaction-processing systems, making them a resource-intensive challenge that may prove easier for smaller market participants to evade than for larger institutions as they beef up their defenses.

Lastly, the over-dependence on automation systems also comes with its share of risks, including false positive alarms that restrict legitimate transactions and friction costs due to the increased burden of relying solely on AI outputs without sufficient human oversight, which can give rise to the propagation of systemic errors or biases from training data at scale until they are identified and rectified. However, in the Indian context, it was found that most of the victims of some of the UPI scams were also deceived into authorising the fraud themselves, which transaction-level AI screening cannot stop all of (Business Standard, 2025).

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

7. Recommendations and Future Directions

The above evidence provides a basis for the following recommendations for financial institutions, regulators, and payment system operators to enhance fraud detection using AI.

However, institutions should implement a layered (hybrid) detection system that uses more than one detection technique – that is, a combination of supervised classification, unsupervised anomaly detection, and graph-based network analysis – because each of these techniques has its own strengths and weaknesses in detecting different fraud typologies.

Second, it is crucial to institutionalise continuous training and monitoring pipelines, rather than considering these actions as maintenance activities, as fraud tactics are evolving, especially through the use of AI, social engineering and deepfakes.

Third, regulators need to further develop options for secure, private data sharing between institutions – for example, India's Central Payments Fraud Information Registry and cross-bank platforms like the RBI's proposed Digital Payments Intelligence Platform (Medianama, 2025).

Fourth, explainable-AI techniques should be prioritised alongside raw predictive performance, particularly for models used in consumer-facing decisions such as transaction declines, to satisfy both regulatory requirements and customer-trust considerations. Where highly complex models such as deep neural networks are used, institutions should pair them with post-hoc explanation methods or maintain simpler, interpretable models for audit and appeal purposes.

Fifth, given the persistent gap between large-institution and SME adoption of AI-based fraud tools, policymakers and industry bodies should support the development of affordable, shared or utility-model fraud-detection services, such as cloud-based, API-accessible platforms, that allow smaller businesses to access AI-grade protection without prohibitive upfront investment.

Sixth, technological defences must be complemented by sustained consumer-education efforts, since a substantial share of fraud, particularly UPI-related fraud in India, continues to succeed through social engineering that induces victims to authorise transactions themselves; the finding that 51 percent of Indian fraud victims did not report incidents (Business Standard, 2025) further suggests a need for simplified, better-publicised reporting channels alongside detection technology.

Lastly, future research should emphasize developing more focussed detection methods that can be adapted to combat generative-AI enabled attacks such as the bypass of identity verification system by deepfakes, and AI-generated phishing content; as of yet, the rate of academic and regulatory progress seems to be lagging behind the pace at which attacks are being enacted in the industry, as reported in recent industry data (CoinLaw, 2026).

8. Conclusion

This paper has examined the expanding role of Artificial Intelligence in securing digital online payment systems against an increasingly sophisticated and rapidly evolving fraud landscape. The evidence reviewed demonstrates that AI-based techniques, particularly ensemble machine-learning methods such as random forest, complemented by unsupervised anomaly detection, deep learning, natural language processing, behavioural biometrics, and graph-based network analysis, consistently achieve high detection

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

accuracy across both academic studies and production deployments, with reported accuracy levels frequently exceeding 92 percent and, in some experimental settings, approaching 100 percent. The Indian UPI ecosystem illustrates the scale at which these technologies must now operate: hundreds of billions of transactions processed annually, alongside fraud losses that, while representing a small percentage of overall transaction value, translate into substantial absolute sums and real harm to individual users, more than half of whom, by some estimates, do not report incidents when they occur.

At the same time, this paper has shown that AI is not a complete or self-sufficient solution. Class imbalance, adversarial adaptation by increasingly AI-literate fraudsters, explainability deficits, data-privacy constraints, and uneven adoption across institutions of different sizes all limit the extent to which AI alone can secure digital payment ecosystems. The emergence of generative-AI-enabled fraud, including deepfake impersonation and AI-crafted phishing content that now characterises a majority of phishing attempts, represents a particularly significant escalation, effectively turning the same technological capability that powers modern fraud defences into a tool for more convincing attacks. This dynamic underscore that AI-based security is best understood not as a fixed solution but as one component of a continuously evolving contest between defensive and adversarial applications of the same underlying technology.

Going forward, the most resilient approach to securing digital payment systems will likely combine layered AI architectures, cross-institutional data-sharing frameworks, explainable-AI methods that satisfy regulatory and consumer-trust requirements, targeted support for smaller institutions facing adoption barriers, and sustained investment in consumer awareness and reporting infrastructure. As digital payment volumes continue to grow across both developed and emerging markets, the capacity of AI-driven fraud detection to adapt at least as quickly as the threats it is designed to counter will remain a central determinant of trust and stability in the global digital-payment ecosystem.

REFERENCES:

1. Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal, 6, 100163. https://doi.org/10.1016/j.dajour.2023.100163
2. Association for Financial Professionals. (2026). Over 75% of US firms experienced payments fraud in 2025, while AI adoption for fraud mitigation lags [Press release]. https://www.financialprofessionals.org/about/learn-more/press-releases/Details/over-75-percent-of-us-firms-experienced-payments-fraud-in-2025-while-ai-adoption-for-fraud-mitigation-lags
3. Business Standard. (2025, June 26). 1 in 5 UPI users faced fraud; 51% victims didn't report, reveals survey. https://www.business-standard.com/finance/news/upi-transaction-fraud-india-survey-one-in-five-users-hit-localcircles-125062601141_1.html
4. Business Today. (2024, June 1). Digital payment frauds surge in India as UPI transactions skyrocket: RBI report. https://www.businesstoday.in/technology/news/story/digital-payment-frauds-surge-in-india-as-upi-transactions-skyrocket-rbi-report-431695-2024-06-01

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 5, September-October 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i05.1510

5. BNY. (2026). How AI is changing payments fraud and fraud prevention. https://www.bny.com/corporate/global/en/insights/ai-and-payments-fraud-an-evolving-landscape.html
6. CoinLaw. (2026). Digital payment fraud statistics 2026: Alarming trends now. https://coinlaw.io/digital-payment-fraud-statistics/
7. Grand View Research. (2026). Fraud detection and prevention market report, 2026-2033. https://www.grandviewresearch.com/industry-analysis/fraud-detection-prevention-market
8. Markets and Markets. (2025, July 28). Fraud detection and prevention market worth $65.68 billion by 2030 [Press release]. https://www.marketsandmarkets.com/PressReleases/fraud-detection-prevention.asp
9. Medianama. (2025, May 30). UPI handles 84% of India's retail payment volume in FY25: RBI. https://www.medianama.com/2025/05/223-upi-84-india-fy25-retail-payment-volume-rbi/
10. Preciado Martínez, P. M., & Náñez Alonso, S. L. (2025). Comparative analysis of machine learning models for the detection of fraudulent banking transactions. Cogent Business & Management, 12(1), Article 2474209. https://doi.org/10.1080/23311975.2025.2474209
11. Press Information Bureau, Government of India. (2025). Digital payment transactions surge with over 18,000 crore transactions in 2024-25. https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2110405
12. PwC India. (2025). Combating payments fraud in India's digital payments landscape. https://www.pwc.in/ghost-templates/combating-payments-fraud-in-Indias-digital-payments-landscape.html
13. Reserve Bank of India. (2024). Annual report 2023-24. Reserve Bank of India.
14. Reserve Bank of India. (2025). Annual report 2024-25. Reserve Bank of India.
15. SEON. (2025, March 5). SEON 2025 digital fraud report reveals surge in fraud prevention budgets & headcounts [Press release]. Business Wire. https://www.businesswire.com/news/home/20250305401152/en/SEON-2025-Digital-Fraud-Report-Reveals-Surge-in-Fraud-Prevention-Budgets-Headcounts
16. U.S. Department of the Treasury. (2024, October). Treasury announces enhanced fraud detection processes, including machine learning AI, prevented and recovered over $4 billion in fiscal year 2024 [Press release]. https://home.treasury.gov/news/press-releases/jy2650
17. U.S. Bank. (2026). How treasury departments use AI to detect and prevent fraud. https://www.usbank.com/corporate-and-commercial-banking/insights/risk/mitigation/treasury-dept-partners-using-ai-to-fight-fraud.html
18. ZIGRAM. (2026). Digital payment fraud statistics 2025. https://www.zigram.tech/resources/digital-payment-fraud-statistics-2025/