International Journal of Science, Strategic Management and Technology
Volume **02** Issue **07** July-2026 | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

The Use of Artificial Intelligence in Banking Fraud Detection: A Systematic Review Using the PRISMA Approach

Dr. Harshita Gupta[1], Aditya Gupta[2]

[1] Devi Ahilya Vishwavidyalaya
[2] Indian Institute of Technology, Patna

https://doi.org/10.55041/ijsmt.v2i7.088

Cite this Article: Gupta, H. & Gupta, A. (2026). The Use of Artificial Intelligence in Banking Fraud Detection: A Systematic Review Using the PRISMA Approach. International Journal of Science, Strategic Management and Technology, *02*(7), 1-9. https://doi.org/10.55041/ijsmt.v2i7.088

Abstract

The rise of digital banking has been accompanied by a corresponding rise in both the scale and the sophistication of financial fraud, thereby highlighting the structural limitations of traditional, rule-based detection systems. This study reviews the existing academic and industry literature on the use of artificial intelligence (AI) and machine learning (ML) in banking fraud detection using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework. Following a clearly defined search and screening process, six full-text sources were used for full qualitative analysis. These include empirical survey research, hybrid statistical modelling, and prior systematic reviews. The study was also supplemented by a total of seventeen secondary references drawn from within those sources' own bibliographies. It was found that supervised learning, anomaly detection, deep learning, and ensemble techniques are the predominant techniques for fraud detection, with the highest improvement of ensemble and hybrid systems compared to single models. Empirical adoption studies rooted in the Technology Acceptance Model consistently identify technological knowledge, perceived usefulness, and organisational training — life more than algorithmic sophistication alone — as the most important predictors of long-term AI use. In addition, the most common barriers in the use of AI include shortage of skilled personnel, expensive computation, and evolving fraud tactics. Ethical and regulatory issues, including data privacy, algorithm bias, model transparency, and compliance with laws such as GDPR, PCI-DSS, and Anti-Money-Laundering law, emerge as a consistent and underdeveloped concern across the literature. The review summarizes that although AI represents substantial improvement in fraud-detection accuracy and speed when compared to rule-based systems, the existing remains geographically concentrated and methodologically descriptive, pointing to a need for more rigorous, cross-institutional, and cross-national research.

Keywords: Artificial Intelligence; Machine Learning; Fraud Detection; Digital Banking; PRISMA; Systematic Review

1. Introduction

Digital banking has fundamentally reshaped how people and firms' access, transmit, and manage money. Through online and mobile banking, digital wallets, and instant payment, banking services are extended to previously underserved populations while cutting the operational cost of routine banking activity (Serifat et al., 2025). Unfortunately, the same convenience created by digital banking has expanded the attack surface available to fraudsters. In fact, currently the digital world is susceptible to identity theft, phishing, card-not-present fraud, transaction manipulation, and money laundering at a scale and speed that manual, rule-based control systems are unable to handle (Omorogbe et al., 2025). Such a situation has facilitated stealing from banks at enormity. In Nigeria alone, the increase in recorded fraud cases between 2019 and 2023 reached 112%, with associated monetary losses increasing by 496% over the same period

(Omorogbe et al., 2025) — a pattern of accelerating loss that recurs, in different magnitudes, across banking markets worldwide.

Traditional fraud detection has been dependent on predefined rules (such as flagging any transaction above a fixed value or multiple withdrawals from different locations within a short window) and manual review of flagged cases by bank staff. Both methods can only detect fraud patterns that have been identified and programmed, and they become less effective when facing high volumes and advanced patterns (Serifat et al., 2025).This problem has led to persistent interest in methods, across both academic research and industry practice, in artificial intelligence and machine learning as a more adaptive, scalable alternative — systems that learn directly from transaction data, detect previously unseen patterns, and improve continuously as new fraud tactics emerge.

Even though the numerous individual studies on this subject, the evidence scatters across different fields (computer science, finance, information systems), geographies, and methodologies (technical algorithm papers, employee-adoption surveys, and industry case studies), making it difficult for a researcher, practitioner, or policymaker to get a unified view of the issue. This study addresses that gap by providing a systematic review of the literature based on AI-driven fraud detection in banking, guided by the PRISMA guidelines.

1.1 Objectives of the Study

- Identifying the most reported types of fraud in digital banking.
- Determining the AI and machine learning techniques used in fraud detection in banking.
- To compile empirical evidences showing how effective AI helps in improving fraud-detection outcomes.
- To examine the organisational, ethical, and regulatory factors shaping the adoption of AI-driven fraud detection.
- To identify gaps in the current literature and propose directions for future research.

1.2 Research Questions

1. Which types of fraud are most coomon in digital banking, in agreement with the reviewed literature?
2. What AI and machine learning techniques are employed for fraud detection, and how are these categorised?
3. What empirical evidence exists on the effectiveness of AI relative to traditional fraud-detection methods?
4. Which organisational, ethical, and regulatory factors influence the usage of AI-driven fraud detection in banks?
5. Which aspects of analysis have not been scrutinized yet?

The structure of this document is given as follows. The second section assesses the theoretical and conceptual literature related to fraud, digital banking, and AI. The third section elaborates on the research methodology of PRISMA. The fourth section presents the findings of the systematic review based on the themes of research questions. The final section summarizes the implications, limitations, and suggestions for future research.

2. Literature Review

2.1 Digital Banking and the Evolving Fraud Landscape

Digital banking is driven by technology, the internet, and customers' needs, motivated by the shift from physical to remote financial service delivery (Omorogbe et al., 2025). It enables users to check balances, transfer funds, apply for loans, and pay their bills from any location, and has also enabled banks to extend a broader range of e-payment services to their customers. However, with this convenience comes, wider range of vulnerability available for exploitation.

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 07 July-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

According to Serifat et al. (2025), digital banking fraud refers to using unauthorised and illegal use of digital banking systems to steal, manipulate, or compromise financial data for personal gain, and identify identity theft, phishing, transaction manipulation, and money laundering as its dominant forms. Omorogbe et al. (2025) found out in their survey of 94 employees working in five Nigerian banks that credit card fraud is the most common type of fraud because of having the highest average (mean = 3.80) score on a 4-point scale, while phishing schemes, identity theft, payment fraud, and forgery were rated 3.72, 3.60, 3.53, and 3.43 respectively.

2.2 Limitations of Rule-Based and Manual Fraud Detection

The conventional approach of fraud detection involves the use of rule-based frameworks where certain criteria, such as high financial transactions or cash saving from numerous points, are defined beforehand and flagged for manual intervention. The two major flaws of the system are that conventional rule-based methods can identify only old types of fraud and the manual process is laborious and slow. According to existing research, this causes a constant issue of working with excessive false positives and undetected fraud (Omorogbe et al., 2025).

2.3 Machine Learning and Deep Learning Approaches

Some researchers in the literature recognize a common classification of the methods of artificial intelligence and machine learning to detect fraud with the following categories: supervised learning, unsupervised learning, deep learning, and ensemble methods (Serifat et al., 2025; Zakaria et al., 2023). Supervised methods make use of already trained models which include decision trees, random forests, and logistic regression. These methods involve analyzing previously collected data on either fraudulent or legit transactions so that the model is able to generate a statistical connection with such input variables as transaction amount, transaction time, and the location of the transactions among others. Meanwhile, the unsupervised methods employ models such as K-means clustering and isolation forest which do not need the training of labelled data to detect transactions that are out of the limits of the learned pattern of behaviour. Lastly, deep learning methods employ such models as convolutional neural networks and long short-term memory neural networks to process information coming from transactions in terms of multimedia and chronological aspects. Natural language processing (NLP) is used to analyse transaction narratives, customer communications, and financial-document text for indicators of fraud, while expert systems encode rule bases derived from domain expertise and offer more transparent, auditable decision logic at the cost of adaptability (Zakaria et al., 2023).

2.4 Adoption, Effectiveness, and Organisational Factors

Beyond technique description, a smaller body of empirical work directly measures AI's effect on fraud-detection outcomes and the organisational conditions shaping its adoption. Grounded in the Technology Acceptance Model (TAM), such studies converge on a shared insight: adoption and continued use of AI-based fraud-detection tools depend less on the sophistication of the underlying algorithm than on employees' technological knowledge, perceived usefulness, and the organisational investment in training (Goyal et al., 2025; Omorogbe et al., 2025). This is discussed at length in Section 4.

2.5 Ethical, Regulatory, and Governance Considerations

A recurring concern across the literature is that AI-based fraud detection must operate within, rather than outside, existing financial regulation. Frameworks cited include the General Data Protection Regulation (GDPR), the Payment Card Industry Data Security Standards (PCI-DSS), and Anti-Money-Laundering (AML) law (Serifat et al., 2025). Algorithmic bias, model transparency, and data privacy are the most frequently cited ethical concerns: models trained on historically biased data risk reproducing and amplifying those biases, and the 'black box' character of many deep learning models undermines customers' and regulators' ability to understand — and contest — high-stakes decisions such as blocking a transaction or freezing an account (Chethan et al., 2024; Serifat et al., 2025).

3. Research Methodology

3.1 Review Design

The methodology applied in the current study follows a systematic literature review guided by the PRISMA method (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Its structure consists of four phases: identifications, screening, eligibility, and inclusion. A narrative synthesis approach was used to combine findings from the studies instead of statistical meta-analysis because the studies included are very different in terms of methodology and therefore cannot be statistically compared.

3.2 Search Strategy

The compilation of records was done through a mixture of the systematic approach of academic and web search, which included the searching of Google Scholar-indexed resources, IEEE Xplore, SpringerLink, Elsevier/ScienceDirect, and ResearchGate in the combination of such terms as 'artificial intelligence', 'artificial intelligence', 'fraud detection', 'banking', 'credit risk', 'anomaly detection', 'cyber security', 'financial technology'. The involvement of both keyword-based search and the texts provided by the researcher is regarded as the initial phase of the review.

3.3 Eligibility Criteria

Inclusion Criteria	**Exclusion Criteria**
Peer-reviewed journal articles, conference papers, or book chapters	Duplicate records identified across multiple search sources
Published in English	Non-English-language publications
Directly addresses AI/ML applications in banking fraud detection, credit risk, or closely related financial security domains	Studies unrelated to banking or financial services
Full text retrievable and reviewable by the researcher	Sources for which only an abstract, landing page, or paywalled/access-restricted link was available

3.4 Study Selection (PRISMA Flow)

Initially, a total of 24 records were found during the review process. Out of these, one record (Goyal et al., 2025) was discarded, as it appeared under two links of different origins. Therefore, the final count of records for examination was 23. Of these, seventeen could not be progressed to full-text eligibility assessment because they were link-only, paywalled, blocked by the host site's access controls, or lacked retrievable full bibliographic detail. This left six records that were retrieved and reviewed in full text; all six met the inclusion criteria and were retained for qualitative synthesis. The seventeen excluded records were not discarded outright: where they were independently cited within the bibliographies of the six core full-text studies, they are referenced in this paper as secondary, contextual citations, but they were not independently verified in full text by the researcher and are treated with corresponding caution. Figure 1 summarises this process.

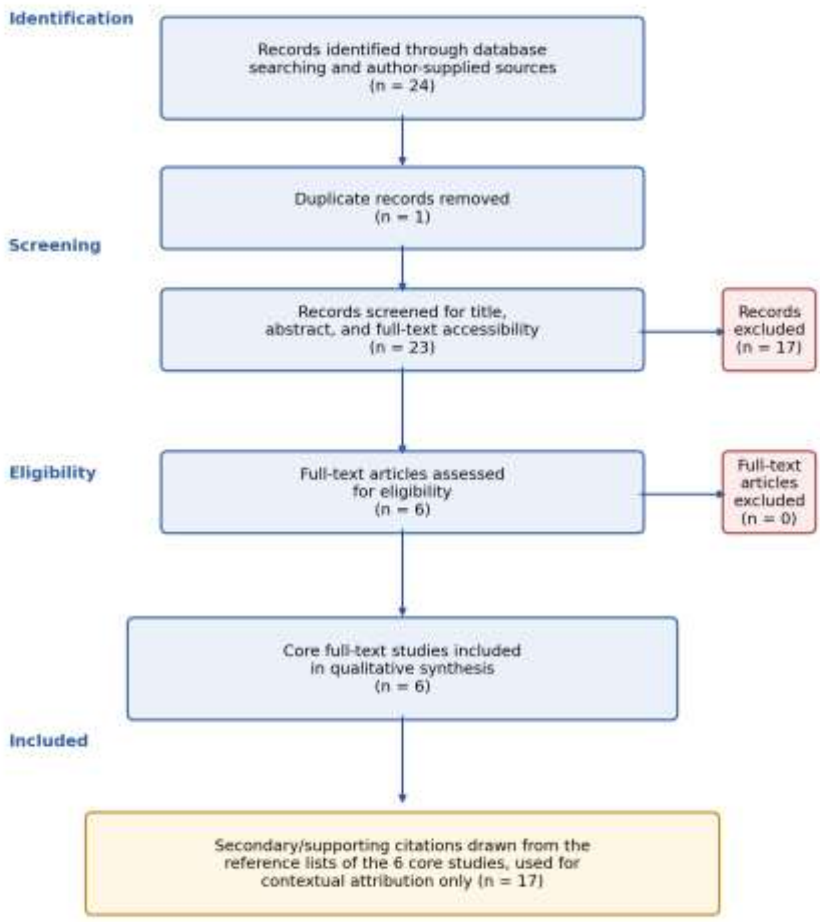

Figure 1. PRISMA flow diagram of the study selection process.

3.5 Data Extraction and Synthesis

For each of the six core studies, the following data were extracted: authors and year, country/context, research design, sample or corpus, and key findings relevant to the five research questions. These are summarised in Table 1 (Section 4.1). Findings were then grouped thematically — fraud typology, AI/ML technique taxonomy, empirical effectiveness evidence, adoption/organisational factors, and ethical/regulatory considerations — corresponding to the structure of the Findings section below.

3.6 Limitations of the Methodology

This review's corpus is smaller and more researcher-curated than a full independent multi-database systematic search would typically produce, and it is geographically concentrated (with Nigeria and India particularly well represented among the six core studies). These limitations are discussed further in Section 5.2 and should be considered when interpreting the generalisability of the findings.

4. Findings and Discussion

4.1 Overview of Included Studies

Study	Context	Design	Key Contribution
Serifat et al. (2025)	Global / conceptual	Narrative review of secondary literature	Taxonomy of AI/ML fraud-detection techniques; ethical, legal, and emerging-technology discussion
Omorogbe et al. (2025)	Nigeria (5 banks, n = 94)	Descriptive survey + linear regression	Empirical fraud typology; AI impact and adoption-barrier rankings; TAM-based hypothesis test
Goyal et al. (2025)	India (n = 271)	Hybrid SEM-ANN survey	TAM-based drivers of continued AI adoption for credit risk and fraud detection
Chethan et al. (2024)	Global (47-study corpus, India-heavy)	Bibliometric + content analysis SLR	Country/year publication and citation trends; opportunities, challenges, applications, and bank-level implementation inventory
Zakaria et al. (2023)	Global / conceptual	Narrative review of secondary literature	Historical development of AI in finance; ML/NLP/expert-system taxonomy; fraud, risk, and behavioural-finance applications
IBM (2025)	Industry / global	Industry explainer	Real-world deployment examples (graph neural networks, LSTM-based risk scoring) at major global banks

4.2 RQ1 — Types of Fraud in Digital Banking

The diverse body of literature reveals the same five types of fraud: credit card fraud, phishing, identity theft, payments fraud, and forgery, with the addition of money laundering in the conceptual discussion. The research done by Omorogbe et al. (2025) yields the most straightforward and ranked proof so far, which indicates the prevalence of credit card fraud (mean = 3.80) and forgery (mean = 3.43), while the other three types did not get a score lower than 3.61 overall score. This suggests that despite the emergence of newer types of digital fraud, long-known types, like credit card fraud and phishing, are still major concerns for specialists at banks that have recently begun to implement AI technologies in order to deal with newly emerging issues of fraud.

4.3 RQ2 — AI and Machine Learning Techniques

This synthesis validates the existence of a four-part taxonomy of techniques, which includes the following: supervised learning (which includes decision trees, random forests, and logistic regression), unsupervised learning (includes K-means clustering, isolation forests, and one-class SVM), deep learning (which includes CNNs and LSTMs), and ensemble methods (which include bagging, boosting, and gradient boosting or XGBoost). An additional layer composed of natural language processing and expert systems facilitates analyses of unstructured transaction narratives as well as common knowledge-based audit rule encoding (Zakaria et al.,2023). According to Serifat et al. (2025), the best performance has been observed when employing ensemble and integrated technology systems, with a study across 25 financial institutions confirming a 15-25% increase in fraud detection rate when using NLP, anomaly detection, and supervised learning rather than only one of these methods.

4.4 RQ3 — Empirical Evidence on Effectiveness

Though there are still relatively few direct empirical investigations into the effects of AI on fraud-detection, findings are uniformly favorable. Omorogbe et al. (2025) analyzed the relationship between AI use and fraud detection in Nigerian banks and found a strong positive correlation of ($\beta = 0.885$, $t = 18.26$, $p < .001$), meaning AI usage accounted for 78% of the variance in the outcome of detection of fraud ($R^2 = .784$). Employees in the banks surveyed stated that enhancing accuracy is the most important benefit of AI (3.98), ahead of speedy response to frauds (millisecond level) and saving from financial losses (3.93). Industry data (IBM, 2025) on the deployment of AI by multinational banks corroborates this and shows that graph neural networks help them to identify fraud rings while LSTM-based models are being employed in the risk scoring process in large payment organizations and credit card issuers.

4.5 RQ4 — Organisational, Ethical, and Regulatory Factors

The results of adoption studies based on the Technology Acceptance Model yield virtually the same outcome in two different national contexts. In India, it was found that individuals' technology knowledge and perceived usefulness (and not just ease of use alone) were the most important factors in ensuring continued AI use among 271 banking experts; attitude was found to act as a partial mediator, while perceived risks (data breaches, algorithmic bias, and compliance with regulations) showed a negative but statistically insignificant correlation with continued use. In Nigeria, a similar finding regarding skilled manpower emerged with a new obstacle to the application of AI-powered fraud detection (mean = 3.96) considered more serious than high computing costs (mean = 3.91) and the challenge of keeping up with changing fraud schemes (mean = 3.81); the most efficient way to eliminate these obstacles was deemed training and development with the mean score of 3.99.

Regarding the ethical and regulatory aspect, existing literature seems to be grounded more in theories than in practical experiences. According to Serifat et al. (2025) and Chethan et al. (2024), the main concerns that need to be resolved are those relating to data protection, algorithmic bias and openness of models ('black box' making). The two papers do not, however, provide quantitative assessment of bias and discrimination in fraud detection systems, which the present review identifies as an important gap (see Section 5.2).

4.6 RQ5 — Institutional Applications

A bibliometric inventory compiled by Chethan et al. (2024) highlights the implementation of AI in India's leading banks. In this context, let us consider three prominent examples: the 'AI assistant' of State Bank of India (with the capability of managing up to 10,000 inquiries per second), the chatbot 'Eva' of HDFC Bank (with 85% success in solving inquiries), and the 'iPal' of ICICI Bank (having about 90% accuracy alongside performing 200 different operations allowing for implementation of robotic process automation). It is important to note that these examples are based not so much on the fraud detection but rather on the customer service capabilities of the banks. However, they are telling us about the institutional arrangements (such as chatbots, predictive analytics, RPA) through which fraud detection systems are implemented and delivered to users.

5. Conclusion

5.1 Summary of Findings

In this review, the purpose was to synthesise literature on the use of AI for fraud detection in digital banking by applying PRISMA methodology. The review covered six core full-text articles, which indicate that use of AI and machine learning enhances accuracy and speed of fraud detection in comparison with conventional rule-based methods, while several methods such as ensemble and integrated methods reach the best results. Credit card fraud, phishing and identity theft are the main fraud types that practitioners talk about. Interestingly enough, the best predictors of successful implementation of AI is mostly organisational, including technological readiness, perceived usefulness and staff training rather than something technical. Ethical and regulatory issues related to fraud detection still exist, but they are often addressed just conceptually.

5.2 Limitations

There are a number of limitations within this review. First, the collection of six main full-text sources is small compared to an independent, multi-database systematic search and from the seventeen additional records, access restrictions lead to them not going beyond the screening stage; therefore, the next version of this review would be facilitated by institutional database access (Scopus, Web of Science, etc.) to provide full texts of the mentioned sources. Second, the body of evidence is concentrated in terms of geography, with India and Nigeria featured particularly well and some major banking markets (North America, Europe, and East Asia) less featured among the full-text literature taken into account here. Third, the fact that the original studies significantly differ in methodology leads to the necessity of using narrative synthesis instead of statistical meta-analysis.

5.3 Directions for Future Research

- Independent, multi-database systematic searches (Scopus, Web of Science, ScienceDirect) to expand the evidence base beyond the six core studies reviewed here.
- Comparative, cross-national adoption studies that test whether the TAM-based drivers identified in India and Nigeria generalise to other regulatory and institutional environments.
- Empirical (not solely conceptual) research quantifying algorithmic bias and discriminatory outcomes in deployed fraud-detection and credit-scoring systems.
- Studies that integrate technical, organisational, and regulatory dimensions of AI fraud detection within a single analytical framework, rather than treating them as separate literatures.
- Longitudinal evaluation of real-world fraud-detection performance (not only customer-service and operational-efficiency metrics) at the institutional level.

References

Anshari, M., Almunawar, M. N., Masri, M., & Hrdy, M. (2021). Financial technology with AI-enabled and ethical challenges. Society, 58(3), 189–195.

Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. Innovative Technology at the Interface of Finance and Operations, I, 223–247.

Chen, X., Li, Z., & Zhang, Y. (2019). Enhancing fraud detection using GAN-augmented data in financial transactions. Journal of Financial Technology, 8(3), 245–260.

Chethan, Dr., Munilakshmi, R., Dr., & Ramesh, L., Dr. (2024). Role of artificial intelligence in banking sector: A systematic literature review. Educational Administration: Theory and Practice, 30(1), 7459–7481. https://doi.org/10.53555/kuey.v30i1.10646

Choi, D., & Lee, K. (2018). An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation. Security and Communication Networks, 2018.

Fernández, A., García, S., & Herrera, F. (2020). Machine learning for banking security: A comparative study of algorithms. Expert Systems with Applications, 116, 200–215.

Ghosh, R., Mitra, P., & Banerjee, A. (2021). Anomaly detection in credit risk using variational autoencoders. International Journal of Data Science and Analytics, 14(1), 87–95.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.

Goyal, K., Garg, M., & Malik, S. (2025). Adoption of artificial intelligence-based credit risk assessment and fraud detection in banking services: A hybrid approach (SEM-ANN). Future Business Journal, 11(1), Article 44. https://doi.org/10.1186/s43093-025-00464-3

IBM. (2025). AI fraud detection in banking. IBM Think. https://www.ibm.com/think/topics/ai-fraud-detection-in-banking

Li, X., Huang, J., Chen, C., & Zhang, Y. (2020). Credit card fraud detection using machine learning: A systematic literature review. Expert Systems with Applications, 164, 113909.

Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. International Journal of Financial Studies, 9(3), 39.

Nkomo, B. K., & Breetzke, T. (2020, March). A conceptual model for the use of artificial intelligence for credit card fraud detection in banks. In 2020 Conference on Information Communications Technology and Society (ICTAS) (pp. 1–6). IEEE.

Olusola, O. J., & Alaba, T. (2020). Comparative analysis of machine learning algorithms for fraud detection in electronic banking transactions. International Journal of Computer Applications, 176(36), 12–19. https://doi.org/10.5120/ijca2020919811

Oluwaseun, A., & Adepoju, S. (2021). Deep learning and natural language processing techniques for fraud detection in African banking. Journal of Applied Computing Research, 21(3), 133–142.

Omorogbe, O. H., Eduje, A. I., Anyanwu, L., Obeka, O. B., Ukaoha, K. C., Izevbuwa, O. G., Ighotuweyin, A. F., & Arenvbaguehita, O. D. (2025). The impact of artificial intelligence (AI) on fraud detection in banks in Edo State. GAS Journal of Engineering and Technology (GASJET), 2(9), 26–35. https://doi.org/10.5281/zenodo.17208855

Schemmel, J. (2020). Artificial intelligence and the financial markets: Business as usual? Regulating Artificial Intelligence, 255–276.

Serifat, O. A., Igah, R. C., Balogun, K. M., Mensah, G. R., & Odai, E. N. (2025). AI-driven fraud detection in digital banking: ML approach for secure and transparent financial transactions. American Journal of Financial Technology and Innovation, 3(1), 135–145. https://doi.org/10.54536/ajfti.v3i1.5168

Silva, R. D. (2021). Calls for behavioral biometrics as bank fraud soars. Biometric Technology Today, 2021(9), 7–9.

Soviany, C. (2018). The benefits of using artificial intelligence in payment fraud detection: A case study. Journal of Payments Strategy & Systems, 12(2), 102–110.

Wang, X., Liu, H., & Yu, Z. (2019). A hybrid approach for fraud detection in online banking transactions. IEEE Access, 7, 64101–64113.

Zakaria, S., Manaf, S. M. A., Amron, M. T., & Suffian, M. T. M. (2023). Has the world of finance changed? A review of the influence of artificial intelligence on financial management studies. Information Management and Business Review, 15(4-SI), 420–432.