RESEARCH PAPER

"Artificial Intelligence in Banking Fraud Detection: A Systematic Review using the Prisma Approach"

Vrushali Shivaji Takate[1*], Manisha Dinesh Balsaraf[2], Shaikh Lubna Mohammadgoush[3]

[1*]Pravara Rural Engineering College, Loni, Email: vrushali.takate@pravara.in
[2]Pravara Rural Engineering College, Loni, Email: manisha.balsaraf@pravara.in
[3]Poona Institute of Management Sciences and Entrepreneurship, Pune, Email: lubna.shaikhmd@gmail.com

***Corresponding Author:**
Vrushali Shivaji Takate
Pravara Rural Engineering College, Loni
Email: vrushali.takate@pravara.in

Abstract
The rapid growth of digital banking, instant payments, mobile applications and card-not-present transactions has improved access to financial services while simultaneously widening the attack surface for fraud. Conventional rule-based systems remain indispensable for regulatory control, yet they are often unable to recognize subtle, coordinated and previously unseen patterns. This paper presents a condensed systematic review of artificial intelligence (AI) in banking fraud detection, organised according to the PRISMA 2020 reporting logic. It synthesizes evidence on supervised learning, anomaly detection, deep learning, graph analytics, natural language processing and hybrid decision systems. The review shows that tree-based ensembles generally perform strongly on structured transaction data; sequential neural networks are useful for behavioural histories; graph methods are promising for mule-account and fraud-ring detection; and anomaly models support discovery where labels are scarce. However, accuracy alone is misleading in highly imbalanced datasets. Precision, recall, F1-score, precision–recall area, false-positive cost, detection latency and financial value saved are more meaningful. The review further identifies concept drift, delayed labels, data leakage, limited public datasets, explainability, privacy and bias as major implementation concerns. It concludes that the most credible banking architecture is layered, adaptive and human-supervised rather than dependent on a single algorithm.
PRISMA 2020 is used as the reporting framework because it improves transparency in identification, screening, eligibility assessment and synthesis of studies.1

Keywords: Artificial intelligence; banking fraud; machine learning; deep learning; anomaly detection; explainable AI; PRISMA; financial technology.

How to cite this article: Takate VS, Balsaraf MD, Mohammadgoush SL. Artificial intelligence in banking fraud detection: a systematic review using the PRISMA approach. Int J Drug Deliv Technol. 2026;16(74s): 2323-2330. DOI: 10.25258/ijddt.16.74s.194

1. Introduction

Banking has shifted from branch-centred service delivery to continuous digital interaction. Customers now transfer funds, open accounts, apply for credit and authenticate payments through interconnected platforms. This transformation has reduced transaction time and expanded financial inclusion, but it has also created opportunities for phishing, account takeover, synthetic identity fraud, merchant collusion, loan-application fraud, money-mule activity and unauthorized transfers. Modern fraud is frequently distributed across accounts, devices and channels; consequently, an isolated transaction may appear legitimate even when the wider pattern is suspicious.
AI-based fraud detection extends statistical monitoring by learning patterns from historical and streaming data and by producing a risk score for each event.2

Artificial intelligence offers a more adaptive approach than static thresholds. Machine-learning systems can combine transaction value, time, location, device identity, merchant category, account age, customer history and beneficiary relationships. The aim is not merely to label a payment as fraudulent, but to prioritize cases, support step-up authentication and direct investigators towards the most consequential alerts. AI is therefore best understood as a decision-support layer embedded within broader governance, cybersecurity and regulatory systems.

1.1 Problem Statement

The central problem is that fraud events are rare, dynamic and costly. A model can appear highly accurate by predicting almost every transaction as genuine, while still missing most fraud. At the same time, a highly sensitive model may produce excessive false positives,

*Author for Correspondence: vrushali.takate@pravara.in

block legitimate customers and overload investigation teams. The research challenge is therefore to identify models and evaluation practices that balance fraud capture, operational cost, customer convenience, interpretability and real-time speed.

1.2 Significance of the Study

This review is significant for banks, regulators and researchers because algorithmic claims are often based on different datasets, validation procedures and metrics. A transparent synthesis helps distinguish laboratory performance from practical usefulness and highlights the conditions under which a technique is likely to succeed or fail.

2. Objectives and Research Questions

The study pursues five objectives: (i) to review AI applications in banking fraud detection; (ii) to identify the principal supervised, unsupervised, deep-learning and graph-based techniques; (iii) to compare their strengths and limitations; (iv) to examine datasets, validation strategies and performance indicators; and (v) to identify methodological, ethical and operational research gaps.

The review is guided by the following questions: Which AI techniques are most frequently used? Which measures provide credible evaluation under class imbalance? What advantages do AI systems offer over conventional monitoring? Which barriers affect deployment in banks? What research priorities are necessary for reliable and responsible fraud detection?

3. Conceptual Background

3.1 Banking Fraud and Its Forms

Banking fraud is intentional deception or unauthorised activity intended to obtain money, credit, confidential information or another financial advantage. Its major forms include credit- and debit-card fraud, account takeover, internet-banking fraud, loan and cheque fraud, insider abuse, identity theft, synthetic identities, phishing, money laundering and coordinated mule-account networks. These forms overlap: stolen identity data may be used to open accounts, obtain credit and move funds through several intermediaries.

3.2 Limitations of Conventional Rules

Rules remain necessary because they translate institutional policy and regulatory obligations into explicit controls. Nevertheless, rules usually depend on known patterns, fixed thresholds and manual maintenance. Fraudsters can adapt amounts and timing to avoid triggers, while legitimate behavioural changes can generate large numbers of alerts. Rules are thus effective for compliance and obvious red flags, but insufficient as the only detection mechanism.

The continuing importance of data-mining and intelligent detection arises from the need to discover complex and changing fraud patterns beyond fixed rules.3

3.3 Role of AI

AI supports predictive classification, clustering, anomaly detection, sequence analysis, network analysis, document screening and behavioural profiling. Its practical contribution lies in ranking risk, reducing the investigative search space and updating detection logic as patterns evolve. It should complement, rather than displace, human judgement and regulatory controls.

4. Research Methodology

4.1 Review Design and Period

The paper adopts a systematic qualitative review design. The principal period covers 2015 to June 2026, while foundational publications are included where necessary to explain established algorithms and evaluation methods. A formal meta-analysis is not attempted because the studies differ substantially in datasets, fraud definitions, sampling, model architecture and reported outcomes.

4.2 Information Sources and Search Strategy

Relevant studies were located through Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, Google Scholar and selected regulatory or institutional publications. Representative search combinations linked AI terms—"machine learning," "deep learning," "anomaly detection," "graph neural network" and "explainable AI"—with fraud terms such as "banking fraud," "credit card fraud," "account takeover," "payment fraud," "loan fraud" and "money laundering."

4.3 Eligibility Criteria

Studies were included when they examined computational intelligence for banking or financial-transaction fraud, reported a recognisable method or dataset, presented empirical or comparative findings and addressed performance, implementation, explainability, privacy or imbalance. Studies were excluded when they focused only on general cybersecurity, non-financial fraud, promotional descriptions, duplicate analyses or results without adequate methodological detail.

4.4 PRISMA Selection Logic

PRISMA Stage	Review Action
Identification	Records located through databases and supplementary searching
Screening	Duplicates removed; titles and abstracts examined
Eligibility	Full texts assessed against inclusion and exclusion criteria
Included	Eligible studies synthesised thematically

The PRISMA approach does not guarantee study quality by itself; rather, it requires transparent reporting of how evidence was located and selected.4

4.5 Quality Assessment and Data Extraction

Each study was examined for clarity of objectives, dataset source and size, representation of fraud, preprocessing, validation, performance measures, baseline comparison, reproducibility, practical relevance and ethical governance. Extracted information included publication year, fraud type, dataset, algorithm, imbalance-handling method, validation design, principal findings, limitations and implementation implications.

5. AI Techniques for Banking Fraud Detection

5.1 Supervised Machine Learning

Supervised learning uses labelled examples of fraudulent and legitimate activity. Logistic regression remains a valuable baseline because it is computationally efficient and interpretable, although it may miss complex non-linear interactions. Decision trees produce understandable rules but can overfit. Support Vector Machines can perform well in high-dimensional spaces but may be difficult to scale to very large transaction streams.

Random Forest reduces the instability of a single tree by aggregating many trees trained on resampled data and feature subsets.5

Random Forest and gradient-boosting models are widely used for structured transaction data. Random Forest handles non-linearity and noisy features, while XGBoost, LightGBM and CatBoost sequentially correct earlier errors and often deliver strong precision–recall trade-offs. Their limitations include parameter tuning, reduced transparency and sensitivity to how training and test data are constructed.

XGBoost is a scalable tree-boosting framework designed for efficient regularised learning on large structured datasets.6

5.2 Comparative Interpretation

No classifier is universally best. Interpretable models are advantageous where decisions require direct explanation; ensembles are useful when predictive performance on tabular data is the priority; and cost-sensitive variants are preferable when the financial consequences of false negatives and false positives differ substantially.

5.3 Unsupervised and Semi-Supervised Learning

Fraud labels are incomplete because detection may be delayed and some fraudulent events remain undiscovered. Unsupervised methods search for unusual behaviour without relying on complete labels. Clustering groups transactions by similarity and highlights distant observations, while one-class models attempt to learn the normal class. Their central limitation is that unusual does not necessarily mean fraudulent; a customer travelling abroad or making an exceptional purchase may appear anomalous but legitimate.

Isolation Forest detects anomalies by isolating observations through random partitions; anomalies generally require fewer partitions than ordinary cases.7

5.4 Deep Learning

Artificial neural networks can represent complex interactions among transaction features. Autoencoders learn to reconstruct normal observations and flag records with high reconstruction error. Recurrent neural networks and Long Short-Term Memory networks examine sequences, making them suitable for sudden changes in spending, beneficiary use or login behaviour. Convolutional architectures can identify local patterns in transformed sequences, while transformers are increasingly explored for long-range dependencies and multimodal context.

Deep models are attractive when large volumes of sequential or unstructured data are available, but they introduce substantial requirements for computation, tuning, monitoring and explanation. Their performance can deteriorate when training data are unrepresentative or when fraud patterns change. For regulated banking decisions, predictive strength must therefore be accompanied by documentation, stress testing and interpretable reason codes.

5.5 Natural Language Processing

Natural language processing can analyse customer complaints, loan applications, transaction narratives, emails, investigator notes and know-your-customer documents. It may identify inconsistent statements, suspicious entities or recurring communication patterns. Text evidence is strongest when combined with transaction, device and network information rather than used in isolation.

5.6 Hybrid Systems

Many fraud schemes are relational. Accounts may share devices, addresses, beneficiaries, merchants or contact details, and funds may circulate through connected entities. Graph analytics represents customers, accounts, devices and transactions as nodes and relationships as edges. Graph-based techniques are particularly promising for detecting fraud rings, mule-account networks, circular transfers, merchant collusion and layering associated with money laundering.

A hybrid system combines regulatory rules, supervised classifiers, anomaly detectors, graph scores, behavioural signals and investigator feedback. This layered arrangement is more practical than exclusive reliance on one model because the components detect different forms of risk. The challenge is to calibrate and combine scores without duplicating alerts or producing conflicting decisions.

Table 1. Comparative Assessment of Major Techniques

Technique	Primary Contribution	Strength	Main Limitation
Logistic regression	Probability-based classification	Interpretability	Limited non-linearity
Random Forest	Tabular transaction classification	Robustness	Reduced transparency
Gradient boosting	High-performance structured modeling	Predictive power	Tuning complexity
Isolation Forest	Unknown-pattern detection	Few labels required	False alerts
Autoencoder	Reconstruction-based anomaly detection	Learns normality	Threshold selection
LSTM/Transformer	Sequential behaviour modelling	Temporal context	Data and compute needs

Graph analytics	Network and fraud-ring detection	Relational insight	Graph construction
Hybrid system	Combined detection layers	Operational resilience	Integration governance

The literature on intelligent financial fraud detection consistently shows that effectiveness depends on matching the model to the data structure and fraud mechanism.8

6. Class Imbalance and Model Evaluation

6.1 The Imbalance Problem

Fraud normally constitutes a very small share of total transactions. Under such conditions, accuracy can conceal failure. For example, a classifier that labels 99.8 per cent of transactions as legitimate may appear successful even when it detects almost no fraud. Researchers therefore use data-level methods such as undersampling, oversampling and the Synthetic Minority Over-sampling Technique (SMOTE), together with algorithm-level methods such as class weighting, cost-sensitive learning, focal loss and threshold adjustment.

SMOTE creates synthetic minority examples by interpolating between neighbouring minority observations, but it must be applied only within the training data to avoid leakage.9

6.2 Essential Performance Measures

Precision measures the proportion of flagged transactions that are genuinely fraudulent. Recall measures the proportion of actual fraud that is detected. The F1-score balances precision and recall, while the area under the precision–recall curve summarises performance across thresholds and is especially informative for rare-event classification. Other necessary indicators include false-positive rate, false-negative rate, Matthews correlation coefficient, detection latency, alert volume and financial value saved.

Precision–recall analysis is generally more informative than ROC analysis when the positive class is rare and the practical interest centres on successful detection of that class.10

6.3 Validation Design

Random train–test splitting can place similar or temporally adjacent transactions in both sets, inflating performance. Fraud systems should preferably train on earlier periods and test on later periods. Cross-validation must preserve class structure and avoid leakage from duplicate accounts, synthetic samples or post-investigation variables. A credible study should also compare against meaningful baselines and report confidence intervals or repeated-run variation where possible.

7. Major Findings of the Review

First, no algorithm is universally superior. Ensemble trees are often strong on tabular transactions, sequential networks are suitable for behavioural histories, and graph methods offer particular value for coordinated fraud. Performance varies with fraud type, feature quality, class distribution, label reliability and validation design.

Second, data quality is more important than algorithmic complexity. Incomplete customer histories, delayed labels, inconsistent time stamps and poorly defined fraud outcomes can undermine even advanced models. Carefully engineered behavioural and relational features frequently produce greater gains than replacing one complex algorithm with another.

Third, banking fraud detection is inherently cost-sensitive. Missing a high-value fraud and incorrectly blocking a routine genuine transaction have different consequences. Thresholds should reflect transaction value, customer profile, investigation capacity and institutional risk tolerance rather than a generic statistical cut-off.

Fourth, explainability is essential for adoption. Investigators and compliance officers require reasons such as a new device, unusual location, sudden amount increase, rapid beneficiary creation or links to suspicious accounts. Explanations support human review, customer communication and model governance.

Fifth, real-time suitability must be measured. A model that produces excellent offline results may be unusable for instant payments if inference is slow or infrastructure requirements are excessive. Throughput, latency, failover behaviour and peak-load performance should accompany conventional statistical metrics.

Finally, AI does not eliminate the need for human expertise. Investigators interpret context, confirm evidence, recognise emerging schemes and provide feedback for model improvement. The most reliable arrangement is collaborative: automated risk scoring narrows attention, while accountable human decision-makers handle ambiguous or high-impact cases.

8. Challenges in AI-Based Banking Fraud Detection

8.1 Concept Drift and Delayed Labels

Fraud patterns change as criminals adapt to controls, payment products and customer behaviour. Models trained on historical data may therefore decay. Labels are also delayed because fraud is confirmed after complaints, chargebacks or investigations. Continuous monitoring, periodic retraining and drift-sensitive evaluation are necessary.

8.2 Limited and Non-Representative Data

Banks rarely release transaction data because of confidentiality, security and competitive concerns. Research consequently depends on a small number of anonymized datasets that may not represent current mobile payments, instant transfers, loan fraud or regional behavior. This limits generalizability and cross-institutional comparison.

8.3 Data Leakage and Reproducibility

Leakage occurs when the model uses information unavailable at the time of decision, when oversampling precedes data splitting, or when duplicates cross validation boundaries. Incomplete reporting of preprocessing, features and hyperparameters further restricts reproducibility. Studies should publish pipelines,

temporal cut-offs and ablation analyses wherever confidentiality permits.

8.4 Privacy, Bias and Security

Fraud systems may process sensitive location, device and behavioural data. Collection should be proportionate, access controlled and auditable. Bias can arise when historical data or proxy variables produce unequal error rates across customer groups or regions. Models must also be tested against deliberate evasion, adversarial manipulation and attempts to infer detection thresholds.

8.5 Operational Burden

Excessive false positives increase investigator workload and inconvenience customers. Successful deployment therefore requires alert prioritization, case-management integration, reason codes, escalation procedures and feedback mechanisms. Model governance is not an administrative add-on; it is part of detection quality.

9. Explainable and Responsible AI

Explainable AI makes model decisions understandable to investigators, auditors and affected customers. Global techniques describe overall feature importance, while local techniques explain an individual transaction. Common methods include feature importance, SHAP values, local surrogate models and rule extraction. An explanation should be faithful enough to support action rather than merely produce a plausible narrative.

Responsible AI extends beyond explanation. It requires clear accountability, human oversight, data-quality controls, privacy protection, fairness testing, independent validation, incident management, model documentation and continuous monitoring. A model should not be considered successful solely because it obtains a high F1-score; it must also be secure, stable, auditable and appropriate for its operational context.

10. Proposed Integrated Fraud-Detection Framework

The review supports a nine-layer framework. Layer 1 acquires data from core banking, card platforms, mobile and internet logs, customer profiles, devices, merchants, beneficiaries and confirmed investigations. Layer 2 cleans, encodes, standardises and deduplicates data. Layer 3 creates behavioural and relational features such as transaction velocity, amount deviation, device novelty, geographic distance, failed-login frequency, account age and network centrality.

Layer 4 applies multiple detectors: business rules, gradient-boosting classification, autoencoder anomaly detection, sequence modelling and graph scoring. Layer 5 fuses these outputs into a calibrated risk value. Layer 6 manages decisions—approve, monitor, authenticate, hold, decline or investigate. Layer 7 generates reason codes. Layer 8 provides human review. Layer 9 feeds confirmed outcomes back into model monitoring and retraining.

This architecture recognises that algorithm selection is only one element of effective fraud control. Data quality, secure engineering, regulatory rules, customer authentication, investigation capacity and feedback are equally important.

Figure 1. Real-Time AI-Based Banking Fraud-Detection Dashboard in a Banking Environment
Source: Author-generated realistic illustration based on the proposed integrated fraud-detection framework.

11. Research Gaps and Future Directions

The literature remains overly dependent on a few public credit-card datasets. Future research should include mobile banking, real-time payment systems, loan fraud, account takeover and multilingual social-engineering data. Studies must also report production-level latency, alert volume, investigator workload and actual financial value protected.

Temporal evaluation is still insufficient. Out-of-time testing, continual learning and explicit drift analysis should become standard. Graph neural networks deserve greater attention because transaction-level classification alone cannot adequately represent mule-account networks and coordinated groups. Cross-bank and cross-country testing is also required to determine whether models generalise beyond a single institution.

Federated learning and privacy-preserving analytics may enable collaboration without direct sharing of sensitive customer records. Cost-sensitive optimisation should replace accuracy-centred model selection. Multimodal systems should integrate transaction histories, text, device signals, behavioural biometrics and graph relationships. Human–AI studies should examine how explanations affect investigator decisions, trust and error correction.

Emerging generative-AI risks also require attention. Deepfake impersonation, automated phishing and synthetic identities can increase the speed and credibility of attacks. Detection research must therefore consider AI as both a defensive capability and a tool available to fraudsters.

12. Managerial and Policy Implications

Banks should select fraud-detection technology using recall, false-positive cost, latency, interpretability, scalability and integration requirements rather than advertised accuracy. Formal governance should involve fraud specialists, data scientists, cybersecurity teams, compliance officers, legal experts and senior management. Independent validation should precede deployment, and post-deployment monitoring should examine drift, fairness, security and operational impact.

13. Limitations of the Review

The review is subject to several limitations. Banking-fraud terminology varies across disciplines, and some relevant publications may use alternative expressions. Access restrictions can limit examination of commercial or proprietary studies. Heterogeneous datasets and metrics prevent direct statistical pooling, while publication bias may favour positive model results.

Finally, fraud strategies and AI technologies evolve rapidly, so the synthesis requires periodic updating. The paper should therefore be read as a systematic qualitative review rather than a statistical meta-analysis.

14. Conclusion

Artificial intelligence has expanded the capacity of banks to analyse large, complex and rapidly changing transaction environments. Random Forest and gradient boosting are often effective for structured data; anomaly models are useful where labels are scarce; sequence models capture behavioural evolution; and graph analytics offers a promising route to identifying coordinated networks. Yet high accuracy alone does not establish practical effectiveness.

Reliable evaluation must address severe class imbalance, temporal change, delayed labels, data leakage and non-representative datasets. Precision, recall, F1-score, precision–recall area, false-positive cost, latency and financial impact provide a more credible basis for comparison. Explainability, privacy, fairness, security and auditability are equally important because fraud detection affects customer access, institutional reputation and regulatory accountability.

The strongest future direction is a hybrid, adaptive and human-supervised system that combines business rules, machine learning, anomaly detection, graph intelligence and expert investigation. Such a system should learn continuously, provide clear reason codes, withstand adversarial behaviour and operate within a formal governance framework. AI should therefore be treated not as a replacement for banking judgement, but as a disciplined decision-support capability that strengthens prevention while preserving fairness and trust.

References

1. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90–113. https://doi.org/10.1016/j.jnca.2016.04.007
2. Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, 134–142. https://doi.org/10.1016/j.eswa.2015.12.030
3. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255. https://doi.org/10.1214/ss/1042727940
4. Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324
5. Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences, 557, 317–331. https://doi.org/10.1016/j.ins.2019.05.042
6. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), Article 15. https://doi.org/10.1145/1541880.1541882
7. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953
8. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
9. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 159–166). IEEE. https://doi.org/10.1109/SSCI.2015.33
10. Fawcett, T., & Provost, F. (1997). Adaptive fraud detection. Data Mining and Knowledge Discovery, 1, 291–316. https://doi.org/10.1023/A:1009700419189
11. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), Article 44. https://doi.org/10.1145/2523813
12. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239
13. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037
14. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413–422). IEEE. https://doi.org/10.1109/ICDM.2008.17
15. Lucas, Y., & Jurgovsky, J. (2020). Credit card fraud detection using machine learning: A survey. arXiv. https://doi.org/10.48550/arXiv.2010.06479
16. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). Curran Associates.
17. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1
18. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data-mining techniques in financial fraud detection: A classification framework and an academic review. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006
19. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

20. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv. https://doi.org/10.48550/arXiv.1009.6119
21. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778
22. Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432
23. Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems, 75, 38–48. https://doi.org/10.1016/j.dss.2015.04.013
24. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47–66. https://doi.org/10.1016/j.cose.2015.09.005
25. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. Data Mining and Knowledge Discovery, 18, 30–55. https://doi.org/10.1007/s10618-008-0116-z