JIRRS
Journal International Review of Research Studies

Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

NEURAL NETWORKS FOR REAL-TIME FINANCIAL FRAUD DETECTION

Deborah Natany Otoni Oliveira
Faculdade Presidente Antônio Carlos, Brazil
Corresponding author: deborahnatany84@gmail.com

Abstract

The accelerating digitalization of financial services has transformed fraud into a systemic global threat, with annual losses estimated at $485.6 billion in 2023 alone — losses that conventional rule-based and statistical detection methods have proven structurally incapable of containing. This article presents a narrative conceptual literature review examining how neural network architectures are redefining real-time fraud detection in financial systems. The review covers the theoretical and epistemological foundations of the field, the principal neural architectures deployed — including Multilayer Perceptrons (MLPs), Long Short-Term Memory networks (LSTMs), Convolutional Neural Networks (CNNs), Autoencoders, Graph Neural Networks (GNNs), and Transformer-based models — and the production infrastructure requirements that constrain their deployment within payment authorization pipelines operating under sub-100-millisecond latency budgets. Critical operational challenges are analyzed in depth, including concept drift, adversarial evasion attacks, class imbalance, algorithmic explainability under GDPR and LGPD regulatory frameworks, and the data privacy constraints that motivate Federated Learning and Differential Privacy approaches. Real-world implementations across credit card networks, instant payment systems — with particular attention to Brazil's PIX ecosystem — anti-money laundering operations, and insurance fraud are documented. The review concludes by mapping the research frontier, encompassing online learning, Large Language Models as fraud detection orchestrators, generative AI weaponized by fraudsters, and the long-term potential of Quantum Machine Learning. Findings indicate that while no single architecture dominates across all fraud typologies, hybrid and ensemble frameworks combining temporal, relational, and anomaly-detection capabilities consistently achieve superior performance, and that the integration of regulatory compliance, explainability, and adversarial robustness alongside predictive accuracy represents the defining challenge for the next generation of production fraud detection systems.

JIRRS
Journal International Review of Research Studies

Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Keywords: fraud detection; neural networks; deep learning; real-time systems; graph neural networks; federated learning; explainable AI; financial security; class imbalance; concept drift.

1. Introduction

The accelerating digitalization of financial services has substantially expanded the attack surface available to malicious actors, transforming fraud into a systemic global threat. Financial fraud encompasses deceptive practices such as credit card fraud, insurance fraud, and money laundering, and global estimates suggest that organizations lose approximately 5% of annual revenues to fraud, equivalent to trillions of dollars worldwide (Ngo et al., 2025). The economic scale is made concrete by institutional data: fraud scams and bank fraud schemes resulted in $485.6 billion in losses globally in 2023 alone, while an estimated $3.1 trillion in illicit funds flowed through the global financial system during the same period (Nasdaq Verafin, 2024). These figures reflect an escalating phenomenon that extends across all verticals of digital finance, eroding institutional trust and imposing mounting costs on financial institutions and consumers alike.

Conventional detection methodologies have proven structurally inadequate in responding to this challenge. Traditional rule-based systems and manual review processes frequently produce elevated false positive and false negative rates, and critically, cannot adapt to evolving fraud patterns without human intervention, requiring domain experts to manually engineer and update detection features (Hilal et al., 2022). This rigidity is particularly consequential in the contemporary threat landscape, where fraudsters exploit automation and advanced technologies to launch large-scale, sophisticated attacks that outpace the cadence at which rule sets can be manually revised (Teixeira et al., 2024). The operational overhead and structural brittleness of legacy approaches have thus created a critical detection gap that demands fundamentally different solutions.

Within this context, the temporal dimension of fraud detection emerges as the decisive factor separating effective prevention from financial loss. Modern payment authorization pipelines operate within strict latency budgets, and the difference between milliseconds and microseconds can represent the contrast between incurring substantial financial losses or avoiding them entirely (Hong et al., 2024). Payment processors typically require scoring decisions within 100 milliseconds to avoid transaction delays, necessitating sophisticated stream processing architectures capable of handling millions of transactions per second while maintaining low latency (Lu et al., 2022). Neural network architectures — including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Graph Neural Networks (GNNs) — have emerged as the dominant paradigm for meeting these

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

demands, enabling real-time anomaly detection and continuous adaptation to evolving fraud patterns (Ngo et al., 2025; Nicholls et al., 2021). This article examines how these architectures are redefining fraud detection in live financial systems, reviewing their theoretical foundations, operational integration, performance benchmarks, and open challenges.

2. Methodological Clarification

2.1 Nature and Scope of the Research

This article constitutes a narrative literature review with a comparative analytical dimension, focusing on the application of neural network architectures to real-time fraud detection in financial systems. The temporal scope covers the period from 2019 to 2024, a window selected to capture the most recent advancements in deep learning applied to fraud detection while reflecting the significant growth in publication output observed from 2022 onwards (Ngo et al., 2025). Geographically, the review is not restricted to a single jurisdiction, encompassing studies from North American, European, and Asian research contexts, given the global nature of digital financial fraud. Studies were included if they employed deep learning techniques — specifically CNNs, LSTMs, GNNs, or Transformer-based architectures — in the context of financial fraud detection, were peer-reviewed, and were published in English. Studies unrelated to financial services, lacking performance evaluation metrics, or relying exclusively on traditional machine learning without neural components were excluded (Abdallah et al., 2022; Teixeira et al., 2024).

2.2 Sources and Databases

The literature search was conducted across major academic databases, including IEEE Xplore, ACM Digital Library, ScienceDirect, Scopus, and Google Scholar, supplemented by preprint repositories such as arXiv for recent unpublished contributions. Database searches were conducted between January and February 2026. Search terms combined keywords related to neural networks, fraud detection, and real-time processing. The initial database search retrieved approximately 312 records across all databases. After removing duplicate entries and screening titles and abstracts for relevance, 198 articles remained for full-text assessment. Following the application of the inclusion and exclusion criteria described in Section 2.1, a final corpus of 74 studies was retained for detailed analysis in this review. Regarding experimental datasets, three benchmarks recur prominently in the literature: the ULB Credit Card Fraud Dataset, a real-world anonymized dataset produced through a collaboration between the Université Libre de Bruxelles and Worldline (Dal Pozzolo,

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

2016); the IEEE-CIS Fraud Detection Dataset, derived from Vesta Corporation's e-commerce transactions (Howard et al., 2019); and PaySim, a synthetic dataset designed to simulate mobile financial transactions for fraud detection research (Lopez-Rojas, 2016). These datasets are widely cited for enabling reproducible benchmarking, but their limitations impose significant constraints on the generalizability of reported findings that merit explicit elaboration across three dimensions. To improve transparency and reproducibility, representative search queries included combinations of the following terms:

("fraud detection" OR "financial fraud") AND ("neural networks" OR "deep learning" OR "graph neural networks" OR "transformers") AND ("real-time" OR "stream processing" OR "online detection").

Equivalent keyword variations were adapted for the syntax of each database.

First, **temporal aging**: the ULB Credit Card dataset captures European cardholder transactions from September 2013, meaning its fraud patterns reflect an attack landscape more than a decade old. The IEEE-CIS dataset, derived from 2017–2019 Vesta Corporation e-commerce data, is similarly dated. Given that fraud tactics evolve continuously — with synthetic identity fraud, account takeover via social engineering, and adversarial ML-based evasion emerging as dominant vectors only in the 2020s — models optimized on these benchmarks may be structurally misaligned with the fraud patterns present in current production environments (Board of Governors of the Federal Reserve System, 2025).

Second, **class imbalance ratios**: the ULB dataset contains 492 fraud cases out of 284,807 transactions, yielding a fraud prevalence of approximately 0.172% — a ratio so extreme that even a classifier predicting all transactions as legitimate achieves 99.8% accuracy. The IEEE-CIS dataset exhibits a similarly skewed distribution, with fraud comprising approximately 3.5% of transactions. PaySim, as a synthetically generated dataset, allows imbalance ratios to be configured by the researcher, introducing variability in reported results that is rarely disclosed explicitly. These ratios mean that standard accuracy metrics are uninformative, and that small differences in F1-score or AUC across studies may reflect differences in resampling strategy rather than architectural superiority (Ngo et al., 2025; Abdallah et al., 2022).

Third, **synthetic vs. real data validity**: PaySim is generated by a simulator calibrated on a real mobile money dataset from a single African operator, and its ability to reproduce the statistical properties of real fraud — particularly the adversarial, adaptive component introduced by human fraudsters — is inherently limited. Models trained or evaluated on PaySim may perform well on the synthetic distribution while failing to generalize to real transaction environments where fraud patterns are shaped

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

by strategic evasion rather than statistical simulation (Lopez-Rojas, 2016; Semmelrock et al., 2025).

Industry reports from the Association of Certified Fraud Examiners (ACFE), Nasdaq Verafin, and the Federal Bureau of Investigation were also consulted to contextualize academic findings within empirical loss data.

Overall, the review process followed a three-stage screening procedure consisting of database retrieval (n = 312), title and abstract screening (n = 198), and full-text eligibility assessment, resulting in a final analytical corpus of 74 studies.

2.3 Methodological Approach

The analytical core of this review involves a comparative evaluation of neural network architectures applied to financial fraud detection, assessing their relative strengths across operational criteria. Performance evaluation follows the metrics most widely adopted in the literature: Precision, Recall, F1-Score, and AUC-ROC, which together provide a multidimensional view of classifier performance on highly imbalanced datasets (Ngo et al., 2025; Abdallah et al., 2022). A particular emphasis is placed on inference latency as an additional evaluation dimension, distinguishing real-time scoring systems — which must deliver decisions within milliseconds to remain operationally viable — from batch processing approaches that analyze accumulated transaction logs post hoc. This distinction is theoretically and practically significant: while batch analysis permits deeper retrospective pattern recognition, real-time inference is the critical requirement for preventing fraudulent transactions before fund disbursement occurs (Hilal et al., 2022). The comparative framework thus addresses not only predictive accuracy but also the architectural trade-offs between model complexity and deployment feasibility under strict latency constraints.

2.4 Methodological Limitations

Several constraints bear on the scope and generalizability of this review. First, access to real financial transaction data is severely restricted by data protection regulations, including the General Data Protection Regulation (GDPR) in the European Union and the Lei Geral de Proteção de Dados (LGPD) in Brazil, which impose stringent limitations on data sharing across institutions and jurisdictions, even for research purposes (Corrêa et al., 2024; Semmelrock et al., 2025). As a result, most experimental studies rely on a limited set of public benchmarks, many of which are dated, heavily anonymized, or synthetically generated, constraining the external validity of reported findings (Board of Governors of the Federal Reserve System, 2025). Second, the literature exhibits a well-documented publication bias toward positive results, with studies reporting high F1-scores and AUC-ROC values on

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

benchmark datasets, while negative outcomes, failed architectures, or results on proprietary data remain largely unpublished (Semmelrock et al., 2025). Third, reproducibility poses a structural challenge: experiments conducted on institutional or proprietary datasets cannot be independently verified, and even studies using public datasets frequently omit sufficient implementation detail to permit exact replication (Semmelrock et al., 2025). These limitations are acknowledged throughout the analysis and inform the interpretation of comparative performance claims presented in subsequent sections. Two dimensions of external validity, however, warrant explicit elaboration beyond their initial acknowledgment.

The first concerns the **transferability gap** between benchmark performance and real banking environments. Models evaluated on the ULB, IEEE-CIS, or PaySim datasets are assessed against static, historically fixed fraud distributions. In production banking environments, the fraud distribution is non-stationary — shaped by an adversarial dynamic in which fraudsters continuously adapt their tactics in response to detection systems — and the feature space available at inference time is substantially richer and more institution-specific than the anonymized or synthetic features available in public benchmarks. A model achieving AUC 0.99 on the ULB dataset may therefore exhibit substantially degraded performance when deployed against live transaction streams at a retail bank, where concept drift, novel fraud typologies, and institution-specific behavioral baselines collectively undermine the statistical assumptions on which benchmark evaluation rests (Board of Governors of the Federal Reserve System, 2025; Semmelrock et al., 2025). The reviewed literature provides limited evidence on this transferability gap because cross-environment validation requires access to institutional data that regulatory constraints make unavailable for publication.

The second dimension concerns **regulatory barriers to cross-institutional model training**. Federated learning and collaborative fraud detection approaches — which would allow institutions to train shared models without centralizing sensitive transaction data — face significant regulatory friction under GDPR in the European Union and LGPD in Brazil. Both frameworks impose restrictions on the processing and transfer of personal financial data that extend, under some regulatory interpretations, to gradient updates and model parameters derived from such data, not merely to the raw data itself (Corrêa et al., 2024). This creates a structural barrier to the cross-institutional collaboration that would be most effective at detecting fraud rings and money laundering networks that span multiple institutions — precisely the fraud typologies for which single-institution models are least adequate. The practical consequence for this review is that the most operationally significant fraud detection challenges are also the least well-represented in the published literature, because the data governance conditions required to study them are incompatible with open publication.

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Another limitation of this review concerns the narrative synthesis methodology. Although structured search and screening procedures were applied, the interpretation and integration of findings remain inherently dependent on the authors' analytical perspective, which may introduce subjective bias in the selection of representative studies and interpretation of research trends.

3. Theoretical Framework and Epistemological Positioning

3.1 Epistemological Positioning

Research in applied data science is predominantly grounded in a positivist paradigm, which treats empirical observation, quantitative measurement, and hypothesis testing as the primary instruments for knowledge generation (Park et al., 2020). In the context of fraud detection, this orientation is reflected in the framing of detection as a pattern classification problem solvable through statistical optimization — an approach that seeks lawlike regularities in transactional data and operationalizes performance in terms of objectively measurable metrics. However, a constructivist counterreading is also applicable: the very definition of what constitutes "fraudulent behavior" is socially and institutionally constructed, shaped by regulatory contexts, historical labeling decisions, and the perspectives of those who annotated training datasets (Pretorius, 2024). This tension is most acutely expressed in the accuracy–interpretability trade-off that pervades neural network applications in finance. High-complexity deep learning models — such as deep neural networks, graph neural networks, and ensemble methods — achieve superior predictive performance but operate as opaque black boxes, generating outputs whose underlying logic is inaccessible to regulators, auditors, and affected individuals (Černevičienė & Kabašinskas, 2024). The epistemological question this raises is not merely technical: if a model cannot explain its decisions, what does it actually "know," and on what basis does it exercise authority over consequential financial determinations?

This epistemological tension has direct practical consequences that extend beyond academic philosophy. The positivist framing of fraud detection as a classification optimization problem tends to render invisible the institutional power dynamics embedded in algorithmic decision-making: when a neural network denies a transaction, blocks an account, or flags a customer for investigation, it exercises a form of algorithmic authority — consequential power over financial access — that is structurally insulated from the contestation mechanisms that govern human decision-makers (Pretorius, 2024). Regulatory frameworks such as GDPR and LGPD recognize this dynamic implicitly through their right-to-explanation provisions, which require that automated decisions with significant effects on individuals be explainable in human-intelligible terms (Corrêa et al., 2024). The constructivist insight further complicates model evaluation: if the training labels that define "fraud" were generated

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

by prior rule-based systems or human reviewers operating under specific institutional biases, then a neural network trained on those labels inherits and potentially amplifies those biases — producing a system that is statistically accurate relative to its training distribution but systematically unjust relative to the population it governs (Černevičienė & Kabašinskas, 2024). This concern is particularly acute in demographically skewed datasets where protected characteristics are correlated with transaction patterns, raising the prospect that high-AUC models simultaneously discriminate against identifiable population subgroups. The epistemological positioning of this review therefore treats performance metrics as necessary but insufficient criteria for evaluating fraud detection systems, and treats interpretability, fairness auditability, and contestability as co-equal design requirements alongside predictive accuracy.

3.2 Theoretical Foundations of Artificial Neural Networks

The theoretical lineage of modern neural networks originates in the work of McCulloch and Pitts (1943), who proposed a simplified mathematical model of the biological neuron — a binary threshold device with multiple inputs and a single output that fired when the weighted sum of its inputs exceeded a predefined threshold (McCulloch & Pitts, 1943). This foundational model established a formal correspondence between neurological computation and Boolean logic. Frank Rosenblatt subsequently extended this framework in 1957 with the Perceptron, which introduced learnable weights adjusted through an iterative error-correction algorithm, enabling the network to classify linearly separable inputs from data rather than from analytical determination alone (Rosenblatt, 1962). The critical theoretical maturation of the field came with the Universal Approximation Theorem which demonstrated that a single hidden-layer feedforward network with a sufficient number of neurons and a non-polynomial activation function can approximate any continuous function to an arbitrary degree of precision (Hornik et al., 1989). This result provides the theoretical justification for the deployment of neural networks in fraud detection: since fraudulent transaction patterns constitute complex, non-linear, high-dimensional functions of input features, a sufficiently expressive network is theoretically capable of modeling them. Deep learning architectures subsequently extended this framework by exploiting depth rather than width, enabling hierarchical feature extraction that scales far more efficiently with problem complexity (Ngo et al., 2025).

3.3 Anomaly Detection Theory

Within the theoretical vocabulary of machine learning, fraud is formally treated as a specific category of anomaly — an observation that deviates significantly from the expected distribution of normal behavior in a given population (Hilal et al., 2022). Anomaly detection methodologies are systematically classified into three paradigms

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

according to the availability and nature of labeled data. Supervised approaches require fully labeled datasets in which each instance is annotated as normal or fraudulent, enabling the training of high-accuracy classifiers; however, they are constrained by their reliance on labeled examples and their limited adaptability to previously unseen fraud typologies (Hilal et al., 2022). Unsupervised approaches require no labeled data and instead model the statistical structure of normal behavior, flagging deviations as potential anomalies — a framework particularly suited to detecting novel fraud schemes, although at the cost of elevated false positive rates (Hilal et al., 2022). Semi-supervised learning occupies a theoretically important intermediate position, training on a small labeled corpus alongside large volumes of unlabeled data, and has proven especially effective in financial fraud contexts where labeled fraud instances are scarce relative to the volume of legitimate transactions (Ngo et al., 2025). From an information-theoretic perspective, anomaly detection can also be understood as the identification of instances with high description complexity relative to a learned model of normal data — that is, observations that require substantially more information to encode than the baseline distributional model predicts (Chandola et al., 2009).

3.4 Theoretical Foundations of Real-Time Processing

The distinction between real-time and batch processing systems has precise theoretical foundations. Real-time systems are formally defined as computational architectures in which correctness depends not only on logical accuracy but also on temporal constraints: a correct result delivered outside its deadline is, by definition, a system failure. This taxonomy distinguishes between hard real-time systems — where deadline violations are catastrophic — and soft real-time systems — where occasional latency overruns are tolerable (Hilal et al., 2022). Payment fraud detection belongs to the soft real-time category, but with de facto hard constraints in practice: major payment processors require end-to-end scoring within approximately 100 milliseconds, and latency exceeding this threshold generates perceptible delays that degrade user experience and trigger transaction abandonment (Lu et al., 2022). Stream processing — the continuous ingestion and analysis of data as it arrives — contrasts with batch processing, which accumulates data over time before performing analysis. While batch systems permit deeper retrospective modeling and are less constrained by latency, they are structurally incapable of intercepting fraud at the moment of transaction authorization. Queuing theory further formalizes these trade-offs: in high-throughput payment systems, the arrival rate of transactions approximates a Poisson process, and system stability requires that processing capacity exceed this arrival rate, otherwise queue lengths grow unboundedly, introducing latency that compounds non-linearly (Hilal et al., 2022).

3.5 The Class Imbalance Problem as a Theoretical Issue

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Fraudulent transactions constitute between 0.1% and 2% of total transaction volume in most financial datasets, creating one of the most structurally challenging problems in applied machine learning: extreme class imbalance (Ngo et al., 2025). The epistemological implications of this asymmetry are profound: a classifier that labels every transaction as legitimate achieves accuracy rates exceeding 99% while possessing zero practical utility. Standard performance metrics such as overall accuracy therefore become meaningless, and detection performance must be evaluated through precision, recall, F1-score, and AUC-ROC — metrics that are explicitly sensitive to minority-class performance (Abdallah et al., 2022). The dominant theoretical response to this challenge is synthetic oversampling. SMOTE (Synthetic Minority Over-sampling Technique), proposed by Chawla et al. (2002), generates synthetic minority class instances by interpolating along line segments connecting each minority observation and its k-nearest neighbors in feature space, effectively expanding the decision boundary around underrepresented fraud instances (Chawla et al., 2002). ADASYN (Adaptive Synthetic Sampling) extends this logic by weighting synthetic generation according to the density of majority-class instances around each minority point: regions where legitimate transactions outnumber fraudulent ones most severely receive proportionally more synthetic fraud samples (He et al., 2008). Both approaches, however, carry theoretical risks — including the generation of noisy synthetic examples, the displacement of decision boundaries into majority-class regions, and the potential for models to learn the artifacts of the synthetic generation process rather than genuine fraud patterns (Semmelrock et al., 2025).

3.6 Algorithmic Bias, Ethics, and Responsibility

The deployment of neural networks in financial fraud detection carries substantial ethical dimensions that extend beyond technical performance. Because these models are trained on historical transaction data, they necessarily encode the distributional patterns and labeling decisions of the past — including any discriminatory practices embedded in earlier detection systems (Ntoutsi et al., 2020). In credit and fraud scoring contexts, models have been documented to produce systematically higher false positive rates for minority racial and ethnic groups, effectively flagging legitimate transactions by these customers as suspicious at disproportionate rates and restricting their access to financial services (Mehrabi et al., 2021). The accountability question compounds this concern: when an automated fraud detection system erroneously blocks a legitimate payment, the decision was made by a model whose weights are distributed across millions of parameters with no humanly interpretable causal chain. Regulatory frameworks such as the EU AI Act and the GDPR's right-to-explanation provision have begun to impose legal obligations for transparency in automated financial decisions (Corrêa et al., 2024). Research in fairness-aware

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

machine learning has responded by proposing fairness constraints — including equalized odds, demographic parity, and counterfactual fairness — that can be embedded as optimization objectives during model training, though achieving these constraints without sacrificing predictive performance on the fraud detection task remains an open theoretical and empirical problem (Mehrabi et al., 2021).

4. Financial Fraud: Landscape and Typology

4.1 Definition and Classification of Financial Fraud

Financial fraud encompasses a broad spectrum of deceptive practices designed to generate illicit financial gain through misrepresentation, concealment, or abuse of financial systems. Among the most prevalent categories is credit card fraud, which is formally subdivided into card-present (CP) and card-not-present (CNP) fraud. CP fraud involves the physical theft or cloning of payment cards to conduct in-person transactions, while CNP fraud — the dominant modality in contemporary digital finance — occurs when stolen card credentials are exploited in online or telephone transactions where the physical card is not required (Razaque et al., 2023). CNP fraud accounted for approximately 71% of all card fraud losses in the United States in 2024, driven by the proliferation of stolen card records traded on dark web marketplaces (Hayashi, 2026). Wire transfer fraud — encompassing systems such as PIX, TED, and SWIFT — constitutes a second critical category, in which fraudsters manipulate payment initiation processes or exploit weaknesses in authentication protocols to redirect funds to fraudulent accounts (Hilal et al., 2022). Account takeover (ATO) fraud represents a closely related modality: unauthorized actors gain control of legitimate accounts through credential stuffing, phishing, or social engineering, subsequently conducting fraudulent transactions or extracting sensitive data (Nicholls et al., 2021). Distinct from ATO, identity fraud involves the creation of entirely synthetic or composite identities — combining real and fabricated personal information — to open fraudulent credit accounts or initiate loan applications (Teixeira et al., 2024). At the systemic level, money laundering involves the concealment of illicitly obtained proceeds through layered financial transactions, while insurance and credit fraud encompass misrepresentation schemes targeting underwriting and lending processes (Abdallah et al., 2022).

4.2 Scale and Statistical Dimensions of the Problem

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

The aggregate economic impact of financial fraud has reached a scale that demands systemic institutional responses. Global estimates suggest that fraud scams and bank fraud schemes produced approximately $485.6 billion in losses in 2023, while an estimated $3.1 trillion in illicit funds transited the global financial system during the same period (Nasdaq Verafin, 2024). In the United States, consumers reported over $12.5 billion in fraud losses to the Federal Trade Commission in 2024, representing a 25% increase from the prior year (Federal Trade Commission, 2024). ATO fraud alone generated nearly $13 billion in losses in 2023, with projections placing this figure at $17 billion by 2025 (Javelin Strategy & Research, 2024). Brazil constitutes a particularly significant case study in the intersection of rapid digital payment adoption and fraud escalation. Since the launch of PIX in November 2020 by the Banco Central do Brasil, the system has grown to process over 63.4 billion transactions annually, representing 52% year-over-year growth (Banco Central do Brasil, 2025). This infrastructure expansion has been accompanied by a commensurate rise in fraud: PIX-related scams generated approximately R$2.7 billion in direct losses in 2024 — a 43% increase from the prior year — while total estimated annual losses from financial fraud in Brazil reached R$297.7 billion, equivalent to approximately 2.5% of the country's GDP (Silverguard, 2024). These figures underscore the structural vulnerability embedded in the rapid scaling of instant payment systems without proportional investment in fraud controls.

These figures, however, represent only the directly measurable component of a broader set of economic and institutional costs that are structurally undercounted in the literature. The indirect costs of financial fraud — encompassing elevated transaction fees passed to consumers, the compliance and operational overhead borne by financial institutions, the chilling effect on digital payment adoption in populations with elevated fraud exposure, and the systemic erosion of institutional trust in payment infrastructure — are estimated to multiply direct losses by a factor of three to five in developed economies (ACFE, 2024). From an institutional economics perspective, financial fraud constitutes a negative externality of digital payment network growth: the expansion of instant payment systems such as PIX, FedNow, and SEPA Instant generates network benefits that accrue broadly across the economy while concentrating fraud risk on individual institutions and consumers who lack the information asymmetry advantages held by organized fraud networks. This dynamic creates underinvestment incentives — individual institutions bear the cost of fraud controls while the benefits of a more secure payment ecosystem are diffused — and provides the economic rationale for regulatory mandates and cross-institutional fraud intelligence sharing frameworks (Hilal et al., 2022; Banco Central do Brasil, 2025). The fraud detection problem is therefore not merely a machine learning optimization challenge but an institutional design problem: effective fraud prevention at the systemic level requires governance structures, data sharing

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

agreements, and regulatory frameworks that align institutional incentives toward collective security, not only technical architectures capable of classification at low latency.

4.3 Evolution of Fraud Tactics

Contemporary fraud is no longer the province of isolated actors exploiting simple security gaps but increasingly reflects the operations of organized criminal networks deploying sophisticated, technology-mediated attack strategies. Fraud rings coordinate large-scale campaigns using automation and shared intelligence, enabling simultaneous attacks across multiple financial institutions and jurisdictions (Nicholls et al., 2021). Social engineering has evolved far beyond rudimentary phishing emails: generative AI now enables the production of highly personalized, grammatically accurate communications that are substantially more difficult for recipients to identify as fraudulent (Ngo et al., 2025). The weaponization of deepfake technology represents a qualitative escalation in threat sophistication: AI-generated synthetic media — including voice cloning, face-swapped video, and full fabricated video conference participants — are being deployed to bypass identity verification systems and authorize fraudulent transactions (Papakostas et al., 2025). A high-profile 2024 case in Hong Kong, in which a finance employee transferred US$25 million following a deepfake video conference impersonating company executives, illustrates the operational maturity of these techniques (Deloitte, 2024). Adversarial attacks — in which fraudsters deliberately craft inputs designed to deceive machine learning detection models — represent a further emerging vector, creating a continuous arms race dynamic between detection systems and the actors seeking to evade them (Ngo et al., 2025).

4.4 Limitations of Traditional Detection Systems

The structural inadequacies of legacy fraud detection systems have been extensively documented in the literature. Rule-based systems — which flag transactions matching a set of manually defined conditions — are inherently rigid: they require domain experts to engineer, test, and update rule sets, a process that cannot keep pace with the speed at which fraud patterns evolve (Al-Hashedi & Magalingam, 2021). Classical statistical models, including logistic regression and decision trees, while interpretable and computationally efficient, impose performance ceilings that reflect their fundamental assumptions. Logistic regression's reliance on linear decision boundaries renders it structurally incapable of capturing the complex, non-linear interaction patterns characteristic of sophisticated fraud (Teixeira et al., 2024). Comparative benchmarks on standard datasets confirm this limitation: logistic regression achieves AUC scores in the range of 0.66 on imbalanced fraud datasets, compared to ensemble and deep learning methods exceeding 0.99 (Abdelghafour et

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

al., 2024). The false positive problem presents an additional operational liability: rule-based systems that are tuned to maximize detection sensitivity generate excessive alerts on legitimate transactions, imposing customer experience degradation and analyst fatigue that collectively reduce system effectiveness (Hilal et al., 2022). Finally, the latency profiles of legacy detection infrastructures are fundamentally incompatible with real-time authorization requirements: systems built on batch-processing architectures introduce delays that exceed the millisecond-range decision windows demanded by modern payment rails, creating an irreducible temporal gap between transaction initiation and fraud identification (Lu et al., 2022).

5. Neural Networks: Core Concepts Applied to Fraud Detection

5.1 Basic Neural Network Architecture

A feedforward neural network is organized into three functional layers: the input layer, which receives raw transaction features as a numeric vector with one node per input dimension; one or more hidden layers, where learned representations are constructed through weighted linear combinations followed by non-linear transformations; and an output layer, which produces the classification decision (Goodfellow et al., 2016). The non-linearity introduced at each hidden layer is the architectural property that enables networks to model complex decision boundaries. The Rectified Linear Unit (ReLU), defined as $f(x) = \max(0, x)$, has become the dominant activation function for hidden layers in deep fraud detection models due to its computational efficiency and resistance to the vanishing gradient problem that hampers learning in deep architectures with sigmoidal activations (Bhujade & Asthana, 2023). In the output layer, binary fraud classification tasks typically employ a single sigmoid neuron, which maps the network's logit to a probability in [0,1] interpretable as the likelihood of fraud (Goodfellow et al., 2016). Training is accomplished through backpropagation — the application of the chain rule to propagate the gradient of the loss function backward through the network — combined with gradient descent or adaptive optimizers such as Adam, which iteratively adjust weights to minimize prediction error on labeled training data (Ngo et al., 2025).

5.2 Fraud Classification as a Learning Task

Fraud detection is formally framed as a supervised binary classification problem: given a transaction and its contextual features, the model estimates P(fraud | transaction, context), the posterior probability that the transaction is fraudulent (Hilal et al., 2022). The standard optimization objective is Binary Cross-Entropy (BCE) loss,

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

which measures the log-divergence between the predicted probability and the true binary label; it penalizes both overconfident incorrect predictions and underconfident correct ones, providing informative gradients throughout training (Goodfellow et al., 2016). In the context of class imbalance, a weighted variant of BCE assigns a proportionally larger penalty to misclassified fraud instances, effectively rescaling the loss contribution of minority-class examples to prevent the optimizer from converging to the degenerate solution of predicting all transactions as legitimate (Sundararajan et al., 2020). Once training is complete, the classifier produces a continuous fraud probability score for each transaction, and a decision threshold converts this score into a binary outcome. The choice of threshold governs the precision-recall trade-off: lowering it increases recall (capturing more true frauds) at the cost of elevated false positives, while raising it improves precision at the cost of missed detections — a calibration decision that must be made in accordance with the asymmetric costs of each error type (Abdallah et al., 2022).

5.3 Feature Engineering for Fraud Detection

The predictive performance of neural network fraud detectors depends critically on the richness and relevance of input features. Behavioral features — including spending patterns, geographic location, time-of-day, device fingerprints, and channel identifiers — capture the statistical regularity of individual cardholders' transactional behavior, enabling the model to identify deviations that are anomalous for a specific user even if they appear plausible in aggregate (Ileberi et al., 2024). Network-relational features encode structural information about the graph of relationships between accounts, IP addresses, beneficiaries, and devices; fraudulent activity frequently exhibits distinctive graph signatures, such as multiple accounts sharing an IP, a beneficiary receiving funds from an unusual number of senders, or newly created accounts with no transaction history (Lu et al., 2022). Temporal features quantify the dynamics of transaction sequences: velocity metrics measuring the number of transactions within rolling time windows, inter-transaction intervals, and seasonal deviations from baseline behavior are particularly informative for detecting coordinated attacks (Xiao et al., 2023). In real-time deployment, computing these features from streaming data requires a feature store architecture — a dedicated infrastructure layer that maintains pre-aggregated behavioral statistics updated continuously from incoming transactions, ensuring that the model receives fresh, low-latency feature values at inference time rather than stale pre-computed snapshots (Ileberi et al., 2024).

5.4 Class Imbalance in Practice

The practical management of class imbalance in fraud detection draws on a complementary toolkit of data-level and algorithm-level interventions. Among

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

oversampling approaches, SMOTE (Chawla et al., 2002) remains the foundational method, generating synthetic minority-class instances by interpolating between existing fraud examples in feature space; Borderline-SMOTE extends this by concentrating synthetic generation on fraud instances that lie near the decision boundary, where misclassification risk is highest. ADASYN (He et al., 2008) further refines this logic adaptively, directing synthetic sample generation proportionally to the local density of majority-class instances. Undersampling approaches work from the opposite direction by removing redundant or misleading majority-class instances: Tomek Links identifies and removes pairs of borderline instances from opposing classes, cleaning the decision boundary region, while Edited Nearest Neighbors (ENN) removes majority-class instances whose class label is inconsistent with the predictions of their nearest neighbors, reducing noise in the training distribution (Batista et al., 2004). Hybrid methods such as SMOTE-ENN and SMOTE-Tomek combine both strategies, yielding improved F1-scores on fraud benchmarks relative to either approach alone — Random Forest with SMOTE-Tomek, for instance, achieves F1-scores of 0.868 on the ULB Credit Card dataset, outperforming models trained on the original imbalanced data (Rtayli & Enneya, 2024). Cost-sensitive learning offers an algorithmic alternative: rather than rebalancing the dataset, it penalizes false negatives more heavily than false positives during training by assigning class-specific weights to the loss function, effectively forcing the optimizer to allocate more of its capacity to correct classification of the minority class (Elkan, 2001). Ensemble methods such as Random Forest and gradient boosting further mitigate imbalance effects by aggregating predictions from multiple learners trained on different data subsets, reducing variance and improving minority-class recall without requiring explicit resampling (Teixeira et al., 2024).

6. Neural Network Architectures for Fraud Detection

The architectures reviewed in this section are evaluated not merely as a catalogue of technical options but within a comparative framework organized around five operationally relevant criteria: **(1) predictive performance** on standard fraud benchmarks (AUC, F1); **(2) inference latency** and compatibility with sub-100-millisecond authorization pipelines; **(3) interpretability** and compliance with GDPR/LGPD explainability requirements; **(4) computational cost** in terms of memory footprint and scalability under production throughput; and **(5) fraud-specific adequacy**, meaning the structural alignment between the architecture's representational capabilities and the dominant fraud patterns it is most suited to detect. No single architecture dominates across all five criteria: MLPs excel on latency and cost but sacrifice relational and sequential expressiveness; GNNs capture fraud ring structures invisible to transaction-level models but impose significant inference

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

overhead; Transformers offer expressive global attention but face quadratic scaling constraints; hybrid ensembles achieve the strongest benchmark performance but at the cost of deployment complexity. The subsections below develop these trade-offs in detail for each architecture, and Table 2 (Section 6.9) provides a consolidated quantitative comparison.

6.1 Multilayer Perceptron (MLP)

The Multilayer Perceptron is the most foundational deep learning architecture applied to fraud detection on tabular transaction data. Its architecture — comprising fully connected input, hidden, and output layers with non-linear activations — enables the modeling of complex feature interactions without presupposing any particular data structure, making it a natural fit for the heterogeneous, high-dimensional feature vectors that characterize payment transactions (Goodfellow et al., 2016). Its principal operational advantages are simplicity and inference speed: MLP models require minimal preprocessing, have compact memory footprints, and can produce classification scores within microseconds, meeting the latency requirements of real-time authorization pipelines (Kasasbeh et al., 2022). Empirical benchmarks on the ULB Credit Card dataset report MLP accuracy of 97.84% and an F1-score of 85.1% for the fraud class, reflecting competitive minority-class detection when combined with appropriate resampling (Kasasbeh et al., 2022). The principal limitation of MLPs is their inability to model sequential dependencies: each transaction is processed as an independent feature vector, and inter-transaction temporal patterns — critical for detecting burst attacks and account takeovers — are structurally invisible to the model (Ngo et al., 2025).

6.2 Recurrent Neural Networks: LSTM and GRU

Recurrent Neural Networks (RNNs) and their gated variants address the fundamental limitation of MLPs by processing transactions as ordered sequences, enabling the model to represent the evolving behavioral state of a cardholder over time. Long Short-Term Memory networks (LSTMs), introduced by Hochreiter and Schmidhuber (1997), overcome the vanishing gradient problem that prevents standard RNNs from retaining information across long sequences through a gating mechanism — comprising input, forget, and output gates — that regulates the flow of information into and out of a persistent memory cell (Benchaji et al., 2021). In fraud detection, LSTMs have demonstrated particular effectiveness at capturing long-range behavioral dependencies: by modeling the typical spending patterns of individual cardholders across historical transaction sequences, the model can identify single anomalous transactions that deviate from an established norm even when isolated features appear innocuous (Benchaji et al., 2021). The Gated Recurrent Unit (GRU) simplifies the LSTM architecture by merging the input and forget gates into a single

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

update gate and eliminating the separate memory cell, reducing parameter count and computational overhead; in fraud detection benchmarks, LSTM- and GRU-based models achieve comparable accuracy (~95.8–95.9%) and precision (~89.5%), with GRU offering lower inference latency that is advantageous for real-time deployment (Alarfaj et al., 2022). Both architectures, however, process sequences serially rather than in parallel, which imposes latency costs that scale with sequence length and constitute a practical constraint in high-throughput transaction environments (Khanh et al., 2024).

6.3 Convolutional Neural Networks (CNN)

Although CNNs were originally developed for image recognition, their application to fraud detection exploits a structural analogy: a temporal sequence of transactions can be represented as a one-dimensional signal in which local patterns — analogous to visual edges and textures — carry discriminative information. In this framing, a 1D-CNN applies learned convolutional filters across sliding windows of the transaction sequence, detecting local temporal patterns such as rapid successive transactions, atypical spending spikes, or geographic inconsistencies within short time intervals (Ibrahim et al., 2025). CNNs' key advantages in this context are parallel computation — filters are applied simultaneously across all positions in the sequence — and strong local pattern extraction, which complements the global sequential modeling of recurrent networks. Comparative evaluations demonstrate that 1D-CNN models achieve F1-scores of approximately 0.96 on the ULB dataset, outperforming standalone LSTMs on detection of concentrated burst-type fraud patterns while underperforming on fraud schemes requiring long historical context (Ibrahim et al., 2025). The most commonly adopted design in practice combines CNN and LSTM components sequentially — with the CNN extracting local feature representations that are subsequently processed by an LSTM — achieving superior performance to either architecture alone (Ibrahim et al., 2025).

6.4 Autoencoders for Unsupervised Detection

Autoencoders represent the dominant unsupervised paradigm for fraud detection, operating on the principle that a network trained exclusively on legitimate transactions will learn a compact latent representation of normal behavior and exhibit high reconstruction error when presented with anomalous inputs (Hilal et al., 2022). This reconstruction-based anomaly scoring requires no fraud labels during training, making autoencoders particularly valuable in cold-start scenarios where labeled fraud instances are scarce or unavailable. The Variational Autoencoder (VAE), which replaces the deterministic latent encoding of a standard autoencoder with a probabilistic latent distribution, provides a theoretically grounded generative model of normal transaction behavior; anomaly scores derived from reconstruction error and

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

divergence from the learned prior distribution have been shown to be competitive with supervised classifiers on standard benchmarks, while additionally enabling the generation of synthetic fraud examples for data augmentation (Alshameri & Xia, 2024; Goodfellow et al., 2014). Hybrid frameworks that combine VAE-based anomaly scoring with supervised classifiers achieve further improvements: on the European Credit Card and IEEE-CIS datasets, a VAE combined with a Graph Attention Network and XGBoost meta-learner achieved F1-scores above 0.980 and AUC reaching 0.995, outperforming baseline models by up to 15% in F1 (Alazizi et al., 2025). The primary limitation of autoencoders is their susceptibility to sophisticated fraud that closely resembles legitimate behavior in feature space, and their performance degrades significantly when fraudulent patterns fall within the distribution of normal transactions (Hilal et al., 2022).

6.5 Graph Neural Networks (GNN)

Graph Neural Networks constitute perhaps the most structurally apt architecture for financial fraud detection because they operate directly on relational data, treating accounts, devices, merchants, and IP addresses as nodes and transactions as edges in a heterogeneous graph (Liu et al., 2024). This representation makes GNNs uniquely capable of detecting fraud rings — coordinated groups of accounts engaged in collective fraudulent activity — whose identifying signatures are visible only at the network level and are structurally invisible to transaction-level classifiers (Devi et al., 2025). The three predominant GNN architectures applied to financial fraud are Graph Convolutional Networks (GCN), which aggregate features from local neighborhoods through spectral convolutions; Graph Attention Networks (GAT), which weight neighborhood contributions through learned attention coefficients, enabling the model to prioritize the most informative relational signals; and GraphSAGE, which employs inductive neighborhood sampling to generalize to nodes not present during training, a critical property for production systems handling new accounts (Liu et al., 2024; Hamilton et al., 2017). Systematic benchmarks confirm the performance advantage of graph-based methods: GAT-based models achieve recall of 0.89 on financial transaction data, compared to 0.78 for Random Forest and 0.81 for XGBoost (Chen & Guestrin, 2016), with the gap reflecting the value of relational context in identifying fraud that is individually indistinguishable but collectively anomalous (Devi et al., 2025). For anti-money laundering specifically, semi-supervised graph learning approaches — which extrapolate from known laundering patterns to detect novel, subtler schemes in unlabeled data — have demonstrated particular effectiveness given the scarcity of labeled examples in this domain (Cheng et al., 2024).

6.6 Transformers and Attention Mechanisms

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

The Transformer architecture has been adapted to financial fraud detection primarily through its self-attention mechanism, which captures dependencies between all pairs of elements in an input sequence simultaneously and without the serial processing constraints of recurrent architectures. In the tabular transaction domain, TabTransformer (Huang et al., 2020) applies multi-head self-attention to categorical feature embeddings, generating contextually enriched representations that encode inter-feature dependencies; FTTransformer (Gorishniy et al., 2021; Vaswani et al., 2017, NeurIPS) extends this framework by embedding both categorical and continuous features as tokens, enabling attention over all feature types and yielding competitive performance against gradient boosting on several benchmarks (Zavitsanos et al., 2025). The computational mechanics underlying this tokenization process — including how transformer architectures handle tokens as discrete processing units and manage inference costs in foundational models — have been formally characterized in the machine learning literature (Salomão, 2026). For sequential transaction modeling, BERT-style pre-training on unlabeled transaction sequences — using masked transaction prediction as a pretext task — produces representations that capture distributional behavioral regularities and transfer effectively to downstream fraud classification with limited labeled data (Luetto et al., 2024). The central operational trade-off of transformer models in fraud detection is between expressive power and inference latency: the quadratic complexity of standard self-attention with respect to sequence length makes real-time deployment with long historical windows computationally demanding, necessitating approximations such as sparse attention or sequence length constraints that partially sacrifice the global receptive field for which Transformers are valued (Ngo et al., 2025).

6.7 Hybrid Models and Ensembles

The limitations inherent in any single architecture have driven the development of hybrid and ensemble frameworks that combine complementary representational strengths. The most prevalent combination in the fraud detection literature pairs LSTM for temporal sequence modeling with MLP for static transaction feature processing, connecting both streams at a fusion layer to produce a joint classification decision that simultaneously captures the user's behavioral history and the intrinsic properties of the current transaction (Ibrahim et al., 2025). CNN-LSTM hybrids, in which the CNN stage extracts local temporal features that are then fed to an LSTM for long-range sequence modeling, consistently outperform either component alone by approximately 2% across standard metrics (Ibrahim et al., 2025). At the systems level, stacking — using the outputs of base neural network models as features for a meta-learner such as XGBoost or logistic regression — has shown consistent advantages over single-model systems: stacked ensembles achieve F1-scores of

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

0.868 on the ULB Credit Card dataset, outperforming individual models trained on the same data (Rtayli & Enneya, 2024). Mixture-of-Experts (MoE) architectures extend this logic dynamically, routing each transaction to the most appropriate specialized model based on learned gating functions; a MoE framework incorporating LSTM, Transformer, and autoencoder experts demonstrated superior performance across all evaluation metrics relative to classical baselines, combining temporal, attentional, and anomaly-detection capabilities within a single inference pipeline (Vo et al., 2025).

Table 1 synthesizes the performance results reported across the architectures reviewed, enabling direct cross-study comparison of the principal evaluation metrics

Architecture	Dataset	AUC	F1	Precision	Recall	Source
MLP	ULB Credit Card	-	0.851	-	-	Kasasbeh et al. (2022)
LSTM	ULB Credit Card	-	-	~0.895	~0.959	Alarfaj et al. (2022)
GRU	ULB Credit Card	-	-	~0.895	~0.958	Alarfaj et al. (2022)
ID-CNN	ULB Credit Card	-	0.960	-	-	Ibrahim et al. (2025)
VAE + GAT + XGBoost	European CC + IEEE-CIS	0.995	0.980	-	-	Alazizi et al. (2025)
GAT	Financial transactions	-	-	-	0.89	Devi et al. (2025)
Stacked Ensemble	ULB Credit Card	-	0.868	-	-	Rtayli & Enneya (2024)

Table 1. Comparative performance of neural network architectures for financial fraud detection across benchmark datasets. Dashes (—) indicate metrics not reported in the original source.

6.8 Critical Assessment of Benchmark Performance Claims

The high performance metrics reported across the fraud detection literature — with AUC values frequently exceeding 0.99 and F1-scores above 0.98 on standard benchmarks — warrant critical scrutiny before their implications are accepted at face

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

value. Three systematic threats to the validity of these results are identifiable in the reviewed literature.

The first is overfitting to benchmark datasets. Models evaluated exclusively on widely reused datasets such as the ULB Credit Card dataset or IEEE-CIS may achieve artificially high scores by exploiting statistical regularities specific to those datasets rather than learning generalizable fraud detection signals. When the same dataset is used across dozens of studies for both development and evaluation, the collective optimization pressure of the research community functions analogously to repeated testing on a held-out set, inflating reported performance beyond what would be achieved on genuinely unseen data (Semmelrock et al., 2025).

The second threat is data leakage, which occurs when information that would not be available at inference time — such as future transaction outcomes, aggregate account-level statistics computed over the full dataset, or label-derived features — is inadvertently included in the training feature set. In fraud detection, temporal leakage is particularly insidious: if feature engineering incorporates statistics computed over time windows that extend beyond the transaction timestamp, the model effectively has access to future information during training, producing evaluation metrics that cannot be reproduced in production deployment (Board of Governors of the Federal Reserve System, 2025).

The third threat is benchmark bias in the published literature. A well-documented publication bias toward positive results means that studies achieving high AUC and F1 values on benchmark datasets are substantially more likely to be submitted and accepted for publication than studies reporting modest or negative results (Semmelrock et al., 2025). This selective reporting creates a distorted picture of the state of the art: the upper tail of benchmark performance dominates the literature, while the realistic distribution of model performance — including failures, degraded results on proprietary data, and performance drops in production — remains largely invisible. Consumers of this literature should therefore treat reported benchmark metrics as optimistic upper bounds rather than expected production performance, and weight more heavily results validated on held-out institutional data or across multiple independent datasets.

6.9 Computational Cost and Inference Trade-offs Across Architectures

Predictive performance metrics alone are insufficient for evaluating fraud detection architectures in production contexts: inference cost, memory footprint, and scalability under throughput constraints are equally determinative of deployment feasibility. The architectures reviewed in this article exhibit substantially different computational profiles that must be weighed against their performance advantages.

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

MLPs offer the most favorable computational profile for real-time deployment: inference requires a fixed sequence of matrix multiplications with complexity linear in the number of parameters, producing classification scores within microseconds on standard hardware with memory footprints typically below 10MB for fraud-scale architectures (Kasasbeh et al., 2022). LSTMs and GRUs introduce sequential processing overhead — hidden state computation must proceed step-by-step through the transaction sequence — yielding inference latency that scales linearly with sequence length and making them approximately 3–5× slower than equivalent MLPs on sequences of 50–100 transactions (Alarfaj et al., 2022). CNNs recover parallelism through simultaneous filter application across all sequence positions, achieving inference speeds closer to MLPs while retaining local temporal pattern extraction; their memory requirements scale with filter count and depth but remain manageable for 1D fraud detection configurations.

GNNs introduce the most significant computational burden among the architectures reviewed. Neighborhood aggregation operations scale with graph density: in transaction graphs where high-degree merchant or device nodes connect to thousands of accounts, aggregation complexity approaches $O(|E|)$ per layer, where $|E|$ is the number of edges. Multi-hop aggregation — required to detect fraud rings spanning indirect relationships — compounds this cost exponentially with hop depth, making real-time GNN inference on large transaction graphs an active engineering challenge that typically requires graph sampling, node pruning, or approximate aggregation to remain within sub-100-millisecond latency budgets (Liu et al., 2024; Devi et al., 2025).

Transformers present a distinct scalability constraint: standard self-attention computes pairwise interactions between all elements of the input sequence, yielding quadratic complexity $O(n^2)$ in both time and memory with respect to sequence length n. For fraud detection with short transaction windows ($n \leq 32$), this cost is manageable; for behavioral modeling over extended histories ($n > 128$), quadratic scaling necessitates sparse attention approximations or sequence truncation that partially sacrifices the global receptive field that motivates Transformer adoption (Ngo et al., 2025). Table 2 summarizes the comparative inference profiles across architectures.

Architecture	Inference Complexity	Relative Latency	Memory Footprint	Parallelizable	Real-Time Viable
MLP	O(P)	Very Low	Low(<10MB)	Yes	Yes
LSTM / GRU	O(T·P)	Medium	Medium	No	With

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

				(Sequential)	constraints
CNN (1D)	O(T·K·P)	Low	Low–Medium	Yes	Yes
Autoencoder	O(P)	Low	Low–Medium	Yes	Yes
GNN	O(|E|·L)	High	High	Partial	With sampling
Transformer	$O(n^2 \cdot d)$	Medium-High	High	Yes	With sparse attention

T = sequence length; P = parameter count; K = kernel size; |E| = edges; L = layers; n = tokens; d = model dimension.

Table 2. Comparative computational profiles of neural network architectures for real-time fraud detection. Latency classifications are relative and assume standard hardware without GPU acceleration.

7. Real-Time Fraud Detection: Architecture and Infrastructure

7.1 Requirements of a Real-Time Detection System

The deployment of neural network fraud detectors within live payment infrastructures is constrained by four interdependent non-functional requirements that collectively define the feasibility envelope of any production system. Latency is the most operationally binding constraint. It represents the elapsed time between transaction submission and a classification decision. Payment processors require end-to-end scoring within approximately 100 milliseconds. This threshold is fast enough to approve or block a transaction before the customer perceives any delay. Interestingly, the bottleneck for this latency is most frequently memory and network speed, rather than the computational inference itself (Bizarro, 2024; Hilal et al., 2022). Throughput imposes a complementary constraint: modern payment networks process thousands of transactions per second, and fraud detection infrastructure must scale horizontally without proportional latency degradation, with production benchmarks demonstrating that optimized architectures can handle up to 228,000 transactions per second while maintaining 99th-percentile response times below 300 microseconds (Okonkwo et al., 2025). Availability requirements for mission-critical payment systems are expressed in terms of "four-nines" (99.99%) or higher uptime targets, since any system downtime translates directly into unmonitored transaction flows and fraud exposure — a requirement demanding cross-regional replication, automated failover, and fault-tolerant state management (Aerospike, 2025). These constraints interact with the CAP theorem — the formal result stating that a distributed system can

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

guarantee at most two of consistency, availability, and partition tolerance simultaneously (Brewer, 2000) — posing a structural design tension between ensuring that fraud flags propagate instantaneously across all nodes (consistency) and maintaining uninterrupted service during network partitions (availability), a tension typically resolved in fraud systems through eventual consistency designs that prioritize availability while accepting bounded staleness in account state.

7.2 Stream Processing Architecture

The architectural backbone of real-time fraud detection systems is built on event-driven stream processing, in which each transaction triggers an immediate pipeline execution rather than waiting for batch accumulation. Apache Kafka has become the de facto standard for transaction event ingestion in this domain: its distributed, fault-tolerant commit log architecture enables high-throughput message delivery with exactly-once processing semantics, decoupling transaction producers from downstream fraud detection consumers and providing a durable event buffer that supports replayability for model retraining (Dev & Usha, 2025). For stream computation, Apache Flink and Apache Spark Streaming represent the dominant frameworks, each with distinct trade-offs. Flink processes events in true streaming mode — one event at a time with native stateful computation and millisecond-level latency — while Spark Streaming employs a micro-batch model that achieves lower average latency in certain configurations but introduces variable delays that make it less suitable for hard latency targets (Ramachandran et al., 2025). Comparative benchmarks confirm that Flink achieves lower tail latencies under high throughput, making it preferable for transaction authorization flows, while Spark's richer ML library ecosystem makes it well-suited for feature engineering pipelines (Ramachandran et al., 2025). Regarding architectural patterns, the Lambda architecture — which maintains parallel batch and streaming pipelines — has been largely superseded in fraud detection deployments by the Kappa architecture, which unifies processing in a single streaming pipeline powered by Kafka, eliminating the synchronization overhead and code duplication inherent in dual-pipeline designs (Dev & Usha, 2025).

7.3 Low-Latency Model Serving

Deploying trained neural network models within the millisecond-range latency budget of payment authorization requires a dedicated serving infrastructure that abstracts the model from the transaction processing pipeline through a low-overhead communication protocol. REST APIs provide simplicity and broad compatibility but introduce serialization overhead that can be prohibitive at scale; gRPC, which uses Protocol Buffers for binary serialization and HTTP/2 multiplexing, reduces this overhead significantly and is increasingly preferred for high-frequency inference endpoints in financial systems (Okonkwo et al., 2025). At the model execution layer,

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

ONNX Runtime provides a cross-framework inference engine that accepts models exported from PyTorch, TensorFlow, and other training frameworks in a standardized format, performing graph optimizations — including operator fusion, constant folding, and quantization — that consistently reduce inference latency relative to native framework runtimes (Bayousef & Johansson, 2025). For GPU-accelerated deployments, NVIDIA TensorRT compiles the model into a hardware-specific execution plan, applying layer fusion, kernel auto-tuning, and precision reduction to INT8 or FP16 representations; benchmarks demonstrate that TensorRT quantization can produce up to a tenfold improvement in inference speed compared to native PyTorch execution on equivalent hardware (Zhou et al., 2023). Model quantization and pruning — techniques that reduce the numerical precision of model weights and eliminate redundant connections — provide complementary latency reductions deployable without GPU infrastructure, enabling lightweight neural networks to meet sub-100ms scoring requirements even on CPU-only serving nodes (Bayousef & Johansson, 2025).

7.4 Real-Time Feature Store

A recurring challenge in production fraud detection systems is training-serving skew — the divergence between the feature representations used during model training and those available at inference time — which arises when offline batch-computed features are inconsistently reproduced in online serving pipelines and systematically degrades model performance in production (Feast, 2023; Tecton, 2023). Feature stores address this challenge by providing a unified infrastructure layer that enforces consistent feature definitions across training and serving environments. The dominant open-source option, Feast, separates offline storage for batch training from an online store — typically backed by Redis or Apache Cassandra — that serves pre-aggregated feature values for low-latency inference; Feast with a Redis backend achieves sub-millisecond feature retrieval latency, meeting the requirements of real-time fraud scoring (Redis, 2025). Tecton, the principal managed commercial alternative, additionally supports streaming feature transformations — computing features from live Kafka streams with freshness guarantees measured in seconds rather than hours — which is essential for fraud detection use cases where behavioral features such as transaction velocity must reflect the cardholder's most recent activity (Redis, 2025). The distinction between online and offline feature computation is architecturally critical: offline features are pre-computed over historical data and materialized into the online store on a scheduled basis, while on-demand real-time features — such as the ratio between the current transaction amount and the user's recent average — must be computed at inference time from the incoming event, requiring the feature store to support hybrid computation pipelines (Feast, 2023).

7.5 Real-Time Decision and Action

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

The output of the neural network inference stage — a continuous fraud probability score — does not directly constitute an operational decision but must be translated into one of several possible actions through a risk decisioning engine. In production payment systems, this translation typically follows a tiered logic: transactions with scores below a low-risk threshold are approved without friction; those above a high-risk threshold are blocked outright; and those falling in an intermediate range trigger step-up authentication challenges such as one-time passwords or biometric verification (Ileberi et al., 2024). In practice, neural network scorers operate alongside rule-based systems rather than replacing them: rules provide computationally cheap first-pass filtering that eliminates obvious fraud patterns — such as known compromised card numbers or IP addresses on sanction lists — reserving expensive neural inference for transactions that pass these preliminary checks, a layered defense architecture that optimizes the overall cost-performance trade-off (Dev & Usha, 2025). A critical but operationally underinvested component of real-time systems is the feedback loop connecting decision outcomes to model retraining. When a flagged transaction is subsequently confirmed as fraud — through chargeback processing, customer dispute resolution, or investigator review — this outcome label should be captured and used to update the training corpus for future model iterations (Ngo et al., 2025). Without this feedback mechanism, models trained on historical data experience distributional drift as fraud patterns evolve, progressively degrading their performance over time; online learning architectures that continuously update model parameters from streaming labeled outcomes represent an active area of research for addressing this limitation (Ngo et al., 2025).

8. Critical Production Challenges

8.1 Data Drift and Concept Drift

The temporal fragility of fraud detection models in production environments constitutes one of the most persistent operational challenges in the field. Two phenomena degrade model performance over time: data drift and concept drift. Data drift is a significant change in the distribution of input features between training and deployment. Concept drift occurs when the relationship between input features and the target label changes. Both can degrade performance independently or together, which is critical because fraud patterns evolve continuously in response to detection pressures (Ngo et al., 2025). In fraud detection specifically, concept drift is particularly acute: fraudsters adapt their behavioral signatures in direct response to detection signals, rendering models trained even months earlier increasingly unreliable. Detection of such drift is accomplished through distributional distance metrics; the Population Stability Index (PSI) is the most widely deployed in financial services,

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

measuring the shift between a reference and current distribution across discretized bins, while Kullback–Leibler (KL) divergence quantifies the information-theoretic distance between the two distributions and is particularly sensitive to large-magnitude distributional changes (Kodakandla, 2024). In a documented production case, PSI-based daily monitoring detected data drift attributable to shifts in user transaction behavior, and the deployment of an adaptive retraining strategy reduced false positives by 30% without degrading recall (Kodakandla, 2024). Retraining strategies exist on a spectrum from periodic retraining on a fixed schedule, through drift-triggered retraining initiated when PSI or performance metrics exceed predefined thresholds, to continuous online learning in which model parameters are updated incrementally from each new labeled observation — an approach that offers the fastest adaptation but introduces risks of catastrophic forgetting and instability under adversarial inputs (Hinder et al., 2024).

8.2 Adversarial Attacks on Fraud Detection Systems

The deployment of machine learning fraud detectors in high-stakes financial environments makes them targets for deliberate adversarial manipulation. Evasion attacks — the dominant adversarial threat in fraud detection — involve crafting transaction inputs in which minimal, carefully optimized perturbations cause the model to misclassify a fraudulent transaction as legitimate, without the manipulation being detectable through standard monitoring (Carminati et al., 2020). The severity of this threat is empirically established: conventionally trained GNN fraud detectors have been shown to be compromised by Projected Gradient Descent (PGD) evasion attacks with Attack Success Rates (ASR) reaching 87.5%, meaning the majority of adversarially crafted fraud instances successfully evade detection (Creswell et al., 2024). The practical challenge of adversarial robustness in fraud is compounded by the constrained nature of the domain: unlike image perturbations, modifications to financial transaction features must remain financially plausible and semantically consistent — a fraudster cannot simply add arbitrary noise to a transaction amount — which constrains the attack space but also limits certain generic defenses (Carminati et al., 2020). The primary defense paradigm is adversarial training, in which adversarially generated examples are incorporated into the training data, exposing the model to the attack manifold during learning; in the financial domain, PGD-based adversarial training has been demonstrated to reduce ASR from 87.5% to 32.0% while simultaneously delivering a 52.3% reduction in expected annual fraud loss (Creswell et al., 2024). Ensemble diversity — ensuring that constituent models make uncorrelated errors — provides a complementary defense, as attacks optimized against one model are less likely to transfer uniformly to others (Rigaki & Garcia, 2023).

8.3 Explainability and Regulatory Compliance

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

The regulatory environment governing automated financial decision-making has intensified demands for model interpretability that are structurally in tension with the opacity of high-performance neural architectures. The GDPR's Article 22 establishes an individual's right to explanation for automated decisions that significantly affect them, while the EU AI Act classifies financial fraud detection systems as high-risk AI applications subject to transparency, documentation, and human oversight requirements; Brazil's LGPD establishes analogous obligations under Articles 20 and 21 (Corrêa et al., 2024). Basel III's model risk management guidelines additionally require that financial institutions be able to demonstrate the soundness and interpretability of risk models to regulators and auditors (Černevičienė & Kabašinskas, 2024). The two most widely adopted post-hoc explainability methods in production fraud systems are SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations). SHAP assigns feature attribution values derived from cooperative game theory that sum to the model's output, providing both globally consistent feature importance rankings and instance-level explanations (Lundberg & Lee, 2017); comparative evaluations demonstrate that SHAP produces more stable and auditably consistent explanations than LIME and is preferable for regulatory documentation and forensic reporting (Raufi et al., 2024; Gimeno-Blanes et al., 2025). LIME constructs a local linear approximation of the model's decision boundary around individual predictions, offering faster computation but lower stability across perturbation samples, making it better suited for operational analyst workflows than for compliance submissions (Raufi et al., 2024). The fundamental tension between interpretability and performance remains partially unresolved: in benchmark comparisons, a hybrid framework combining a stacking ensemble with SHAP-guided feature selection achieved 98.3% accuracy while maintaining regulatory explainability, but the addition of explainability constraints generally imposes a measurable performance cost on the underlying classifier (Obeng et al., 2025).

8.4 False Positives and Customer Experience

The false positive problem in fraud detection carries significant and frequently underestimated business costs beyond its direct operational burden. Every legitimate transaction blocked or subjected to unnecessary friction imposes measurable damage to customer relationships: industry data indicate that 40% of consumers who experience a false positive will reduce their use of the affected payment method, and 11% will abandon the issuer entirely (Javelin Strategy & Research, 2024). The structural challenge is that false positive rate and fraud recall are inversely related through the decision threshold: any threshold reduction that captures more fraud necessarily increases legitimate transaction friction, creating an optimization problem with asymmetric and institution-specific cost structures (Abdallah et al., 2022). Strategies for reducing the false positive rate without proportional sacrifice in recall

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

operate on several axes. Behavioral profiling — maintaining longitudinal models of individual cardholder spending patterns — enables the system to contextualize transactions against the specific user's baseline rather than a population average, substantially reducing anomaly flags on legitimate but unusual transactions (Ileberi et al., 2024). Threshold personalization, in which the classification threshold is calibrated per customer segment or channel based on historical false positive rates, risk tolerance, and transaction value, allows institutions to optimize the fraud-friction trade-off at a granular level rather than applying a single institution-wide threshold (Hilal et al., 2022). Contextual risk scoring that integrates real-time signals — device fingerprint, geolocation, time of day, and session behavior — further increases specificity by providing the model with situational context that distinguishes legitimate anomalies from fraudulent ones (Ngo et al., 2025).

8.5 Data Privacy and Security in ML Pipelines

The sensitivity of financial transaction data imposes structural constraints on the training of fraud detection models that extend beyond model design to the architecture of the entire ML pipeline. Federated Learning (FL) has emerged as the primary technical paradigm for addressing the cross-institutional data sharing problem: rather than aggregating raw transaction data in a central repository, FL allows each financial institution to train on its local data, sharing only model weight updates with a central aggregation server, such that no raw transaction records ever leave the originating institution (Zhu et al., 2024). Cross-silo federated learning — in which the clients are institutions rather than individual devices — is particularly well-suited to the financial sector, where each institution holds large, stable datasets and operates under strong regulatory compliance requirements derived from GDPR, LGPD, and sector-specific frameworks (Emmanuel, 2025). However, FL alone does not eliminate privacy risk, as model gradient updates have been demonstrated to be vulnerable to gradient inversion attacks that can partially reconstruct training data (Zhu et al., 2024). Differential Privacy (DP) addresses this residual risk by injecting calibrated noise into gradient updates during training — typically via the DP-SGD algorithm — providing a formal mathematical guarantee that the probability of inferring any individual data point from the model is bounded by the privacy parameter ε (Zhu et al., 2024). Empirical evaluations of DP-enabled FL for financial fraud detection confirm that this combination maintains competitive detection performance: at a privacy budget of $\varepsilon = 5.74$, DP-FL achieves accuracy within 3% of standard FL while providing substantially stronger privacy guarantees (Khursheed et al., 2025). The principal remaining challenge is the accuracy-privacy trade-off: as the privacy budget tightens, the noise magnitude required for formal privacy guarantees increasingly degrades model utility, and finding the optimal operating point for a given regulatory context and performance requirement remains an active research problem (Zhu et al., 2024).

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

9. Use Cases and Real-World Implementations

9.1 Credit Card Fraud Detection

The most extensively documented industrial implementations of neural network fraud detection are those of the major card networks. Mastercard's Decision Intelligence platform scores approximately 143 billion transactions annually through a real-time machine learning pipeline that evaluates the probability of fraud for every authorization request submitted by member banks (Mastercard, 2024). The system's architecture combines generative AI with graph technology, using each cardholder's historical merchant visit sequence as a behavioral prompt to score the plausibility of the current transaction against established patterns. The reported outcomes are operationally significant: the AI-enhanced system detects three times the volume of fraudulent transactions compared to prior approaches, while simultaneously reducing false positives tenfold — outcomes that translate into documented reductions in fraud losses for issuers and measurably fewer declined legitimate transactions for consumers (Amazon Web Services, 2024). A subsequent enhancement incorporating generative AI graph scanning further doubled the detection rate for compromised cards and reduced false positive alerts by up to 200%, while accelerating identification of at-risk merchants by 300% (Mastercard, 2024). Visa has pursued a parallel trajectory, investing $100 million in a generative AI venture fund and deploying proprietary real-time scoring systems targeting, among other attack vectors, enumeration attacks — automated bot operations that test stolen card credential variants at scale — which the company attributes approximately $1.1 billion annually in industry-wide fraud losses (PYMNTS, 2024).

9.2 PIX Fraud and Open Banking in Brazil

Brazil represents one of the most challenging and analytically instructive environments for studying real-time fraud detection in instant payment infrastructure. The PIX system, launched by the Banco Central do Brasil (BCB) in November 2020, processes over 42 billion transactions annually and has achieved near-universal adoption among Brazil's banked population (QED Investors, 2025). Its defining operational characteristic — settlement under ten seconds, 24/7 — is simultaneously its principal security vulnerability: once a fraudulent PIX transfer is executed, funds are dispersed across mule account chains in under nine minutes in more than half of reported cases, severely constraining the window for interception (W Fintechs, 2025). In response to escalating fraud losses, the BCB has implemented a layered institutional framework: the Special Refund Mechanism (MED) enables victims to request fund blocks and recovery across up to five levels of chained transfers under

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

the MED 2.0 update effective February 2026; Resolution No. 6 mandates real-time sharing of fraud intelligence among all Pix participants; and Normative Instruction No. 491 imposes device registration and behavioral controls specifically targeting first-time and unregistered device access (Banco Central do Brasil, 2025; Feedzai, 2024). Interviews with major Brazilian banks — including Banco do Brasil, Sicredi, and Banrisul — confirm that production fraud prevention strategies combine multifactor authentication with AI-driven behavioral detection models, with institutions uniformly citing the speed of criminal operations as the primary bottleneck limiting fraud recovery (Bertoldi et al., 2025). Digital-native institutions such as Nubank — with over 105 million users — face the additional challenge of operating anti-fraud systems at consumer bank scale without legacy infrastructure, making them early adopters of stream processing architectures and online learning frameworks (QED Investors, 2025).

9.3 Money Laundering Detection with GNNs

Anti-money laundering (AML) represents the domain in which Graph Neural Networks have demonstrated the most pronounced performance advantage over preceding methodologies, owing to the fundamentally relational structure of money laundering operations. Money laundering schemes are defined by layered transactions designed to obscure the provenance of illicit funds, producing graph-level signatures — such as fan-out patterns from source accounts, circular transfer chains, and intermediary accounts with atypically high in-degree — that are structurally invisible to transaction-level classifiers but detectable through neighborhood aggregation in GNN architectures (Cheng et al., 2024). A landmark study applying heterogeneous GNNs to real-world bank transaction data from DNB, Norway's largest bank, demonstrated that GNN-based detection substantially outperforms traditional rule-based AML systems, which struggle to capture the complex relational dependencies among customers, accounts, and transactions that characterize laundering operations (van Essen et al., 2025). At the institutional level, GNN-based AML systems have achieved documented superiority over prior methods: the FlowScope algorithm, which models transaction networks as directed graphs and tracks fund flows from source to destination, detected money laundering patterns that evaded rule-based monitoring by targeting the layering phase — the most sophisticated stage of the laundering process — where large deposits, internal transfers, and sequential withdrawals are hardest to flag individually (Cheng et al., 2024). Continual graph learning approaches, which update GNN parameters incrementally as new laundering tactics emerge without catastrophic forgetting of prior patterns, represent the current research frontier for production-grade AML (Deprez et al., 2025).

9.4 Fraud Rings and Coordinated Fraud

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Fraud rings — organized groups of accounts, devices, and identities acting in coordinated fashion to amplify fraud volume — represent a category of attack that exposes the fundamental limitation of transaction-level models and demands graph-aware detection. Their defining characteristic is that no single transaction or account in the ring is necessarily anomalous in isolation; the signal emerges only from the structural properties of the network of interactions (Devi et al., 2025). The standard detection pipeline for fraud rings combines community detection — typically using algorithms such as Louvain clustering or spectral methods to identify tightly connected account subgraphs — with GNN-based classification that assigns fraud scores to individual nodes conditioned on the properties of their detected community (Liu et al., 2024). In e-payment environments, this approach has demonstrated recall rates of 0.89 for coordinated fraud, compared to 0.78 for Random Forest operating on the same data, with the performance gap reflecting the value of relational context for identifying accounts whose individual behavior is legitimate but whose network position is anomalous (Devi et al., 2025). In the PIX ecosystem specifically, the BCB's MED 2.0 infrastructure incorporates graph-based chained blocking — tracing and freezing funds across up to five account layers — which operationalizes, at a regulatory infrastructure level, the same graph traversal logic that underlies academic GNN fraud ring detectors (W Fintechs, 2025).

9.5 Insurance and Credit Fraud

Neural network applications in insurance and credit fraud detection have matured substantially in recent years, with deep learning methods consistently outperforming classical classifiers across both domains. In auto insurance, a CNN-LSTM hybrid model trained on standard insurance claim datasets achieved 89.6% accuracy and 90.7% precision, demonstrating that sequential feature extraction is beneficial even in claim-level data where temporal patterns encode sequences of reported events, geographic movements, and repair shop interactions (Ming et al., 2024). For healthcare insurance — where fraudulent billing schemes involve collusive relationships among patients, providers, diagnoses, and services — GNN architectures have emerged as the dominant methodological paradigm, as the relational structure of healthcare fraud maps naturally onto a heterogeneous graph in which each entity type becomes a distinct node class (du Preez et al., 2024). Heterogeneous GNN architectures such as HINormer and HybridGNN have demonstrated measurable performance advantages over tabular models including gradient boosting and MLP on Medicare fraud detection tasks, with the relational learning capacity of GNNs capturing provider-patient collusion rings that are structurally invisible to record-level classifiers (Diamanti et al., 2025). In the credit domain, deep learning models for loan application fraud — where fraudsters exploit synthetic identity construction and stolen personal information — have achieved F1-

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

scores exceeding 0.98 on benchmark datasets, with transformer-based models showing particular strength in modeling the sequential application behavior that distinguishes genuine from fraudulent credit seekers (Ngo et al., 2025). Across both sectors, the combination of supervised deep learning with graph-based relational modeling represents the current performance frontier, while regulatory demands for explainability introduce persistent tension between model complexity and deployment compliance (du Preez et al., 2024).

10. Research Frontiers and Future Directions

10.1 Federated Learning for Cross-Institutional Collaboration

The most structurally significant limitation of current fraud detection models — their confinement to the transaction data of a single institution — is increasingly addressed by Federated Learning (FL), which enables multiple competing financial institutions to train collaborative models without any raw customer data leaving individual systems (Emmanuel, 2025). In the cross-silo FL paradigm, each bank computes model gradient updates locally and transmits only these updates to a central aggregation server, which combines them through algorithms such as FedAvg or FedProx into a shared global model (Gad et al., 2025). Empirical results confirm the performance advantage of this approach: a federated graph learning framework combining a Convolutional Feedforward Neural Network and Generative Adversarial Network achieved recall improvements of 11.87% to 33.9% over the non-federated baseline on the IEEE-CIS dataset, with AUC gains exceeding 3%. A federated model trained with SMOTE integration maintained accuracy of 99% for 5–50 participating clients with AUC reaching 1.00, results that are competitive with centralized approaches while preserving strict data confidentiality (Alhamad et al., 2025). The principal implementation challenges are heterogeneous data distributions — each institution's transaction data reflects distinct customer demographics and product mixes — and the risk of gradient inversion attacks, which can partially reconstruct training data from shared updates, motivating the integration of Differential Privacy mechanisms as a complementary safeguard (Zhu et al., 2024).

10.2 Online Learning and Continuous Adaptation

A fundamental limitation of periodically retrained batch models is that their parameters are frozen between retraining cycles, rendering them blind to emerging fraud patterns until the next update. Online learning — in which model parameters are updated incrementally after each new labeled observation — directly addresses this temporal rigidity by enabling continuous adaptation to shifting data distributions

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

(Hinder et al., 2024). For streaming fraud detection specifically, an autonomous learning framework leveraging incremental learning, online clustering, and self-evolving models demonstrated superior performance to static batch models under embedded drift events by recalibrating feature weights and adjusting thresholds in real time, simulated across batches of 10,000 transactions (Ogundipe, 2025). The primary open-source frameworks supporting production online learning are River — a Python library merging the former creme and scikit-multiflow projects, offering drift detection, anomaly detection, and classification algorithms designed for one-sample-at-a-time processing — and Vowpal Wabbit, developed under co-sponsorship of Microsoft Research and Yahoo Research, which employs the hashing trick for memory-efficient feature encoding and supports online versions of classification, regression, and reinforcement learning (Halford, 2022). The principal theoretical risk of online learning in adversarial domains is catastrophic forgetting — the progressive overwriting of knowledge about historical fraud patterns as parameters adapt to recent data — which requires mitigation through experience replay, selective parameter regularization, or ensemble strategies that preserve model diversity across temporal windows (Hinder et al., 2024).

10.3 Large Language Models in Fraud Detection

Large Language Models (LLMs) are emerging as complementary instruments in fraud detection, contributing capabilities that are structurally orthogonal to those of tabular neural classifiers. Their primary value lies in processing unstructured textual content: LLMs deployed for adverse media screening in Know Your Customer (KYC) and Anti-Money Laundering (AML) workflows can analyze news articles, regulatory filings, and social media at scale, disambiguating entities with common names and ranking the relevance of flagged content with a contextual sophistication unavailable to keyword-based systems (Fenergo, 2024). In fraud classification on tabular transaction data, LLMs demonstrate competitive performance specifically in low-data regimes: by serializing transaction records as natural language prompts and leveraging zero-shot and few-shot in-context learning, LLMs can achieve meaningful classification accuracy without the large labeled datasets required by supervised deep learning models, a property particularly valuable for novel fraud typologies where labels are initially scarce (Luetto et al., 2024). A Retrieval-Augmented Generation (RAG) architecture — in which an LLM queries a dynamically updated policy knowledge base to contextualize each transaction against current fraud typologies and institutional rules — enables fraud systems to incorporate new threat intelligence without model retraining, addressing the temporal rigidity of static classifiers (Ravi et al., 2025). The emerging paradigm of LLM-as-orchestrator takes this further, positioning an LLM as a reasoning engine that coordinates multiple specialized

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

detection agents — graph anomaly detectors, sequence classifiers, rule engines — directing their outputs toward a coherent fraud investigation workflow (Taktile, 2024).

10.4 Generative AI Weaponized by Fraudsters — and Countermeasures

The democratization of generative AI has fundamentally altered the threat landscape by placing previously expert-level forgery capabilities into the hands of low-skill actors. Deepfake-related fraud attempts in fintech increased by 700% in 2023 alone, and projected AI-enabled fraud losses in the United States are estimated to reach $40 billion annually by 2027, up from $12.3 billion in 2023 (Deloitte, 2024). The most consequential attack vector in KYC systems is not video deepfake presentation but injection attacks — in which fraudsters intercept the API call between an application and the KYC verification provider and substitute a synthetic biometric stream, bypassing liveness detection entirely because the attack never engages the camera (Reality Defender, 2025). Synthetic identity fraud, in which entirely fabricated personas are constructed from combinations of real and AI-generated data, accounted for $3.1 billion in U.S. lender losses in 2023 and is growing at over 20% annually (Transunion, 2024). Countermeasures constitute an active and rapidly evolving research frontier: cross-modal detection systems that correlate voice and facial biometrics simultaneously are more robust to unimodal spoofing; the European technical specification CEN/TS 18099 and forthcoming ISO 25456 standard are formalizing Injection-Attack Detection (IAD) requirements alongside traditional Presentation Attack Detection (PAD); and multi-layered frameworks combining biometric checks, behavioral monitoring, document metadata forensics, and adaptive risk-based authentication represent the current institutional consensus on defensible KYC architecture (Reality Defender, 2025; Papakostas et al., 2025).

10.5 Quantum Machine Learning in Fraud Detection

Quantum Machine Learning (QML) represents the most speculative but theoretically significant emerging frontier in fraud detection research. The core theoretical proposition is that quantum algorithms can access exponentially larger feature spaces than classical counterparts through superposition and entanglement, potentially enabling the detection of fraud patterns that are computationally intractable for classical neural networks operating on high-dimensional transaction data (Corli et al., 2025). The Quantum Support Vector Machine (QSVM) is the most empirically validated QML algorithm in the financial fraud domain: a landmark study applying QSVM to real card payment data using IBM Quantum hardware via the Qiskit software stack demonstrated that a hybrid classical-quantum ensemble incorporating QSVM-derived features outperformed both pure classical and pure quantum approaches, with QSVM providing a complementary exploration of the feature space that improved ensemble accuracy (Grossi et al., 2022). A comparative evaluation of

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

QML models on financial transaction data found that QSVM achieved an F1-score of 0.91 and precision of 0.96, competitive with the best classical LSTM model at F1 0.94, while Quantum Neural Networks exhibited significant training instability attributable to the barren plateau optimization problem on current Noisy Intermediate-Scale Quantum (NISQ) hardware (Farahmand, 2025). The critical constraint on practical deployment is hardware maturity: current quantum processors are limited to tens to hundreds of noise-prone qubits, necessitating drastic dimensionality reduction of transaction feature spaces that sacrifices the very representational richness that motivates the quantum approach (Corli et al., 2025). Realistic assessments place fault-tolerant quantum hardware capable of providing genuine computational advantage over classical deep learning in fraud detection on a horizon of ten to fifteen years, positioning current QML research in fraud as primarily foundational — establishing theoretical frameworks, algorithmic primitives, and hybrid classical-quantum pipelines that will be deployable as hardware matures (Corli et al., 2025).

11. Conclusion

The evidence surveyed across this review establishes neural networks not merely as one tool among many but as the structural foundation of modern financial fraud detection. The performance ceiling of rule-based and classical statistical systems has been empirically and repeatedly demonstrated, and the complexity, volume, and adaptability of contemporary fraud operations demand detection architectures capable of modeling non-linear, high-dimensional, and relational patterns at scale. GNNs, LSTM-based sequence models, hybrid ensembles, and Transformer-based frameworks have each demonstrated measurable and substantial advantages over their predecessors in controlled benchmarks and documented production deployments alike (Ngo et al., 2025; Hilal et al., 2022).

Yet predictive accuracy, however high, is a necessary but insufficient condition for effective fraud prevention. The temporal constraint is non-negotiable: a correct fraud classification delivered after fund disbursement has the practical value of a post-mortem — the damage is done. Real-time detection within sub-100-millisecond authorization windows, sustained at the throughput of modern payment networks with four-nines availability, defines the operational frontier that separates academically interesting models from deployable systems (Bizarro, 2024; Lu et al., 2022). Meeting this frontier simultaneously with the demands of regulatory explainability under GDPR and LGPD, algorithmic fairness toward vulnerable consumer groups, data privacy preservation through Federated Learning and Differential Privacy, and minimal false positive friction for legitimate customers constitutes the defining multidimensional

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

optimization problem of the field (Černevičienė & Kabašinskas, 2024; Corrêa et al., 2024).

The arms race dimension of fraud detection imposes a further constraint that no static model can satisfy: fraudsters adapt continuously, deploying adversarial evasion techniques, generative AI-fabricated identities, and deepfake-enabled KYC bypass at an accelerating rate (Papakostas et al., 2025; Deloitte, 2024). This dynamic renders any deployed model a depreciating asset. Robust MLOps infrastructure — encompassing drift monitoring, triggered retraining, online learning pipelines, and feedback loops that recapture labeled outcomes from resolved disputes — is therefore not an operational convenience but a strategic necessity for maintaining detection efficacy over time (Kodakandla, 2024; Hinder et al., 2024). Priority research directions include the maturation of cross-institutional Federated Learning, LLM-orchestrated multi-agent detection frameworks, and the continued development of adversarially robust GNN architectures, all of which address dimensions of the problem that current production systems handle inadequately. The longer-term integration of Quantum Machine Learning awaits hardware maturity but merits continued foundational investment. Ultimately, however, technology is a necessary but not sufficient condition for systemic fraud resilience. The most sophisticated detection architecture can be undermined by inadequate governance structures, siloed institutional cultures that resist fraud intelligence sharing, regulatory frameworks that lag behind threat evolution, and insufficient investment in the human expertise required to interpret, audit, and override model decisions. The algorithm must be embedded within a broader organizational and institutional context in which governance, accountability, and adaptive capacity are treated as coequal imperatives to technical performance.

References

Abdallah, A., Maarof, M. A., & Zainal, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. Applied Sciences, 12(19), 9637. https://doi.org/10.3390/app12199637

Abdelghafour, E. B., Mohamed, C., Noura, A., & Abdelhamid, B. (2024). Enhancing credit card fraud detection using a stacking model approach and hyperparameter optimization. International Journal of Advanced Computer Science and Applications, 15(10). https://doi.org/10.14569/IJACSA.2024.01510110

Aerospike. (2025). Real-time fraud detection for payments. https://aerospike.com/blog/real-time-fraud-detection/

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access, 10, 39700–39715. https://doi.org/10.1109/ACCESS.2022.3166891

Alazizi, A., Habrard, A., & Jacquenet, F. (2025). Dual sequential variational autoencoders for fraud detection. In Advances in Intelligent Data Analysis (pp. 14–26). Springer. https://doi.org/10.1007/978-3-030-44584-3_2

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 40, 100402. https://doi.org/10.1016/j.cosrev.2021.100402

Alhamad, R., Alfawareh, H., & Drabasa, S. (2025). Federated learning for credit card fraud detection: A privacy-preserving approach with SMOTE optimization. ResearchGate. https://doi.org/10.13140/RG.2.2.24028.72323

Alshameri, F., & Xia, R. (2024). An evaluation of variational autoencoder in credit card anomaly detection. Big Data Mining and Analytics, 7(3), 718–729. https://doi.org/10.26599/BDMA.2023.9020035

Amazon Web Services. (2024). Using AWS AI and ML services to detect and prevent fraud: Mastercard case study. https://aws.amazon.com/solutions/case-studies/mastercard-ai-ml-testimonial/

Banco Central do Brasil. (2025). Pix statistics and regulatory framework. https://www.bcb.gov.br/en/financialstability/pixstatistics

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, 6(1), 20–29. https://doi.org/10.1145/1007730.1007735

Bayousef, M., & Johansson, M. (2025). Performance and energy efficiency benchmarking of deep learning inference frameworks. Electronics, 14(2), 336. https://doi.org/10.3390/electronics14020336

Benchaji, I., Douzi, S., & El Ouahidi, B. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. Journal of Big Data, 8(1), 151. https://doi.org/10.1186/s40537-021-00541-8

Bertoldi, P., Rodrigues, M., & Souza, L. (2025). A taxonomy of Pix fraud in Brazil: Attack methodologies, AI-driven amplification, and defensive strategies. arXiv. https://arxiv.org/abs/2511.20902

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Bhujade, R. K., & Asthana, S. (2023). An extensive review of ReLU and sigmoid function in multiple hidden layer back propagation neural network model. International Journal of Applied Engineering and Technology, 5(2), 72–80.

Bizarro, P. (2024). Latency in machine learning: What fraud prevention leaders need to know. Feedzai. https://www.feedzai.com/blog/latency-in-machine-learning-what-fraud-prevention-leaders-need-to-know/

Board of Governors of the Federal Reserve System. (2025). A Bayesian simulator for payment card fraud detection (Finance and Economics Discussion Series 2025-017). Federal Reserve. https://doi.org/10.17016/FEDS.2025.017

Brewer, E. A. (2000). Towards robust distributed systems. Proceedings of the 19th Annual ACM Symposium on Principles of Distributed Computing, 7. https://doi.org/10.1145/343477.343502

Carminati, M., Polino, M., Continella, A., Lanzi, A., Maggi, F., & Zanero, S. (2020). Evasion attacks against banking fraud detection systems. Proceedings of the 23rd International Symposium on Research in Attacks, Intrusions and Defenses, 290–303. https://www.usenix.org/system/files/raid20-carminati.pdf

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review, 57, 216. https://doi.org/10.1007/s10462-024-10854-8

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

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

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2024). Graph learning-empowered financial fraud detection: Progress and future directions. Intelligent Computing, 3, 0146. https://doi.org/10.34133/icomputing.0146

Corli, S., Moro, L., Dragoni, D., Macaluso, A., & Prati, E. (2025). Quantum machine learning algorithms for anomaly detection: A review. arXiv. https://arxiv.org/abs/2408.11047

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2024). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns, 4(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Creswell, M., Petrov, D., & Hughes, R. (2024). Adversarial machine learning in finance: Developing resilient AI models to counter fraudster evasion attacks on US bank security systems. International Journal of Computer Applications, 13(12), 111–137.

Dal Pozzolo, A. (2016). Adaptive machine learning for credit card fraud detection [Doctoral dissertation, Université Libre de Bruxelles].

Deloitte Center for Financial Services. (2024). Deepfakes and fraud risk in financial services. Deloitte. https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html

Deprez, B., Vanderschueren, T., Baesens, B., Verdonck, T., & Verbeke, W. (2025). Advances in continual graph learning for anti-money laundering systems: A comprehensive review. arXiv. https://arxiv.org/abs/2503.24259

Dev, R. S., & Usha, J. (2025). Event-driven fraud detection pipeline: Real-time processing with Kafka, ksqlDB & Apache Flink. International Journal of Computer Applications, 187(60), 13–18. https://doi.org/10.5120/ijca2025925872

Devi, R. R., Manoharan, P., & Suresh, A. (2025). Reinforcement learning with graph neural network fusion for real-time financial fraud detection. Scientific Reports, 15, 8201. https://doi.org/10.1038/s41598-025-25200-3

Diamanti, N., Malandri, L., & Seveso, A. (2025). Fraud detection and explanation in medical claims using GNN architectures. Scientific Reports, 15, 16312. https://doi.org/10.1038/s41598-025-22910-6

du Preez, A., Bhattacharya, S., Beling, P., & Bowen, E. (2024). Fraud detection in healthcare claims using machine learning: A systematic review. Artificial Intelligence in Medicine, 103, 103061. https://doi.org/10.1016/j.artmed.2024.103061

Elkan, C. (2001). The foundations of cost-sensitive learning. Proceedings of the 17th International Joint Conference on Artificial Intelligence, 973–978.

Emmanuel, M. (2025). Federated learning for privacy-preserving financial fraud detection. SSRN. https://doi.org/10.2139/ssrn.5399668

JIRRS Journal International Review of Research Studies

Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Farahmand, D. (2025). Comparative analysis of quantum machine learning models and classical deep learning frameworks for anomaly detection in financial transactions. ResearchGate. https://www.researchgate.net/publication/397471761

Feast. (2023). What is a feature store? https://feast.dev/blog/what-is-a-feature-store/

Federal Trade Commission. (2024). Consumer sentinel network data book 2024. FTC. https://www.ftc.gov/reports/consumer-sentinel-network

Feedzai. (2024). BCB Normative No. 491: How Brazil can strengthen Pix fraud prevention. https://www.feedzai.com/blog/bcb-normative-no-491-how-brazil-can-strengthen-pix-fraud-prevention/

Fenergo. (2024). Harness AI's large language models for enhanced anti-fraud defenses. BAI. https://www.bai.org/banking-strategies/harness-ais-large-language-models-for-enhanced-anti-fraud-defenses/

Gad, A., Zaher, M., & Hassan, F. (2025). Enhancing privacy in IoT-enabled digital infrastructure: Evaluating federated learning for intrusion and fraud detection. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC12114921/

Gimeno-Blanes, F. J., Cuenca-Jimenez, P., & Blanes-Selva, V. (2025). Explainable AI for forensic analysis: A comparative study of SHAP and LIME in intrusion detection models. Applied Sciences, 15(13), 7329. https://doi.org/10.3390/app15137329

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34, 18932–18943.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems (NeurIPS), 27.

Grossi, M., Ibrahim, N., Radescu, V., Loredo, R., Voigt, K., Von Altrock, C., & Rudnik, A. (2022). Mixed quantum-classical method for fraud detection with quantum feature selection. IEEE Transactions on Quantum Engineering, 3, 1–12. https://doi.org/10.1109/TQE.2022.3213474

Halford, M. (2022). The future of River. https://maxhalford.github.io/blog/future-of-river/

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems (NeurIPS), 30.

Hayashi, F. (2026). New data on card-present and card-not-present fraud rates in the United States. Federal Reserve Bank of Kansas City Payments System Research Briefing. https://www.kansascityfed.org/research/payments-system-research-briefings/new-data-on-card-present-and-card-not-present-fraud-rates-in-the-united-states/

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. Proceedings of the 2008 IEEE International Joint Conference on Neural Networks, 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Hinder, F., Vaquet, V., & Hammer, B. (2024). One or two things we know about concept drift: A survey on monitoring in evolving environments. Part A: Detecting concept drift. Frontiers in Artificial Intelligence, 7, 1330257. https://doi.org/10.3389/frai.2024.1330257

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hong, X., Zheng, C., Zang, M., Perreault, L., Bensoussane, R., & Zilberman, N. (2024). In-network machine learning for real-time transaction fraud detection. University of Oxford. https://eng.ox.ac.uk/media/tutlpnpf/hong2024mind.pdf

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Howard, J., & others. (2019). IEEE-CIS fraud detection [Dataset]. Kaggle. https://www.kaggle.com/competitions/ieee-fraud-detection

Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). TabTransformer: Tabular data modeling using contextual embeddings. arXiv. https://arxiv.org/abs/2012.06678

Ibrahim, A., Abdelhamid, A., & El-Bastawissy, A. (2025). Analysis and visualization of fraud detection patterns through data mining and classification using MLP and

JIRRS Journal International Review of Research Studies

Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

hybrid deep learning model. Egyptian Informatics Journal, 30, 100591. https://doi.org/10.1016/j.eij.2025.100591

Ileberi, E., Sun, Y., & Wang, Z. (2024). Real-time contextual AI for proactive fraud detection in financial services. Journal of Information Systems Engineering and Management, 9(4). https://doi.org/10.55267/iadt.07.13765

Javelin Strategy & Research. (2024). 2024 identity fraud study: The virtual battleground. Javelin Strategy & Research.

Kasasbeh, B., Aldabaybah, B., & Ahmad, H. (2022). Multilayer perceptron artificial neural networks-based model for credit card fraud detection. Indonesian Journal of Electrical Engineering and Computer Science, 26(1), 362–373. https://doi.org/10.11591/ijeecs.v26.i1.pp362-373

Khanh, T. T. B., Minh Le, T., & Nguyen, T. H. (2024). Deep learning-based financial fraud detection with temporal sequence modeling. Journal of Computer Engineering in Intelligent Management, 3(2). https://doi.org/10.58915/jceim.v3i2.82

Khursheed, M., Azmat, F., & Larijani, H. (2025). Privacy-preserving federated credit risk models: Evaluating differential privacy and homomorphic encryption techniques. Scientific Reports, 15, 3691. https://doi.org/10.1038/s41598-025-34536-9

Kitchenham, B. (2004). Procedures for performing systematic reviews (Technical Report TR/SE-0401). Keele University.

Kodakandla, N. (2024). Effective MLOps strategies for addressing data drift in real-time machine learning systems. International Journal of Science and Research Archive, 12(1), 3127–3139. https://doi.org/10.30574/ijsra.2024.12.1.0724

Liu, Z., Dou, Y., Yu, P. S., Deng, Y., & Peng, H. (2024). Financial fraud detection using graph neural networks: A systematic review. Expert Systems with Applications, 238, 122156. https://doi.org/10.1016/j.eswa.2023.122156

Lopez-Rojas, E. (2016). PaySim: A financial mobile money simulator for fraud detection [Dataset]. Kaggle. https://www.kaggle.com/datasets/ealaxi/paysim1

Lu, M., Chen, J., Cheng, W., Guo, H., & Li, Z. (2022). BRIGHT: Graph neural networks in real-time fraud detection. arXiv. https://arxiv.org/abs/2205.13084

Luetto, C., di Stasi, S., Giudici, P., & Raffinetti, E. (2024). Application of LLMs to fraud detection. World Journal of Advanced Research and Reviews, 24(3). https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-1586.pdf

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 30.

Mastercard. (2024). Mastercard accelerates card fraud detection with generative AI technology. https://newsroom.mastercard.com/news/press/2024/may/mastercard-accelerates-card-fraud-detection-with-generative-ai-technology/

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5(4), 115–133. https://doi.org/10.1007/BF02478259

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35. https://doi.org/10.1145/3457607

Ming, R., Abdelrahman, O., Innab, N., & Ibrahim, M. H. K. (2024). Enhancing fraud detection in auto insurance and credit card transactions: A novel approach integrating CNNs and machine learning algorithms. PeerJ Computer Science, 10, e2088. https://doi.org/10.7717/peerj-cs.2088

Nasdaq Verafin. (2024). 2024 global financial crime report. Nasdaq. https://verafin.com/nasdaq-verafin-global-financial-crime-report/

Ngo, T. T. B., Le, T. M., & Nguyen, T. H. (2025). Year-over-year developments in financial fraud detection via deep learning: A systematic literature review. arXiv. https://arxiv.org/abs/2502.00201

Nicholls, J., Kuppa, A., & Le-Khac, N. A. (2021). Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. IEEE Access, 9, 163965–163986. https://doi.org/10.1109/ACCESS.2021.3134076

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in data-driven artificial intelligence systems: An introductory survey. WIREs Data Mining and Knowledge Discovery, 10(3), e1356. https://doi.org/10.1002/widm.1356

Obeng, S., Yussif, S., & Asante, M. (2025). Financial fraud detection using explainable AI and stacking ensemble methods. arXiv. https://arxiv.org/html/2505.10050v1

JIRRS Journal International Review of Research Studies

Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Ogundipe, T. (2025). Autonomous learning models for adaptive fraud detection in real-time payment environments. Global Knowledge Academy. http://globalknowledgeacademy.com/index.php/gna/article/download/62/139

Okonkwo, C. J., Okeke, G. M., & Okonkwo, B. N. (2025). Backend latency optimization in real-time fraud detection systems. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(1), 3295–3308.

Papakostas, D., Soupionis, Y., Sarigiannidis, P., & Voulgaridis, K. (2025). Organisational challenges in US law enforcement's response to AI-driven cybercrime and deepfake fraud. Laws, 14(4), 46. https://doi.org/10.3390/laws14040046

Park, Y. S., Konge, L., & Artino, A. R. (2020). The positivism paradigm of research. Academic Medicine, 95(5), 690–694. https://doi.org/10.1097/ACM.0000000000003093

Pereira Costa, G. (2026). The Role of Game Engines in the Democratization of Digital Game Development. Journal International Review of Research Studies, 1(02), 1-17. https://doi.org/10.66104/8yn0e657

Pretorius, L. (2024). Demystifying research paradigms: Navigating ontology, epistemology, and axiology in research. The Qualitative Report, 29(10), 2698–2715. https://doi.org/10.46743/2160-3715/2024.7632

PYMNTS. (2024). Mastercard uses AI to solve "jigsaw" of credit card fraud. https://www.pymnts.com/fraud-prevention/2024/mastercard-uses-ai-to-solve-jigsaw-of-credit-card-fraud/

QED Investors. (2025). The frontlines of fraud: How Brazil is becoming a global testbed for financial crime prevention. https://www.qedinvestors.com/blog/the-frontlines-of-fraud-how-brazil-is-becoming-a-global-testbed-for-financial-crime-prevention

Ramachandran, A., Kumar, S., & Patel, R. (2025). A comparative study on real-time data streaming for fraud detection using Kafka with Apache Flink and Apache Spark. Procedia Computer Science. https://doi.org/10.1016/j.procs.2025.03.001

Raufi, B., Finnegan, C., & Longo, L. (2024). A comparative analysis of SHAP, LIME, ANCHORS, and DICE for interpreting a dense neural network in credit card fraud detection. In L. Longo, S. Lapuschkin, & C. Seifert (Eds.), Explainable Artificial

JIRRS Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Intelligence. xAI 2024. Communications in Computer and Information Science (Vol. 2156, pp. 1–20). Springer. https://doi.org/10.1007/978-3-031-63803-9_20

Ravi, A., Msahli, M., Qiu, H., Memmi, G., Bifet, A., & Qiu, M. (2025). Advanced real-time fraud detection using RAG-based LLMs. arXiv. https://arxiv.org/abs/2501.15290

Razaque, A., Frej, M. B. H., Bektemyssova, G., Amsaad, F., Almiani, M., Alotaibi, A., Jhanjhi, N. Z., Amanzholova, S., & Alshammari, M. (2023). Credit card-not-present fraud detection and prevention using big data analytics algorithms. Applied Sciences, 13(1), 57. https://doi.org/10.3390/app13010057

Reality Defender. (2025). How deepfakes exploit KYC verification systems. https://www.realitydefender.com/insights/how-deepfakes-exploit-kyc-verification-systems

Redis. (2025). Feature stores for real-time AI/ML: Benchmarks, architectures, and case studies. https://redis.io/blog/feature-stores-for-real-time-artificial-intelligence-and-machine-learning/

Rigaki, M., & Garcia, S. (2023). A survey of transfer learning for adversarial robustness in deep neural networks. ACM Computing Surveys, 56(1), 1–37. https://doi.org/10.1145/3600190

Rosenblatt, F. (1962). Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan Books.

Rtayli, N., & Enneya, N. (2024). A soft voting ensemble learning approach for credit card fraud detection. PLOS ONE, 19(2). https://doi.org/10.1371/journal.pone.0299023

Salomão, P. E. A. (2026). Tokens as computational units in data science and machine learning: Mathematical foundations, transformer architecture, inference economy, and caching systems in foundational models. Journal International Review of Research Studies, 1(02), 1–27. https://doi.org/10.66104/kxf7hk05

Semmelrock, L., Kern, R., & Veasna, S. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. AI Magazine. https://doi.org/10.1002/aaai.70002

Silverguard. (2024). 2024 Pix scam study. https://silverguard.com.br

JIRRS Journal International Review of Research Studies

Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Sundararajan, M., Taly, A., & Yan, Q. (2020). Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning, 3319–3328.

Taktile. (2024). How LLMs are becoming investigative partners in fintech fraud detection. https://taktile.com/articles/llms-investigative-partners-fraud-detection

Tecton. (2023). What is a feature store? https://www.tecton.ai/blog/what-is-a-feature-store/

Teixeira, P., Rodrigues, J., & Silva, C. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11, 1185. https://doi.org/10.1057/s41599-024-03606-0

Transunion. (2024). H1 2024 state of omnichannel fraud report. https://transunion.com/fraud-report

van Essen, T., Floris, M., & Verbeke, W. (2025). Finding money launderers using heterogeneous graph neural networks. Finance Research Letters, 73, 106713. https://doi.org/10.1016/j.frl.2025.106713

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30.

Vo, T. T., Nguyen, L. H., & Tran, T. A. (2025). Advancing fraud detection with hybrid AI: A MoE, RNN, and transformer approach. Journal of Intelligence and Engineering, 3(3). https://doi.org/10.58915/jie.v3i3.54

W Fintechs. (2025). Fraud in Pix and the evolution of anti-fraud infrastructure. https://wfintechs.substack.com/p/152-en

Xiao, F., Gu, Y., Li, Y., Wang, C., Zhang, X., Li, X., & Tang, R. (2023). MINT: Detecting fraudulent behaviors from time-series relational data. Proceedings of the VLDB Endowment, 16(12), 3610–3623. https://doi.org/10.14778/3611540.3611548

Zavitsanos, E., Kelesis, D., & Paliouras, G. (2025). Calibrating TabTransformer for financial misstatement detection. Applied Intelligence, 55, 3. https://doi.org/10.1007/s10489-024-05861-9

Zhou, Y., Rong, H., & Khatri, V. (2023). TensorRT implementations of model quantization on edge SoC. Proceedings of the IEEE International SoC Conference. https://par.nsf.gov/servlets/purl/10488646

JIRRS
Journal International Review of Research Studies
Received: 01/02/2026 - **Accepted:** 05/03/2026
Vol: 01.03
DOI: 10.66104/xkad2306
Pages: 1-49

Zhu, H., Chen, F., & Li, X. (2024). Privacy issues, attacks, countermeasures and open problems in federated learning: A survey. Connection Science, 36(1). https://doi.org/10.1080/08839514.2024.2410504