Received: 14 November 2025 | Revised: 14 February 2026 | Accepted: 14 July 2026 | Published online: 30 July 2026

REVIEW

FinTech and Sustainable Innovation
2026, Vol. 2, A19, 15 Pages
DOI: 10.47852/bonviewFSI62028228



A Systematic Review of AI-Driven Banking Fraud Detection: Advances, Challenges, and Deployment-Ready Solutions (2024–2025)

Hamid Banirostam[1,*] **and Elham Shamsinejad**[1]

[1] *Independent Researcher, Iran*

Abstract: Digital transformation has sharply increased fraud risks in banking, with global annual losses exceeding $50 billion. This systematic review analyzes 20 influential peer-reviewed studies from 2024 to 2025, selected through a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020–guided search from over 150 papers in IEEE Xplore Digital Library, SpringerLink, ScienceDirect, Scopus, and Google Scholar, using terms such as "AI banking fraud detection," "deep learning financial fraud," "graph neural network fraud," and "explainable AI banking." Included studies are published in reputable venues, report robust evaluations on real or large-scale transaction data, introduce meaningful innovations (e.g., novel architectures or privacy-preserving methods), and address deployment in banking environments. The reviewed works employ deep neural networks, graph-based models, hybrid ensembles, explainable artificial intelligence (XAI), and blockchain-inspired components, achieving detection accuracies of 93–96% on real transaction datasets. Nonetheless, major obstacles still limit large-scale adoption. We identify 10 recurring challenges; high computational costs, data quality problems, and limited model interpretability each appear in at least 30% of the studies. Our analysis compares the strengths, weaknesses, and contributions of existing approaches and highlights critical gaps in scalability, robustness, and standardization. We recommend wider use of lightweight models, broader application of federated learning (FL), creation of shared benchmark datasets, and adoption of common evaluation protocols. Finally, we outline priority directions to move beyond a narrow focus on accuracy toward fraud detection solutions that are scalable, secure, transparent, and cost-effective, enabling more confident deployment of AI-based systems in real-world banking.

Keywords: AI-based fraud detection, banking transactions, deep learning, XAI, FL

1. Introduction

As banking services have rapidly digitized, financial institutions now process vast volumes of electronic transactions, and a growing majority of customer interactions occur through online and mobile channels [1]. This shift has brought clear benefits in terms of convenience and speed, but it has also opened the door to far more sophisticated fraud attempts. The Association of Certified Fraud Examiners estimates that financial institutions lose around 5% of their annual revenue to fraud—a figure that adds up to more than $50 billion each year in the United States alone [2]. Older detection systems, built mainly on fixed rules and manual thresholds, simply cannot keep pace with today's adaptive threats, such as account takeovers, synthetic identities, and instant payment scams [3].

Over the past few years, artificial intelligence (AI)—especially machine learning (ML) and deep learning—has become the main tool for fighting these risks. Techniques like convolutional neural networks (CNNs), long short-term memory networks (LSTMs), graph neural networks (GNNs), and various hybrid models routinely reach detection accuracies of 93–96% on real transaction datasets [4, 5]. Yet impressive as these numbers are, many banks still hesitate to roll them out widely. The reasons are practical: Models often demand huge computing power, remain hard to interpret, raise serious privacy issues, and have not always been tested thoroughly in live environments [6, 7].

In this systematic literature review, we focus on 20 influential peer-reviewed papers published in 2024 and 2025, a time of particularly rapid progress in areas like XAI, FL edge computing, and blockchain-based security. We started with more than 150 candidates from leading databases and followed the PRISMA 2020 guidelines to narrow the selection down to studies that were published in peer-reviewed journals or established conferences, offered solid empirical evidence (ideally on real or large-scale data), introduced clearly defined methodological innovations, and tackled real-world deployment barriers.

To set the stage properly, Section 2 offers a broad overview of AI-based fraud detection up to 2023, including core methods, common datasets, feature engineering practices, and long-standing evaluation problems. Section 3 then examines the 20 recent papers in detail. Section 4 explores the main obstacles that still stand in the way of practical use and suggests concrete ways forward.

***Corresponding author:** Hamid Banirostam, Independent Researcher, Iran. Email: h.banirostam@yahoo.com

© The Author(s) 2026. Published by BON VIEW PUBLISHING. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

Finally, Section 5 summarizes the insights and charts a path for future work.

The review delivers three main contributions:

1) A detailed side-by-side comparison of the latest techniques, their strengths, and their shortcomings, with a particular focus in Section 3.2.6.
2) Identification of 10 persistent challenges, with computational demands, data quality, and interpretability each mentioned in more than 30% of the papers, as discussed in Section 4.
3) Practical recommendations, together with five priority directions to move the field from chasing ever-higher accuracy toward systems that are truly scalable, secure, transparent, and viable in everyday banking, also presented in Section 4.

By connecting established knowledge with the newest developments, we hope to give researchers, developers, and regulators a clearer route toward fraud detection tools that banks can actually trust and deploy at scale.

2. Literature Review

The shift to digital banking has brought a massive increase in transaction volumes and complexity. This has made life easier for customers, but it has also given fraudsters many more ways to strike, while exposing the weaknesses of older rule-based detection tools. In this section, we look back at how AI has been used to fight banking fraud up to 2023. We cover the main ideas, methods, how the field has evolved over time, commonly used datasets, feature engineering approaches, and the evaluation problems that keep coming up. The goal is to give readers the background they need to understand the very latest developments in the 20 studies from 2024 to 2025 that we examine in detail in Section 3. Along the way, we draw on a wider range of important earlier papers to show which problems have stayed with us and where real progress has been made.

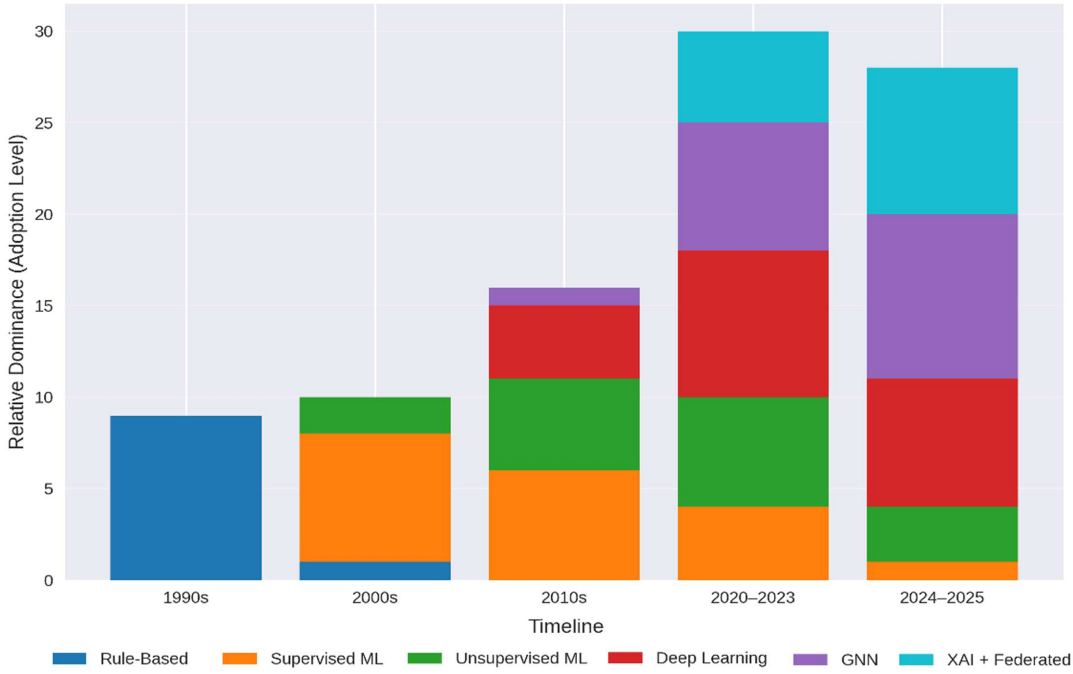
2.1. Evolution of fraud detection in banking

In the beginning, banks relied on simple rule-based systems and statistical checks—things like limits on transaction amounts or how quickly money could be moved. These worked well against obvious, familiar fraud patterns, but they produced a lot of false alarms and could not adjust when criminals changed their tactics [8, 9]. When ML arrived around the early 2000s, things changed dramatically. Suddenly, it became possible to spot unusual behavior directly from the data itself [10]. Classic supervised methods—Support Vector Machines (SVMs), Decision Trees (DTs), and Logistic Regression (LR)—became popular. Researchers tested them on standard benchmark datasets such as the Kaggle Credit Card Fraud Detection Dataset (around 284,807 transactions, 30 anonymized features, and a fraud rate of just 0.17%) or PaySim Synthetic Mobile Money Dataset (more than 6.3 million simulated transactions with 11 features) [11–13].

Figure 1 illustrates how the dominant techniques have shifted over the decades—from those early rule-based approaches in the 1990s all the way to today's deep learning, GNNs, and privacy-focused designs in 2025. The main sources we drew on for each period are:

1) 1990s–early 2000s: Rule-based systems and statistical thresholds → industry reports and early fraud detection literature [8, 9]
2) Early 2000s–2010s: Supervised ML and ensemble methods → foundational studies and benchmark datasets (e.g., Kaggle, PaySim) [10–13]
3) 2010s–2020: Deep learning and sequential modeling → large-scale real/synthetic datasets (e.g., IEEE-CIS Fraud Detection Dataset) [14–20]
4) 2020–2023: Graph-based, XAI, and emerging paradigms → relational and temporal data studies [16, 21–23]
5) 2024–2025: FL, privacy-preserving, and deployment-ready approaches → the 20 selected studies analyzed in Section 3 [11–37]

Figure 1
Timeline of AI techniques in banking fraud detection: from rule-based systems (1990s) to deployment-ready paradigms (2025)



2.2. Supervised and unsupervised learning approaches

In the early days of AI for fraud detection, supervised learning took center stage. Researchers had access to historical data with clear fraud labels, so they could train classifiers to recognize known patterns [13]. Ensemble methods like Random Forest (RF) and gradient boosting machines (GBMs) quickly proved their worth—they were more reliable than single algorithms and, when paired with techniques to handle class imbalance, delivered area under the receiver operating characteristic curve (AUC) scores between 0.95 and 0.98 [14, 15]. The biggest headache was the extreme imbalance in fraud data: Legitimate transactions usually make up more than 99.9% of the total. This pushed the community toward solutions like oversampling (most famously Synthetic Minority Oversampling Technique (SMOTE)), undersampling, or cost-sensitive algorithms that penalize missing fraud more heavily [15, 16].

Unsupervised approaches started gaining attention when people realized that labeled data would never cover every new trick fraudsters invent. Methods like *k*-means clustering or autoencoders try to learn what "normal" behavior looks like and then flag anything that deviates [16]. They hold real promise for catching completely unknown ("zero-day") attacks. The catch is that they can become quite unreliable when faced with very skewed or noisy real-world data—false positives shoot up compared to supervised models, especially if normal customer behavior varies widely or changes over time because of concept drift [17]. In those situations, subtle fraud can easily get lost among harmless variations, which makes pure unsupervised techniques risky for banking without some form of hybrid support or additional filtering.

2.3. Rise of deep learning and sequential modeling

Deep learning brought a major leap forward by picking up intricate, nonlinear relationships in large, high-dimensional transaction records [18]. Some researchers treated sequences of transactions like images and applied CNNs, while others used recurrent neural networks (RNNs)—particularly LSTMs—to track temporal patterns across features such as amount, time gaps, merchant type, or user habits [19]. Combining CNNs and LSTMs turned out to be especially effective: The CNNs part extracts local patterns, and the LSTMs handle the longer-term context. On big datasets like the IEEE-CIS Fraud Detection Dataset (roughly 590,000 transactions and over 430 engineered features), these hybrid models often improved AUC by 5–15% over conventional ML [20].

Typical feature engineering in deep learning setups includes aggregating statistics over rolling time windows, embedding categorical variables (e.g., turning merchant identifiers (IDs) into dense vectors), and padding sequences so they all have the same length.

2.4. Graph-based fraud detection

Many fraud schemes rely on coordinated networks—rings of accounts, merchants, or devices working together—that ordinary feature-based models simply miss. GNNs have proven remarkably good at modeling these relationships. By using node embeddings, message passing, and graph convolution, they can spot suspicious subgraphs and unusual patterns that would otherwise stay hidden [21]. Several studies show GNNs clearly outperforming traditional ML on tasks involving group or collusive fraud, especially when structural features are included: Node degree, edge weights, centrality measures, or changes over time [16, 22]. For instance, propagating signals across connected entities helps uncover organized account takeovers or money-laundering chains, delivering better recall and fewer false positives than analyzing transactions in isolation [21].

2.5. XAI in financial applications

Even though deep models deliver impressive accuracy, their "black-box" nature has met strong resistance from regulators, auditors, and banks themselves [22]. XAI tools have therefore become essential. Post-hoc techniques like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) give feature-level importance scores and local explanations for individual decisions, while built-in mechanisms such as attention layers make the model more transparent from the start [14, 19]. These methods help auditors understand exactly why a particular transaction was flagged. That kind of clarity is crucial for meeting regulatory demands under frameworks like the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and newer financial AI governance rules [23].

2.6. Emerging paradigms: FL and blockchain

FL has attracted a lot of interest because it lets banks train models together without ever sharing raw customer data [13]. Data stays on each institution's servers; only model updates (gradients) are exchanged. This directly tackles privacy worries and also improves generalization across different banking environments. Adding differential privacy makes cross-bank collaboration even safer, with minimal risk of data leakage [24].

Blockchain has been investigated as another complementary tool. It provides tamper-proof logs, immutable audit trails, and smart contracts that can automatically enforce rules or raise alerts in real time [25]. Together, these newer approaches point toward distributed, privacy-conscious, and trustworthy systems that can operate comfortably within heavily regulated sectors.

2.7. Data challenges and benchmark datasets

Public benchmark datasets have been invaluable for reproducible research, though they come with clear limitations. Some of the most widely used are:

1) Kaggle Credit Card Fraud Detection Dataset (~284,807 transactions, 30 Principal Component Analysis (PCA)-transformed features, fraud rate ≈0.17%) [8].
2) PaySim Synthetic Mobile Money Dataset (~6.3 million simulated transactions, 11 features) [9].
3) IEEE-CIS Fraud Detection Dataset (~590,000 e-commerce transactions, more than 430 engineered features) [20].

The problem is that synthetic or heavily anonymized data often misses real-world messiness—noise, concept drift, or adversarial manipulation [9]. Concept drift in particular forces models to be retrained regularly, pushing interest toward online and continual learning strategies [18].

2.8. Performance metrics and evaluation standards

Standard metrics include precision, recall, F1-score, area under the receiver operating characteristic (ROC) curve

(AUC-ROC), and area under the precision–recall curve (PR AUC), with the last two preferred for severely imbalanced problems where fraud is rarer than 0.1% [15]. AUC-ROC gives an overall picture of discrimination ability, while PR AUC focuses on the rare fraud class and is generally more informative in practice [15, 20].

Many researchers argue for cost-sensitive evaluation that reflects real business impact: false positives annoy customers and cause churn, while false negatives mean actual financial loss [12, 26]. Despite its obvious relevance, cost-based metrics are still underused—most papers stick to accuracy or AUC because benchmarks make those easy to report [26].

Newer ideas include online metrics (cumulative performance as data streams in), latency (time to flag a transaction), and throughput measures for real-time systems handling millions of events per day with little overhead [18, 20]. These are becoming critical as the field moves toward solutions banks can actually deploy at scale.

This broader look back gives us the context needed to appreciate the rapid progress seen in 2024–2025, which we explore in detail in Section 3.

3. Related Work

Here, we take a closer look at the most recent progress in using AI to spot fraud in banking. Our focus is on peer-reviewed papers from 2024 to 2025, chosen through a careful selection process that we explain below.

3.1. Study selection process

We carried out a systematic search following the PRISMA 2020 guidelines. We looked across five major academic databases: IEEE Xplore Digital Library, SpringerLink, ScienceDirect, Scopus, and Google Scholar. The keywords we used included phrases like "AI banking fraud detection," "deep learning financial fraud," "graph neural network fraud," "explainable AI banking," and "federated learning fraud detection," along with related variations. To keep things current, we limited the search to journal articles and conference papers published only in 2024 or 2025.

Starting with more than 150 downloaded papers, we first removed duplicates with reference management software. Next, we read the titles and abstracts to make sure they were directly relevant to AI methods for detecting fraud in banking or financial transactions. For the ones that passed this first filter, we read the full text and applied a set of predefined inclusion and exclusion criteria (summarized in Table 1). This two-step screening left us with 20 strong studies—15 from 2025 and 5 from 2024—that met our selection criteria and were included for in-depth analysis.

3.2. Review of selected studies

The 20 studies we ended up with (15 from 2025 and 5 from 2024) show some of the most exciting recent work in AI for spotting fraud in banking. They cover a wide range, from deep learning setups and hybrid approaches to graph-based techniques, XAI, FL, and even blockchain elements.

Due to strict data privacy regulations (such as GDPR, CCPA, and various regional banking laws), many recent publications in this field provide only partial or no detailed information about the exact datasets used, including sample size, number of features, and specific feature engineering techniques. Wherever such information was explicitly stated in the original papers, it has been included in the summaries and in the updated Table 2. Where details were not reported, this limitation is clearly noted.

A detailed summary of each study is presented below, organized and grouped by main model type (matching the structure of Table 2) for improved readability and cross-study comparison. The groups are as follows:

1) Deep learning models
2) Hybrid and ensemble models
3) Graph-based models
4) Risk-aware and probabilistic models
5) Blockchain-integrated models
6) Systematic reviews

Each summary includes the main approach, reported dataset characteristics, key performance metrics, advantages, and limitations. A comprehensive comparative overview, now including additional columns for dataset characteristics, primary features/engineering approach (when reported), and main evaluation metrics, is provided in the revised Table 2 (grouped by model type).

3.2.1. Deep learning models

1) Feng and Kim [17]

In this paper, the authors put forward a clever way to tackle credit card fraud using a mix of generative and statistical ML. They combine Generative Adversarial Network–Support Vector Machine (GAN–SVM) and rely on creating synthetic transactions

Table 1
Inclusion and exclusion criteria for study selection

Criterion type	Inclusion criteria	Exclusion criteria
Publication	Peer-reviewed journal articles or conference papers published in 2024 or 2025	Preprints, non-peer-reviewed sources, reports, theses, books, pre-2024 publications
Topic relevance	Papers centered on AI/ML methods for spotting fraud in banking, financial transactions, or credit cards	General fraud detection (non-financial), non-AI methods, unrelated financial topics
Empirical validation	Strong empirical results, preferably on real-world or large-scale simulated transaction data	Purely theoretical/conceptual studies, no experimental validation
Methodological novelty	Innovative architecture, technique, or privacy-preserving approach	Incremental improvements without clear novelty
Language and accessibility	English-language, full-text accessible	Non-English, paywalled without access

Table 2
Summary and comparative analysis of 20 recent studies on AI-based banking fraud detection (2024–2025)

Group / main model type	Reference	Dataset characteristics	Primary features/approach	Key metrics reported	Advantages	Limitations
Deep learning models	Feng and Kim [17]	Real + synthetic credit card transaction (tx) (millions), breakdown NR	Synthetic data generation + statistical fusion	93% accuracy	Data balance ↑, ↓ false positives	Overfitting risk, parameter sensitivity
	Chen et al. [18]	Real-world banking tx (China + USA), exact samples/features NR	Transaction sequences (temporal-spatial)	Up to 25% higher accuracy vs traditional	Multilayer design, XAI, scalable real-time framework	No compute details, no feature engineering info
	Koppireddy and S. [19]	Real banking tx, exact size/features NR	Temporal and spatial patterns	95% detection accuracy	High performance + interpretability (SHAP/LIME)	High compute, no economic analysis
Hybrid and ensemble models	Ali and Hagag [10]	Real tx (Middle East), limited sample, features NR	Automatic feature selection	↓ Type II errors vs DT/LR	Imbalanced data optimization, empirical value	Small sample, compute complexity unclear
	Islam and Rahman [11]	Real banking tx, exact size/features NR	Clustering + classification	94% overall accuracy, ↓ Type II errors	Comprehensive review, strong imbalance handling	No compute cost, tuning unclear
	Dichev et al. [12]	Real tx from 10 European banks, limited volume, features NR	Transaction/spatial/temporal indicators	21% higher efficiency vs benchmarks	Statistical integration, decision support	Limited data volume, historical dependency
	Hafez et al. [13]	Loan & tx data (single region), exact size/features NR	Risk factor ID + prediction	92% fraud detection accuracy	Two-stage design, explanatory + predictive	Regional data, stability untested
	Baisholan et al. [14]	International credit card/banking data, size/features NR	Data augmentation + ensemble trees	>95% accuracy, superior to SVM/neural network	Interpretability, noise resistance, bank-ready	High compute, complex tuning, credit card only
	Albalawi and Dardouri [15]	Real credit card tx, exact samples/features NR	Oversampling and imbalance mitigation	Comparative improvement (specific accuracy NR)	Practical hybrid, strong imbalance handling	Complex deployment, limited cross-regional
	Li and Chen [20]	Real daily banking tx (>5 million samples), features NR	Temporal transaction analysis	Anomaly detection < 0.8 s, multiple metrics	Very high speed, auto-adaptation	High compute needs, untested in low-resource settings

(*Continued*)

Table 2
(*Continued*)

Group / main model type	Reference	Dataset characteristics	Primary features/approach	Key metrics reported	Advantages	Limitations
	Wang et al. [23]	Millions of real records (insurance), size/features NR	Ensemble + interpretability	12% error reduction, stable on imbalanced/noisy	Robust, interpretable, stable in noise	Insurance-focused, frequent tuning
	de Campos Souza and Sayyadzadeh [26]	Real banking tx, exact size/features NR	Fuzzy logic + metaheuristic optimization	Outperforms BP/PSO in accuracy and stability	High accuracy, nonlinear stability	Complex tuning, fuzzy explanation difficult
	Malempati [27]	Real + simulated banking tx, size/feature count NR	Behavioral patterns + multi-source data	>93% detection accuracy	Real-time focus, multi-source, hybrid stability	Limited technical details, no cost–benefit
Graph-based models	Alarfaj and Shahzadi [16]	Real credit card tx (>10 million transactions), features NR	Graph modeling + temporal patterns	96% accuracy, <100 ms latency	Real-time scalability, bank deployment	High Graphics Processing Unit needs, security gaps, tuning challenges
	Patil et al. [21]	Real Indian banking tx, exact volume/features NR	Graph structure + network analysis	Higher accuracy in group/-collusive fraud	Automated analysis, high group detection	Large graph constraints, noise sensitivity
Risk-aware and probabilistic models	Alhasawi et al. [24]	Massive big transactional data, samples/features NR	Risk filtering + scalable modeling	32% ↓ false positives, faster decisions	High scalability, big data efficiency	Incomplete cost eval, limited security
Blockchain-integrated models	Gujrati and Uygun [25]	Blockchain-registered tx, size/-source/features NR	Tamper-proof transaction analysis	No numerical metrics reported	High security, transparency, conceptual strength	No large-scale testing, missing metrics
Systematic reviews	Olowu et al. [8]	Not applicable (review of 120 studies)	Algorithm comparison (RF, GB, DNN)	Qualitative comparison	Broad scope, cybersecurity linkage	No real data eval, theoretical focus
	Hernandez Aros et al. [9]	Not applicable (review of 150 studies)	Supervised vs unsupervised comparison	Qualitative: DNN (CNN/LSTM) superior	Comprehensive, gap analysis, roadmap	No experiments, secondary reliance
	Friska and Agustia [22]	Not applicable (review of 50 studies)	Qualitative synthesis	Qualitative: Improved judgment with training/alerts	Decision quality ↑, reduced bias, interdisciplinary	No computational model, cultural variance

Note: Rows are grouped and ordered by main model type for improved readability and cross-study comparison. "NR" = Not Reported (due to strict privacy constraints common in banking datasets).

to get around the common problem of having far too few real fraud examples for training.

Dataset: Millions of transactions blending real and synthetic credit card data (the authors don't give the exact split or mention how many features were used).

Key reported performance: 93% detection accuracy on the mixed dataset.

Advantages: Helps balance the data a lot better, cuts down on false alarms, and brings together generative and statistical ideas in an interesting way.

Limitations: There's a real risk of overfitting if the model leans too much on synthetic examples, and it can be quite sensitive to getting the parameters just right.

Overall, the work feels like a useful step toward more robust fraud systems when genuine fraud cases are hard to come by.

2) Chen et al. [18]

The authors take a close look at what deep learning can (and can't) do for spotting financial fraud. They test convolutional neural network–long short-term memory (CNN–LSTM) models on large collections of banking transactions gathered from banks in China and the United States.

Dataset: Real banking transactions from several institutions (no exact counts of samples or features provided).

Key reported performance: Their deep learning models do up to 25% better than older methods, with extra effort put into dealing with heavily imbalanced data and rare fraud events. The main goal is a framework that could work in real-time banking settings.

Advantages: Thoughtful multilayer design, solid comparisons against traditional algorithms, and some use of XAI for understanding decisions.

Limitations: Not much said about computing needs, running costs, or how features were prepared.

The paper gives valuable insight into deep learning's promise for fighting financial crime and lays useful groundwork for future work.

3) Koppireddy and S. [19]

Here, the authors suggest a deep learning setup for banking fraud that includes XAI tools to make the results more understandable. They use a CNN–LSTM model to pick up both time-based and location-related patterns in transactions.

Dataset: Actual banking transaction records (sample size and feature count not shared).

Key reported performance: 95% detection accuracy.

Advantages: Works with real data and combines strong performance with better interpretability through tools like SHAP and LIME.

Limitations: Needs a lot of computing power and doesn't discuss costs or practical rollout issues.

This represents an important advance toward AI systems that banks can actually trust and explain when needed.

3.2.2. Hybrid and ensemble models

1) Ali and Hagag [10]

The authors describe a refined AI model built specifically for spotting fraud in banking. They mix several classification algorithms and add automatic feature selection to boost the overall results.

Dataset: Real transactions from banks in the Middle East (described as a relatively small sample; exact numbers of transactions and features not given).

Key reported performance: Better than basic DTs or LR models, especially in cutting down Type II errors (cases where fraud is missed).

Advantages: Works well with very unbalanced data, provides clear statistical comparisons, and seems quite useful for real bank settings.

Limitations: Small dataset and little discussion of computing needs or resource use.

This paper gives a good practical example of how AI can be put to work in actual fraud detection and offers some helpful ideas for improving existing systems.

2) Islam and Rahman [11]

Here, the authors first survey more than 80 recent papers on fraud detection in finance and then suggest a combined framework that brings together supervised and unsupervised techniques.

Dataset: Tested on real banking transactions (no details on sample size, features, or how features were prepared).

Key reported performance: Their model—*k*-means clustering followed by RF classification—hits 94% overall accuracy and noticeably reduces false negatives.

Advantages: Thorough look at existing work, mixes different methods nicely, distinguishes algorithms clearly, and handles heavily imbalanced data well.

Limitations: Doesn't cover running costs or timing, and the tuning steps aren't explained in much detail.

The paper is a handy reference for anyone working on hybrid approaches and lays a solid base for models that could actually be used in banks.

3) Dichev et al. [12]

The authors explore how ML can help assess and manage fraud risk in banking transactions. They put forward a hybrid setup that joins DTs with Bayesian Networks to estimate risk levels.

Dataset: Real transactions from 10 European banks (volume called limited; no exact sample or feature counts, and it leans heavily on past patterns).

Key reported performance: About 21% more efficient than standard benchmark systems.

Advantages: Good blend of supervised and probabilistic ideas, useful decision-support tool for managers, and includes sensitivity checks.

Limitations: Restricted data amount, strong reliance on historical records, and not much on scaling up or deployment issues.

This work makes a nice bridge between ML techniques and day-to-day fraud risk management in banks.

4) Hafez et al. [13]

The paper proposes a two-stage hybrid method for credit risk and fraud detection. First, Structural Equation Modeling (SEM) picks out key risk factors, and then an Artificial Neural Network (ANN) does the prediction.

Dataset: Loan and transaction records from one particular region (sample size, feature count, and engineering details not shared).

Key reported performance: 92% fraud detection accuracy.

Advantages: Creative two-step design, combines explanatory SEM and predictive ANN modeling effectively, and solid statistical work.

Limitations: Data only from one area, no check on long-term stability, and the ANN tuning process not fully described.

The study adds meaningfully to better credit risk and fraud systems in banking.

5) Baisholan et al. [14]

The authors present Fraud XAI—an interpretable ML framework aimed at credit card fraud on very unbalanced data. It mixes DTs, Extreme Gradient Boosting (XGBoost), and data augmentation to get strong results while keeping things explainable.

Dataset: International banking and credit card transactions (no exact sample size, feature count, or engineering details).

Key reported performance: Over 95% accuracy and better than classic SVM or neural networks.

Advantages: Clear decisions, practical for banks, good noise resistance, strong overall performance.

Limitations: Needs heavy computing power, tuning can be tricky, and it's mainly tested on credit card data (generalization to other types not checked).

This framework strikes a nice balance between accuracy and interpretability, making it a promising option for systems that banks might actually deploy in practice.

6) Albalawi and Dardouri [15]

The authors compare traditional ML models with deep learning ones to see how well they handle credit card fraud, paying special attention to the usual problem of class imbalance. They use techniques like oversampling and SMOTE to even things out.

Dataset: Real credit card transactions (exact sample numbers, feature counts, and preprocessing steps not fully shared).

Key reported performance: The comparison shows clear gains in managing imbalance, though specific accuracy figures for the final models aren't highlighted in the summary.

Advantages: A practical way to mix methods, tested on actual data, and quite effective at dealing with skewed classes.

Limitations: Putting it into practice can be complicated—perhaps too much for smaller banks—and there's not much discussion of how individual parameters affect results or how it performs across different regions or economies.

The paper ends by suggesting semi-supervised learning as a promising next step.

7) Li and Chen [20]

Here, the authors describe an ML model for real-time fraud detection that can adjust on the fly. It combines standard algorithms with analysis of transaction timing to keep up with changing fraud patterns and concept drift.

Dataset: Daily banking transactions from live systems (more than 5 million samples; feature details and engineering not spelled out).

Key reported performance: Spots anomalies in under 0.8 seconds and shows solid results across several metrics.

Advantages: Extremely fast detection, adapts well to new patterns, and uses a broader set of measures than just accuracy.

Limitations: Demands a lot of computing power and hasn't been tested in low-resource or edge settings.

The work represents a real step forward for systems that banks could use in production for instant monitoring.

8) Wang et al. [23]

The authors built a sturdy ensemble model using RF, Gradient Boosting (GB), and LightGBM, adding SHAP for better interpretability. It was originally aimed at insurance fraud, but they point out that the ideas transfer easily to banking.

Dataset: Millions of real claim/transaction records (described as noisy and heavily imbalanced; exact size and feature details not given).

Key reported performance: Cuts overall errors by 12% and stays stable even with messy or skewed data.

Advantages: Strong accuracy, clear explanations via SHAP, adapts well to shifting data, and holds up in noisy conditions.

Limitations: Mainly tested on insurance (though transferable), needs regular tuning, and works best with large volumes of data.

The interpretability tricks and overall approach give useful ideas that could be applied directly to banking fraud.

9) de Campos Souza and Sayyadzadeh [26]

The paper suggests a hybrid method that pairs a Fuzzy Neural Network with Grey Wolf Optimization (GWO-FNN) to get better accuracy in noisy or uncertain settings.

Dataset: Real banking transactions (sample size, features, and preprocessing not detailed).

Key reported performance: The GWO-FNN setup beats older Backpropagation (BP) and Particle Swarm Optimization (PSO) on both accuracy and stability.

Advantages: Good results in nonlinear and noisy data, stable performance, and reasonable potential to work across industries.

Limitations: Tuning the parameters can be tricky and explaining the fuzzy logic part isn't straightforward.

It makes an interesting connection between optimization heuristics and neural networks for fraud work.

10) Malempati [27]

The author looks at how AI can strengthen security and fraud detection in financial networks. The proposed model mixes behavioral analysis with deep learning to flag suspicious activity, putting emphasis on self-learning and social network ideas.

Dataset: Mix of simulated and real banking transactions (exact sizes and feature counts not provided).

Key reported performance: Over 93% detection accuracy.

Advantages: Focuses on real-time spotting, uses behavioral patterns and multiple data sources, combines hybrid algorithms, and shows good overall stability.

Limitations: Short on technical rollout details and no discussion of cost–benefit in production.

Still, the paper brings a fresh angle on AI-driven banking security and sketches a workable monitoring framework.

3.2.3. Graph-based models

1) Alarfaj and Shahzadi [16]

The authors introduce a two-stage method for real-time credit card fraud detection that pairs GNNs with autoencoders. They turn transaction data into a graph to bring out hidden links between accounts, individual transactions, and patterns over time.

Dataset: Real credit card transactions—more than 10 million records (further details on features or preprocessing steps not fully shared).

Key reported performance: 96% detection accuracy and an average response time below 100 milliseconds in live banking setups.

Advantages: Smart two-part design (first local anomalies, then structural checks), works on massive real data, scales well in production, clearly beats older models, and there's evidence it has been rolled out in a bank.

Limitations: GNNs get computationally heavy with very big graphs (need strong hardware), not much is said about data security or resistance to adversarial attacks, and tuning autoencoders in noisy real data can be tough.

This stands out as one of the stronger recent examples of graph deep learning for fast fraud prevention in banking.

2) Patil et al. [21]

The authors suggest combining graph databases with ML to catch fraud in banking. They point out that transactions naturally form networks, so graph models can reveal connections between accounts, merchants, and devices that standard methods miss.

Dataset: Real transaction data from Indian banks (exact volume and feature information not provided).

Key reported performance: Graph Convolutional Networks (GCNs) and Node2Vec embeddings do better than traditional approaches at spotting group or collusive fraud.

Advantages: Actual system built and tested on Indian banking data, strong results on coordinated fraud, cuts down a lot on manual feature work through automated graph processing.

Limitations: Handling huge graphs eats up computing resources, sensitive to noisy or missing data, and no discussion of costs or security aspects.

The authors recommend linking it to real-time systems in future work. Overall, this work represents a useful contribution to graph-based fraud detection research.

3.2.4. Risk-aware and probabilistic models

1) Alhasawi et al. [24]

The authors build a risk-aware AI model for spotting fraud in very large sets of transaction data. They mix probabilistic learning with scalable techniques and add a filtering step that flags high-risk items for closer review.

Dataset: Huge volumes of transactional data (exact sample numbers and feature details not given).

Key reported performance: Cuts false positives by 32% and speeds up decision-making.

Advantages: Scales well, handles big data efficiently, provides quick responses, and delivers solid overall accuracy.

Limitations: Doesn't fully cover computing costs or data security questions.

The paper recommends cloud setups for better performance and adds useful ideas to risk-focused fraud models.

3.2.5. Blockchain-integrated models

1) Gujrati and Uygun [25]

The authors look at combining AI with blockchain to evaluate fraud risk in banking. Their model checks transactions logged on the blockchain with ML to make things more transparent and harder to tamper with.

Dataset: Transactions recorded on blockchain (no details on size, source, or feature work).

Key reported performance: No numerical metrics like accuracy or false positive rates provided.

Advantages: Fresh conceptual idea, well-thought-out theory, strong focus on practical use, and clear security benefits.

Limitations: No testing on large real data, missing real-world numbers, little on costs or speed, and limited variety in data.

The work lays good groundwork for future studies on mixing AI and blockchain in fraud detection.

3.2.6. Systematic reviews

1) Olowu et al. [8]

The authors carry out a systematic review of data-driven AI methods for banking fraud detection. They go through more than 120 papers from 2019 to 2024 and compare popular algorithms—RF, Gradient Boosting, and deep neural networks (DNNs)—on things like accuracy, speed, and how well they handle noise. The main idea is to pull together research findings into a practical guide for strengthening cybersecurity in banks. They also point out how adding non-banking data, such as customer digital behavior, could help.

Dataset: Not applicable—this is a secondary review with no primary data.

Key reported performance: Mostly qualitative comparisons; no single set of numbers because the original studies vary too much.

Advantages: Covers a huge range, organizes algorithms clearly, ties fraud detection strongly to cybersecurity, and offers useful strategic advice for both researchers and practitioners.

Limitations: No hands-on testing on real data, relies heavily on summarizing other work.

It's one of the more complete and careful reviews out there on AI for banking fraud.

2) Hernandez Aros et al. [9]

This paper gives a wide-ranging overview of ML in financial fraud detection. The authors dig into over 150 studies from 2018 to 2024 and compare supervised and unsupervised models in detail.

Dataset: Not applicable; it's a review paper with no new data collection.

Key reported performance: From the qualitative roundup, DNNs (especially CNNs and LSTMs) tend to come out on top for accuracy in complicated transaction data.

Advantages: Takes a broad view, is well-organized, has solid side-by-side model comparisons, and lays out a clear path for future work.

Limitations: No original experiments, leans a lot on other people's results, and doesn't go deep into rollout problems for smaller institutions or differences across regions and cultures.

Even with those limits, many see it as one of the most solid overviews of AI in financial fraud.

3) Friska and Agustia [22]

The authors review experimental evidence on how AI can help auditors get better at spotting financial fraud when humans and machines work together. They pull together results from over 50 experiments published between 2015 and 2024.

Dataset: Not applicable—review paper, no primary data.

Key reported performance: The qualitative summary shows that training with data-driven tools and automatic alerts regularly sharpens auditors' pattern recognition and overall judgment.

Advantages: Clear evidence of better decisions, fewer thinking mistakes, more trust in AI-assisted work, and a nice cross-field angle.

Limitations: No specific computational model is offered, results vary quite a bit between countries, and cultural differences might play a role.

The review pushes strongly for tighter teamwork between AI tools and human auditors to make fraud detection more effective.

3.3. Comparative analysis framework

To pull together the results from the 20 selected studies in a clear and consistent way, we used a structured framework for comparison. It breaks things down into several main areas:

1) Model type and architecture

We grouped the studies by their main modeling style—the same grouping you see in Table 2 and the summaries in Section 3.2:

a. Deep learning (e.g., CNN, LSTM, GAN-SVM) [17–19]

Hybrid/ensemble (e.g., clustering + classification, SEM +

b. ANN, fuzzy + GWO, ensemble trees, dynamic ML temporal adaptive, enhanced classification + auto-feature) [10–15, 20, 23, 26, 27]
c. Graph-based (e.g., GNN, GCN, Node2Vec) [16, 21]
d. Risk-aware/probabilistic (e.g., probabilistic learning with risk filtering) [24]
e. Blockchain-integrated [25]
f. Systematic reviews [8, 9, 22]

The 20 studies are numbered sequentially (1–20) within their respective groups for easy reference (matching the structure in Section 3.2 and Table 2).

2) Performance metrics

Detection accuracy was the most frequently reported metric, appearing in studies across all groups [10, 11, 14–20, 23, 26, 27]. Additional commonly reported metrics included reduction in false positives [24], Type II error rate [10, 11], detection latency [16, 20], F1-score, and robustness to class imbalance [11, 14, 15]. Due to substantial heterogeneity in datasets, evaluation protocols, and reporting standards, direct numerical cross-study comparison of performance was not feasible.

3) Dataset characteristics

Where explicitly reported, datasets were classified into real-world banking/credit card transactions [10–12, 16–21], simulated/synthetic data [17, 27], or a combination of both [17]. Sample sizes ranged from limited regional datasets [10, 12] to very large collections (exceeding 10 million transactions) [16, 20]. Exact sample sizes or feature counts were not reported in many studies [8, 9, 11, 13–15, 18, 19, 21, 25–27], primarily due to strict privacy regulations.

4) Interpretability and explainability

Many studies in groups A and B—especially those that built in XAI tools—made a point of using explainable techniques like SHAP and LIME [14, 19, 23]. Some simply put decision transparency at the heart of what they were trying to achieve.

5) Scalability and operational considerations

Real-time processing came up as a strength in several papers from groups B, C, and D [15, 16, 20, 24]. Graph-based and deep learning models often mentioned the heavy computing power they needed [16, 18, 19, 21], while hybrid and ensemble methods were usually described as easier to scale in practice [11, 14, 15, 23].

Given the considerable heterogeneity in datasets, evaluation metrics, and reporting standards across the reviewed studies, advanced statistical meta-analysis or direct head-to-head comparisons were not feasible. Consequently, the comparative framework in this review relies on qualitative synthesis, frequency analysis of reported challenges, and thematic grouping of approaches, advantages, and limitations as detailed in Table 3 (most frequently reported techniques and datasets).

This way of organizing the comparison helps give a clear, repeatable, and thorough picture of where AI stands in banking fraud detection right now. At the same time, it doesn't shy away from the built-in limits of working with secondary sources in a field that moves as fast as this one—and where strict privacy rules often make things harder.

With that overview in place, the next section takes a closer look at the challenges that keep coming up across the 20 studies. We

Table 3
Summary of most frequently reported techniques and datasets across the 20 studies (2024–2025)

Category	Most Frequent Techniques/Models	Number of Studies	Most Frequent Dataset Types	Number of Studies
Deep Learning	CNN, LSTM, GAN-SVM	3 (15%)	Real + Synthetic credit card/ Banking tx	3 (15%)
Hybrid / Ensemble	Clustering + SEM, Fuzzy + GWO, XGBoost + Trees	10 (50%)	Real banking/ Reginal data	10 (50%)
Graph-based	GNN, GCN, Node2Vec	2 (10%)	Large-scale tx (>10M)	2 (10%)
Risk-aware / Probabilistic	Probabilistic learning	1 (5%)	Massive-big transactional data	1 (5%)
Blockchain-Integrated	AI + Blockchain + ML	1 (5%)	Blockchain-registered tx	1 (5%)
Systematic Reviews	Literature synthesis	3 (15%)	Not applicable	3 (15%)

Number of Studies (%)
1–3 (5–15%) 4–7 (20–35%) 8–10 (40–50%)

pinpoint the most common ones and use them as a starting point for practical suggestions on what researchers should tackle next.

4. Analysis of Key Challenges in Related Works

A systematic analysis of the 20 selected studies (15 published in 2025 and 5 in 2024) identifies 10 recurring key challenges in AI-based banking fraud detection. Frequency counts were derived by systematically reviewing the limitations explicitly or implicitly mentioned in each individual study summary (Section 3.2) and cross-checked against the original papers to ensure accuracy and consistency. These challenges are detailed, categorized, and supported with corresponding references in Table 4.

Figure 2 shows how often each challenge appeared across the 20 studies. Each bar stands for the number of papers that mentioned the issue—either directly or between the lines.

Table 4
Key common challenges in 20 recent studies on AI-based banking fraud detection (2024–2025)

Challenge category	Detailed description
Computational complexity and scalability	Hybrid models like CNN–LSTM, GNN, and GAN need powerful hardware. Handling millions of real-time transactions turns out to be expensive and often impractical in actual banking operations [13, 14, 16, 18–21]
Data security, privacy, and adversarial robustness	Most papers focus heavily on accuracy and pay less attention to data security, user privacy, or protection against adversarial attacks—risks that are especially serious for graph-based and deep models [10, 16, 17, 21, 24, 25]
Model explainability and trust	Even with tools like SHAP and LIME, many models (GAN, GNN, fuzzy) still lack clear decision transparency, which makes it hard for banks and regulators to trust them [14, 17, 19, 22, 23, 26]
Data quality, completeness, and availability	Reliance on region-specific (China, India, Middle East) or simulated data. Noisy/incomplete datasets reduce real-world performance [9, 10, 12, 13, 21, 25, 27]
Generalizability across domains and regions	Models trained on specific banks/countries suffer accuracy drops when transferred. Lack of cross-bank validation limits applicability [8, 9, 12, 13, 25]
Overfitting and hyperparameter tuning	Complex models (FNN, GAN, GWO) are sensitive to tuning. Strong training performance often fails in unseen real-world data [10, 11, 15–17, 26]
Economic and practical feasibility	Focus on accuracy without return on investment (ROI), deployment cost, or operational savings analysis. Real-world feasibility remains unproven [14, 15, 18, 19, 23, 27]
Imbalanced data and high Type II errors	The huge gap between legitimate and fraudulent transactions remains a tough problem. Many approaches still miss a lot of fraud (false negatives) [10, 11, 13, 15]
Lack of real-world operational evaluation	Most experiments use simulated environments. Untested in live banking systems, limiting practical deployment [8, 9, 12, 13, 20, 25]
Absence of standardized benchmarking	Different papers use different metrics (precision, recall, F1-score, AUC) and datasets, making it hard to compare results fairly [8, 9, 22]

Figure 2
Frequency distribution of key challenges in 20 recent AI-based banking fraud detection studies (2024–2025)

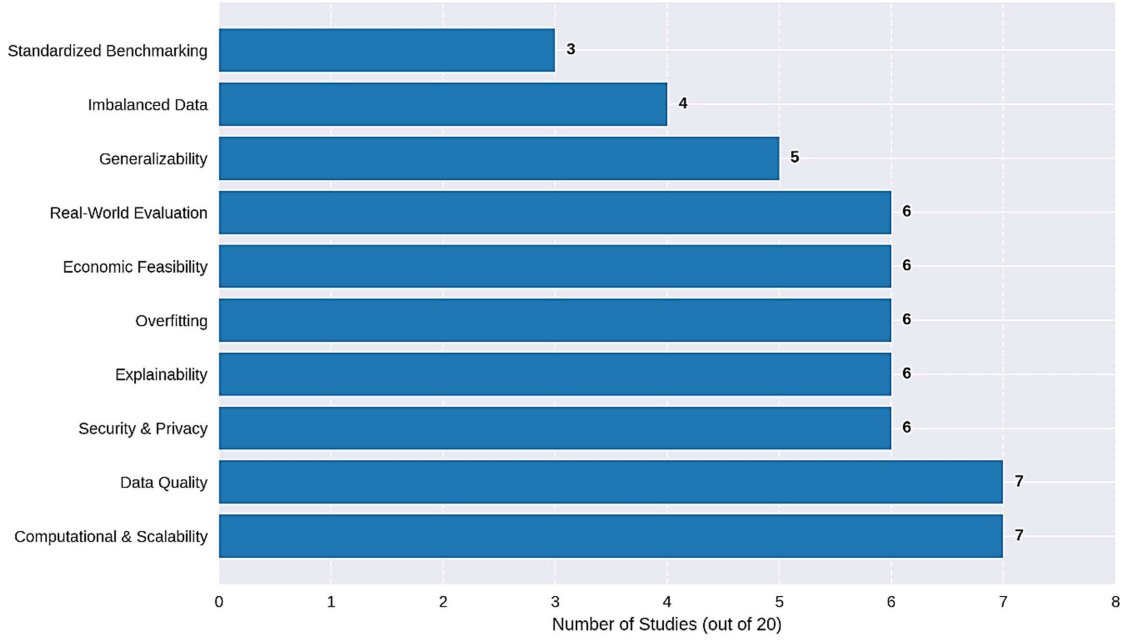


Table 5
Key challenges and proposed recommendations for future AI-based banking fraud detection research (based on 20 studies, 2024–2025)

Challenge	Brief description	Recommendations
Computational complexity and scalability	CNN-LSTM, GNN, and GAN models need a lot of computing power, making real-time handling of millions of transactions expensive and often impractical in banks [13, 14, 16, 18–21]	1. Switch to lighter models (e.g., TinyML). 2. Spread the load with cloud or edge computing. 3. Build modular designs that scale more easily.
Data security and adversarial robustness	Many papers put accuracy first and give less attention to privacy, security, or protection against adversarial attacks—especially serious issues for deep and graph models [10, 16, 17, 21, 24, 25]	1. Add defenses against adversarial examples. 2. Use FL together with differential privacy. 3. Run real-world security checks.
Model explainability and trust	Even with tools like SHAP and LIME, models such as GAN, GNN, and fuzzy ones often lack clear explanations, which makes banks and regulators hesitant to adopt them [14, 17, 19, 22, 23, 26]	1. Build in XAI methods (SHAP, LIME). 2. Create interactive dashboards to show decisions. 3. Set up standard reporting rules for AI in banking.
Data quality and availability	Datasets are frequently noisy, incomplete, or limited to certain regions, hurting real-world reliability [9, 10, 12, 13, 21, 25, 27]	1. Set up shared banking data pools. 2. Generate balanced synthetic data with GANs. 3. Agree on common data cleaning steps.
Generalizability across regions	Models trained on single-country/bank data suffer performance drops when transferred [8, 9, 12, 13, 25]	1. Use transfer learning and domain adaptation. 2. Build datasets from multiple countries. 3. Test under different regulatory settings.
Overfitting and hyperparameter tuning	Complex models (FNN, GAN, GWO) are sensitive to tuning and fail in real-world data [10, 11, 15–17, 26]	1. Try AutoML or Bayesian optimization for tuning. 2. Stick to strict cross-validation. 3. Keep model designs simpler.
Economic and practical feasibility	Studies report accuracy without ROI, deployment cost, or real savings analysis [14, 15, 18, 19, 23, 27]	1. Conduct a cost–benefit analysis alongside accuracy. 2. Define economic key performance indicators (e.g., ROI, savings). 3. Simulate real-world deployment scenarios.
Imbalanced data and Type II errors	Vast legitimate vs. rare fraudulent transactions cause high false negatives [10, 11, 13, 15]	1. Apply resampling (SMOTE, adaptive synthetic sampling approach for imbalanced learning). 2. Use cost-sensitive loss functions. 3. Explore semi-supervised learning.
Lack of real-world validation	Most models tested in labs, not live banking systems [8, 9, 12, 13, 20, 25]	1. Run pilot deployments in real banks. 2. Build shared testing platforms. 3. Enable online learning for adaptation.
Absence of standardized benchmarking	Diverse metrics (precision, F1-score, AUC) and no common datasets hinder comparison [8, 9, 22]	1. Develop standard evaluation frameworks. 2. Create global benchmark datasets. 3. Define hybrid performance metrics.

The key findings of this study are summarized as follows:

1) The two challenges that came up most often were "computational complexity and scalability" and "data quality, completeness and availability"—both in seven studies (35% of the total).
2) Issues around data security, model interpretability, overfitting, economic analysis, and real-world operational testing each showed up in 6 papers (30%).
3) Generalizability across different domains or regions was noted in five studies (25%), while handling imbalanced data remained a stubborn problem in four papers (20%).
4) The least mentioned challenge was the lack of standardized evaluation frameworks—only three studies (15%)—but it still matters a lot because it makes comparing results across papers so difficult.

On average, each study raised about 3.2 challenges, which shows there is quite a bit of overlap in the problems researchers are running into. Overall, this pattern makes it clear that future work should put more effort into improving operational scalability, data security, model interpretability, and testing in real banking environments.

To help overcome the limitations we found and close the gap between lab results and real banking use, Table 5 lists practical suggestions for each of the 10 main challenges (from Table 4). These ideas are meant to point future work toward AI systems that are easier to scale, safer, more understandable, and actually feasible to put into practice in fraud detection.

4.1. Data quality and privacy

One of the biggest roadblocks to building scalable fraud detection systems is the shortage of good, privacy-safe real-world datasets. Synthetic data helps fill some gaps, but it often misses important real-life details like shifting customer behavior, deliberate attacks from adversaries, or changes over time—problems pointed out in about 40% of the papers we looked at [8, 9].

Sharing data between banks is key for training stronger federated models [13, 24, 33–37], but rules like GDPR and CCPA make it almost impossible without very careful anonymization. Fortunately, newer methods are making progress here: advanced anonymization for big data [28], fast in-memory clustering for streaming transactions [29], parallel processing setups [30], and improved FL with built-in privacy [33–37]. These approaches produce datasets that keep the useful patterns (like transaction sequences or merchant connections) while meeting standards like k-anonymity and differential privacy. In short, they help close the data quality gap and make the kind of multi-bank cooperation possible that next-generation systems will need.

4.2. Security and adversarial robustness

Adversarial attacks—things like poisoning training data or inverting models to steal information—can seriously undermine federated or cloud-based fraud systems. These risks came up in roughly 30% of the studies that dealt with shared training setups [18, 24]. Insider threats add another layer of worry in distributed environments.

One useful defense is continuous authentication through biometric behavioral profiling [31], which sets up dynamic trust checks and spots unusual user activity as it happens. On top of that, trust-based security for cloud platforms [32] and newer FL designs that are better at handling adversarial interference [33, 35–37] bring in features like reputation scoring, node-level anomaly detection, and secure aggregation. These stop bad data from sneaking into the training process.

Taken together, these measures make models much tougher against evasion and poisoning attacks, helping keep AI fraud detection reliable when it's running in real banking operations.

5. Conclusion

This systematic review of 20 recent peer-reviewed studies (15 from 2025 and 5 from 2024) shows clear progress in AI-driven banking fraud detection, with leading models achieving 93–96% accuracy on real-world transaction data. However, several recurring challenges remain prominent in more than 30% of the papers: Computational complexity and scalability (seven studies, 35%), data quality and availability (seven studies, 35%), security and adversarial robustness (six studies, 30%), and model interpretability (six studies, 30%). These issues underline a persistent gap between strong laboratory performance and practical deployment in financial institutions.

High accuracy alone is insufficient for adoption. Banks require AI systems that are scalable, secure, transparent, generalizable across regions, and economically justified.

Many studies report promising results on "real-world" data, but these datasets are often limited to single institutions or regions, heavily anonymized under privacy regulations, or partly synthetic—making truly diverse, large-scale, cross-jurisdictional data difficult to obtain.

Privacy-preserving techniques like FL must navigate varying global regulations (e.g., GDPR in Europe, CCPA in the United States, Personal Information Protection Law in China, Personal Data Protection Act in Southeast Asia, and emerging laws in the Middle East and beyond), which complicate cross-bank data sharing and collaborative training.

In conclusion, the future lies not in marginal accuracy gains but in building deployment-ready, trustworthy, and sustainable AI solutions that banks can confidently adopt across diverse regulatory and operational environments. The five prioritized future directions in Table 5 directly target these most frequent challenges:

1) Privacy-preserving FL and multi-national data sharing (top priority) to address data quality, privacy, and generalizability (seen in >30% of studies),
2) Lightweight, edge-deployable models to solve computational complexity and scalability (the most common barrier),
3) Integrated XAI with adversarial robustness to improve interpretability and trust (a major adoption hurdle),
4) Cost-aware evaluation to prove economic feasibility,
5) Adaptive online learning to ensure long-term robustness against evolving fraud patterns and concept drift.

Recommendations

To close the gap between current research and real banking use, future work should prioritize these directions—ordered by frequency and severity of challenges in Section 4:

1) Privacy-preserving FL and multi-national data sharing (top priority)—directly tackle data quality, privacy, and generalizability issues seen in more than 30% of the papers, while handling different international rules.
2) Lightweight models for edge deployment (e.g., TinyML)—address the most common complaint: Computational demands and scalability problems in real-time processing.
3) XAI combined with strong adversarial defenses—boosts interpretability, builds trust, and protects against attacks, which remain big barriers to adoption.
4) Cost-focused evaluations—including ROI and real savings calculations to show economic value in banking settings.
5) Adaptive online learning—to stay robust against shifting fraud tactics and concept drift over the long term.

These steps provide researchers with a practical path to shift from accuracy-focused studies toward AI fraud tools that are truly scalable, secure, transparent, and ready for banks to deploy.

Acknowledgments

We would like to thank the anonymous reviewers and the editorial team for their thoughtful comments and helpful suggestions.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Hamid Banirostam: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Elham Shamsinejad:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing.

References

[1] Oduro, D. A., Okolo, J. N., Bello, A. D., Ajibade, A. T., Fatomi, A. M., Oyekola, T. S., & Owoo-Adebayo, S. F. (2025). AI-powered fraud detection in digital banking: Enhancing security through machine learning. *International Journal of Science and Research Archive*, *14*(3), 1412–1420. https://doi.org/10.30574/ijsra.2025.14.3.0854

[2] Association of Certified Fraud Examiners. (2024). *Occupational fraud 2024: A report to the nations*. https://www.acfe.com/-/media/files/acfe/pdfs/rttn/2024/2024-report-to-the-nations.pdf

[3] Vasant, M., Ganesan, S., & Kumar, G. (2025). Enhancing E-commerce security: A hybrid machine learning approach to fraud detection. *FinTech and Sustainable Innovation*, *1*, A7. https://doi.org/10.47852/bonviewFSI52024882

[4] Wu, Y., Wang, L., Li, H., & Liu, J. (2025). A deep learning method of credit card fraud detection based on continuous-coupled neural networks. *Mathematics*, *13*(5), 819. https://doi.org/10.3390/math13050819

[5] Banirostam, H., Banirostam, T., Pedram, M. M., & Rahmani, A. M. (2024). Analysis and evaluation of various fraud detection methods for electronic payment cards transactions in big data. *Journal of Signal Processing Systems*, *96*(12), 849–870. https://doi.org/10.1007/s11265-025-01947-w

[6] Btoush, E. A. L. M., Zhou, X., Gururajan, R., Chan, K. C., & Alsodi, O. (2025). Enhancing credit card fraud detection with a stacking-based hybrid machine learning approach. *PeerJ Computer Science*, *11*, e3007. https://doi.org/10.7717/peerj-cs.3007

[7] Shamsinejad, E., Banirostam, H., Banirostam, T., Pedram, M. M., & Rahmani, A. M. (2025). Providing and evaluating a model for big data anonymization streams by using in-memory processing. *Knowledge and Information Systems*, *67*(9), 7791–7824. https://doi.org/10.1007/s10115-025-02417-2

[8] Olowu, O., Adeleye, A. O., Omokanye, A. O., Ajayi, A. M., Adepoju, A. O., Omole, O. M., & Chianumba, E. C. (2024). AI-driven fraud detection in banking: A systematic review of data science approaches to enhancing cybersecurity. *GSC Advanced Research and Reviews*, *21*(2), 227–237. https://doi.org/10.30574/gscarr.2024.21.2.0418

[9] Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. *Humanities and Social Sciences Communications*, *11*(1), 1130. https://doi.org/10.1057/s41599-024-03606-0

[10] Ali, A. H., & Hagag, A. A. (2024). An enhanced AI-based model for financial fraud detection. *International Journal of Advanced and Applied Sciences*, *11*(10), 114–121. https://doi.org/10.21833/ijaas.2024.10.013

[11] Islam, M. S., & Rahman, N. (2025). AI-driven fraud detections in financial institutions: A comprehensive study. *Journal of Computer Science and Technology Studies*, *7*(1), 100–112. https://doi.org/10.32996/jcsts.2025.7.1.8

[12] Dichev, A., Zarkova, S., & Angelov, P. (2025). Machine learning as a tool for assessment and management of fraud risk in banking transactions. *Journal of Risk and Financial Management*, *18*(3), 130. https://doi.org/10.3390/jrfm18030130

[13] Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*, *12*(1), 6. https://doi.org/10.1186/s40537-024-01048-8

[14] Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). FraudX AI: An interpretable machine learning framework for credit card fraud detection on imbalanced datasets. *Computers*, *14*(4), 120. https://doi.org/10.3390/computers14040120

[15] Albalawi, T., & Dardouri, S. (2025). Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation. *Frontiers in Artificial Intelligence*, *8*, 1643292. https://doi.org/10.3389/frai.2025.1643292

[16] Alarfaj, F. K., & Shahzadi, S. (2025). Enhancing fraud detection in banking with deep learning: Graph neural networks and autoencoders for real-time credit card fraud prevention. *IEEE Access*, *13*, 20633–20646. https://doi.org/10.1109/ACCESS.2024.3466288

[17] Feng, X., & Kim, S.-K. (2025). Statistical data-generative machine learning-based credit card fraud detection systems. *Mathematics*, *13*(15), 2446. https://doi.org/10.3390/math13152446

[18] Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2026). Deep learning in financial fraud detection: Innovations, challenges, and applications. *Data Science and Management*, *9*(2), 100162. https://doi.org/10.1016/j.dsm.2025.08.002

[19] Koppireddy, C. S., & S., V. D. R. V. D. (2025). Fraud detection in banking: A deep learning approach with explainable AI. *Journal of Soft Computing Paradigm*, *7*(3), 258–275. https://doi.org/10.36548/jscp.2025.3.004

[20] Li, F., & Chen, Z. (2025). Dynamic quantification anti-fraud machine learning model for real-time transaction fraud detection in banking. *Discover Computing*, *28*(1), 59. https://doi.org/10.1007/s10791-025-09549-7

[21] Patil, A., Mahajan, S., Menpara, J., Wagle, S., Pareek, P., & Kotecha, K. (2024). Enhancing fraud detection in banking by integration of graph databases with machine learning. *MethodsX*, *12*, 102683. https://doi.org/10.1016/j.mex.2024.102683

[22] Friska, R., & Agustia, D. (2025). Enhancing auditor judgment quality: A review of evidence from experimental research. *Journal of Risk and Financial Management*, *18*(3), 115. https://doi.org/10.3390/jrfm18030115

[23] Wang, Z., Chen, X., Wu, Y., Jiang, L., Lin, S., & Qiu, G. (2025). A robust and interpretable ensemble machine learning model for predicting healthcare insurance

fraud. *Scientific Reports*, *15*(1), 218. https://doi.org/10.1038/s41598-024-82062-x

[24] Alhasawi, Y., Almtrf, A. A., & Asad, M. (2025). A federated approach to scalable and trustworthy financial fraud detection. *Security and Privacy*, *8*(5), e70099. https://doi.org/10.1002/spy2.70099

[25] Gujrati, R., & Uygun, H. (2025). Integrating artificial intelligence and blockchain for assessing the financial risk of fraud in banking sector. *PromptAI Academy Journal*, *4*, e089.

[26] de Campos Souza, P. V., & Sayyadzadeh, I. (2025). GWO-FNN: Fuzzy neural network optimized via grey wolf optimization. *Mathematics*, *13*(7), 1156. https://doi.org/10.3390/math13071156

[27] Malempati, M. (2025). *Artificial intelligence-driven fraud detection systems and the future of security in financial networks.* SSRN. https://doi.org/10.2139/ssrn.5331008

[28] Sethupathy, U. K. A. (2025). Risk-aware AI models for financial fraud detection: Scalable inference from big transactional data. *International Journal of Intelligence Science*, *15*(4), 162–183. https://doi.org/10.4236/ijis.2025.154009

[29] Shamsinejad, E., Banirostam, T., Pedram, M. M., & Rahmani, A. M. (2024). Anonymizing big data streams using in-memory processing: A novel model based on one-time clustering. *Journal of Signal Processing Systems*, *96*(6), 333–356. https://doi.org/10.1007/s11265-024-01920-z

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

[31] Silva Atencio, G. (2025). Effective cybersecurity strategies for mitigating remote work and IoT risks in enterprises. *FinTech and Sustainable Innovation*, *1*, A13. https://doi.org/10.47852/bonviewFSI52025962

[32] Polytarchos, E. (2025). Credit card fraud detection through deep learning and real-time data streams: A comparison and new directions. *FinTech and Sustainable Innovation*. https://doi.org/10.47852/bonviewFSI52026108

[33] Xia, Z., & Saha, S. C. (2025). FinGraphFL: Financial graph-based federated learning for enhanced credit card fraud detection. *Mathematics*, *13*(9), 1396. https://doi.org/10.3390/math13091396

[34] AbouGrad, H., & Sankuru, L. (2025). Online banking fraud detection model: Decentralized machine learning framework to enhance effectiveness and compliance with data privacy regulations. *Mathematics*, *13*(13), 2110. https://doi.org/10.3390/math13132110

[35] Abbassi, H., El Mendili, S., & Gahi, Y. (2025). Adaptive, privacy-enhanced real-time fraud detection in banking networks through federated learning and VAE-QLSTM fusion. *Big Data and Cognitive Computing*, *9*(7), 185. https://doi.org/10.3390/bdcc9070185

[36] Aljunaid, S. K., Almheiri, S. J., Dawood, H., & Khan, M. A. (2025). Secure and transparent banking: Explainable AI-driven federated learning model for financial fraud detection. *Journal of Risk and Financial Management*, *18*(4), 179. https://doi.org/10.3390/jrfm18040179

[37] Zheng, Y. (2025). Bank data protection and fraud identification based on improved adaptive federated learning and WGAN. *Scientific Reports*, *15*(1), 23006. https://doi.org/10.1038/s41598-025-06807-y

How to Cite: Banirostam, H., & Shamsinejad, E. (2026). A Systematic Review of AI-Driven Banking Fraud Detection: Advances, Challenges, and Deployment-Ready Solutions (2024–2025). *FinTech and Sustainable Innovation*, *2*, A19. https://doi.org/10.47852/bonviewFSI62028228