World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Review Article)

Machine Learning Models for Multi Sector Fraud Detection in Public and Private Financial Systems

Daniel Tom Agara [1, *], Samuel Sunday Omotoso [2] and Oluwatosin Junior Alabi [1]

[1] *Department of Business Administration, University of Arkansas, Fayetteville.*

[2] *Department of Information Systems, University of Arkansas, Fayetteville.*

World Journal of Advanced Research and Reviews, 2026, 29(02), 1345-1354

Publication history: Received on 13 January 2026; revised on 21 February 2026; accepted on 24 February 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.2.0397

Abstract

Fraud detection is a critical challenge across public and private financial systems due to increasing transaction volumes, evolving adversarial behavior, and strict regulatory constraints. While machine learning has significantly improved fraud detection performance, existing research largely focuses on single-sector applications, limiting its effectiveness in addressing cross-sector fraud patterns. This paper presents a comprehensive review of machine learning models for fraud detection in multi-sector financial systems, covering banking, insurance, and public-sector domains. We examine supervised, unsupervised, semi-supervised, deep learning, and graph-based approaches, highlighting their strengths and limitations in real-world settings. In addition, we analyze key challenges related to data imbalance, label uncertainty, evaluation metrics, scalability, privacy, and explainability. Finally, the paper identifies open research directions, including cross-sector generalization, privacy-preserving collaboration, scalable graph learning, and continual adaptation to evolving fraud strategies. This review provides a structured reference for researchers and practitioners seeking to design effective, responsible, and scalable fraud detection systems.

Keywords: Fraud Detection; Machine Learning; Multi-Sector Financial Systems; Graph-Based Learning; Privacy-Preserving Analytics

1. Introduction

Fraud in financial systems remains a persistent and costly problem for both public and private institutions. The rapid expansion of digital financial services, including online banking, mobile payments, and e-government platforms, has significantly increased transaction volumes and system complexity. While these developments have improved efficiency and accessibility, they have also created new opportunities for fraudulent activities that span multiple sectors, such as banking, insurance, taxation, public procurement, and social welfare programs (Dal Pozzolo et al., 2022; Hernández et al., 2024). The multi-sector nature of modern fraud complicates detection efforts, as fraudulent behavior often exploits interactions across institutions rather than remaining confined to a single domain. Historically, fraud detection systems have relied on rule-based approaches and expert-defined heuristics. Although effective for detecting well-known fraud patterns, these systems struggle to adapt to evolving fraud strategies and typically require frequent manual updates. As fraudsters continuously modify their behavior to bypass static rules, rule-based systems often exhibit high false-negative rates and delayed detection. Furthermore, overly conservative rules may generate large numbers of false positives, increasing investigation costs and reducing trust in automated systems (Bahnsen et al., 2023). These limitations have driven growing interest in data-driven and machine learning (ML) approaches for fraud detection.

Machine learning techniques have demonstrated strong potential for improving fraud detection accuracy by learning complex, non-linear patterns from large-scale transactional data. Supervised learning methods, including logistic

[*] Corresponding author: Daniel Tom Agara

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

regression, decision trees, random forests, and gradient-boosted models, remain widely used due to their strong performance and relative interpretability (Carcillo et al., 2022). Deep learning models, such as recurrent neural networks (RNNs) and transformer-based architectures, have further enhanced the modeling of sequential transaction behavior and temporal dependencies (Fiore et al., 2023). In parallel, unsupervised and semi-supervised methods, including isolation forests and autoencoders, have been explored to identify anomalous behavior in settings where labeled fraud data are scarce or delayed (Pumsirirat & Yan, 2022). More recently, graph-based learning has emerged as a powerful paradigm for fraud detection. Financial fraud often involves networks of related entities, such as shared devices, identities, or transaction pathways, which are naturally represented as graphs. Graph analytics and graph neural networks (GNNs) have shown strong performance in identifying organized fraud rings and collusive behavior that may not be detectable using transaction-level features alone (Cheng et al., 2023; Weber et al., 2022). These approaches are particularly relevant for multi-sector fraud scenarios, where illicit activities may span multiple institutions or administrative domains.

Despite significant progress, existing research largely treats fraud detection as a sector-specific problem. Most studies focus on individual domains—such as credit card fraud, insurance claim fraud, or tax evasion—using datasets and evaluation protocols tailored to a single application context. However, real-world fraud increasingly exhibits cross-sector characteristics, including identity reuse, coordinated abuse of public and private systems, and long-term fraud campaigns that evolve over time (Hernández et al., 2024). The lack of a unified, cross-sector perspective limits the ability of current models to generalize and detect complex fraud patterns that transcend institutional boundaries.

Several challenges further complicate the application of ML to multi-sector fraud detection. Extreme class imbalance is a defining characteristic of fraud datasets, with fraudulent events often accounting for less than one percent of all transactions, making model training and evaluation difficult (Dal Pozzolo et al., 2022). Label noise and delayed ground truth, common in both public and private sectors, reduce the reliability of supervised learning approaches. Additionally, data heterogeneity across sectors—differences in schemas, identifiers, and data quality—poses significant barriers to model reuse and transfer learning. Privacy and regulatory constraints, particularly in public financial systems, further restrict data sharing and necessitate privacy-preserving learning strategies (Kairouz et al., 2021; Yang et al., 2023). Beyond technical considerations, ethical and regulatory concerns play a critical role in fraud detection, especially in public-sector applications. Automated fraud detection systems may have direct consequences for individuals, including denial of benefits or legal investigations. Recent studies have emphasized the importance of explainability, fairness, and human oversight to mitigate potential harms arising from opaque or biased models (Barredo Arrieta et al., 2022). As a result, explainable AI techniques and human-in-the-loop frameworks are increasingly viewed as essential components of trustworthy fraud detection systems.

This paper provides a comprehensive review of machine learning models for fraud detection with a specific focus on multi-sector financial systems. Unlike prior surveys that concentrate on individual domains, this review synthesizes approaches across public and private sectors, highlighting common challenges, methodological trends, and emerging solutions. The main contributions of this paper are threefold: (1) a structured taxonomy of ML-based fraud detection techniques applicable across sectors; (2) an analysis of data, evaluation, and regulatory challenges unique to multi-sector settings; and (3) a discussion of open research directions, including graph-based learning, privacy-preserving collaboration, and explainable fraud detection. The remainder of this paper is organized as follows. Section 2 reviews fraud typologies across sectors, Section 3 surveys machine learning approaches, Section 4 discusses evaluation practices, Section 5 highlights open challenges and future research directions, and Section 6 concludes the paper.

2. Fraud Typologies Across Financial Sectors

Fraud in financial systems encompasses a broad range of deceptive activities designed to obtain unauthorized financial benefits. The nature of fraud varies substantially across sectors due to differences in transaction processes, regulatory frameworks, data availability, and incentive structures. In multi-sector environments, fraud schemes may exploit interactions between public and private systems, making detection more complex than in single-domain settings. A clear understanding of sector-specific fraud typologies is therefore essential for developing effective machine learning–based detection models and for interpreting their outputs in real-world deployments (Hernández et al., 2024). This section reviews the most prevalent fraud typologies across major financial sectors, including banking and payments, insurance, and public financial systems, and highlights common patterns that motivate cross-sector analytical approaches.

2.1. Banking and Payment Systems Fraud

Banking and payment systems represent one of the most mature domains in fraud detection research due to the high volume of transactions and the direct financial impact of fraudulent activity. Common fraud types in this sector include credit and debit card fraud, account takeover, identity theft, merchant fraud, and unauthorized electronic funds transfers (Dal Pozzolo et al., 2022). These frauds are often characterized by rapid execution, requiring near real-time detection to minimize losses. Transaction data in banking systems are typically structured and high frequency, including attributes such as transaction amount, time, location, merchant category, payment channel, and device identifiers. While this richness supports advanced analytics, it also creates challenges related to class imbalance, as fraudulent transactions often represent less than one percent of total activity (Carcillo et al., 2022). Furthermore, legitimate user behavior can vary widely, making it difficult to distinguish anomalous but benign transactions from actual fraud. A notable trend in banking fraud is the increasing sophistication of fraudsters, who leverage social engineering techniques such as phishing and malware to gain access to user credentials. These attacks blur the boundary between cybercrime and financial fraud, requiring detection systems to incorporate behavioral and contextual signals rather than relying solely on transaction-level anomalies (Fiore et al., 2023).

2.2. Insurance Fraud

Insurance fraud constitutes a significant source of financial loss for insurers and policyholders alike. Unlike payment fraud, insurance fraud often unfolds over longer time horizons and involves more complex interactions between claimants, service providers, and intermediaries. Common typologies include false claims, exaggerated claims, staged accidents, premium evasion, and identity misrepresentation (Bahnsen et al., 2023).

Insurance data are typically heterogeneous and semi-structured, combining numerical attributes (e.g., claim amounts), categorical variables (e.g., policy type), textual descriptions, and sometimes multimedia evidence such as images or medical reports. The detection of insurance fraud is further complicated by delayed labeling, as investigations may take months to conclude, and by the subjective nature of some claims, which can blur the distinction between opportunistic fraud and legitimate error. Machine learning applications in insurance fraud detection often emphasize pattern recognition across claims histories and relationships among entities, such as shared addresses, service providers, or vehicles. These relational characteristics make insurance fraud a natural candidate for network-based and graph-oriented analysis, particularly when detecting organized fraud rings that operate across multiple policies or insurers (Cheng et al., 2023).

2.3. Public and Government Financial Fraud

Fraud in public financial systems presents distinct challenges due to the scale, societal impact, and regulatory sensitivity of government programs. Common forms of public-sector fraud include tax evasion, welfare and social benefit fraud, procurement fraud, subsidy abuse, and corruption-related activities (Hernández et al., 2024). Unlike private-sector fraud, the consequences of false positives in public systems can be severe, potentially resulting in wrongful denial of benefits or legal action against individuals. Public-sector fraud often involves complex administrative processes and long-term behavioral patterns rather than isolated transactions. For example, welfare fraud may involve misreporting income or household composition over extended periods, while procurement fraud may involve collusion among suppliers and officials. These characteristics complicate the use of purely transactional detection models and require longitudinal and relational analysis. Data availability in public systems is frequently constrained by privacy regulations and institutional silos. Government datasets may be incomplete, inconsistently formatted, or distributed across multiple agencies, limiting opportunities for centralized analysis. Additionally, public-sector fraud detection systems are subject to heightened scrutiny regarding transparency, fairness, and accountability, necessitating explainable and auditable decision-making processes (Barredo Arrieta et al., 2022).

2.4. Cross-Sector Fraud Patterns

While sector-specific fraud typologies exhibit distinct characteristics, many modern fraud schemes operate across institutional and sectoral boundaries. Cross-sector fraud patterns include identity reuse across banking, insurance, and welfare systems; coordinated abuse of public benefits followed by laundering through private financial channels; and organized fraud rings that exploit gaps between regulatory jurisdictions (Hernández et al., 2024). These cross-sector schemes highlight the limitations of isolated detection systems that operate within a single organizational context. Fraudsters may exploit inconsistencies in identity verification, reporting standards, or detection thresholds across institutions to evade detection. As a result, there is growing interest in analytical approaches that can integrate signals from multiple sectors while respecting privacy and legal constraints. From a data perspective, cross-sector fraud detection requires the alignment of heterogeneous datasets with differing schemas, identifiers, and temporal resolutions. This alignment challenge complicates feature engineering and model transferability, motivating research into representation learning and federated analytics that can operate across decentralized data sources (Yang et al., 2023).

2.5. Implications for Machine Learning–Based Detection

The diversity of fraud typologies across sectors has important implications for the design of machine learning models. High-frequency payment fraud may favor real-time, transaction-level classifiers, while insurance and public-sector fraud often require longitudinal and relational analysis. Models trained on one sector may not generalize well to others due to differences in data distributions, fraud incentives, and operational constraints. Moreover, the societal impact of fraud detection varies by sector, influencing acceptable trade-offs between false positives and false negatives. In public financial systems, minimizing harm to legitimate beneficiaries may take precedence over maximizing detection rates, whereas private financial institutions may prioritize loss reduction. These differing objectives underscore the need for adaptable and context-aware ML frameworks capable of supporting multi-sector deployment. fraud typologies differ substantially across financial sectors but increasingly exhibit cross-sector characteristics that challenge traditional detection approaches. A comprehensive understanding of these typologies provides the foundation for reviewing machine learning methods capable of addressing the complexity, scale, and ethical considerations of multi-sector fraud detection, which is the focus of the subsequent sections.

3. Literature Review

The application of machine learning (ML) to fraud detection has grown substantially as financial systems have become more digitized, interconnected, and data-intensive. Researchers have explored a wide range of learning paradigms to address the unique characteristics of fraud data, including extreme class imbalance, evolving adversarial behavior, heterogeneous data sources, and strict regulatory requirements. This section reviews the major categories of ML approaches used for fraud detection, with emphasis on their strengths, limitations, and relevance to multi-sector financial systems.

3.1. Supervised Learning Methods

Supervised learning techniques have historically formed the foundation of ML-based fraud detection systems. These methods rely on labeled historical data to learn decision boundaries between fraudulent and legitimate transactions. Early studies employed classical classifiers such as logistic regression, decision trees, and support vector machines due to their simplicity and interpretability. Despite their limitations in modeling complex patterns, such models remain relevant in highly regulated environments where transparency and auditability are critical (Carcillo et al., 2022). More recent literature has demonstrated the effectiveness of ensemble-based supervised models, particularly random forests and gradient boosting machines. Gradient boosting frameworks such as XGBoost and LightGBM have become dominant in both academic studies and industrial applications due to their ability to handle non-linear feature interactions, missing values, and large-scale datasets (Dal Pozzolo et al., 2022). Several studies report that these models achieve state-of-the-art performance on benchmark fraud datasets, especially when combined with cost-sensitive learning techniques that explicitly account for the asymmetric costs of false positives and false negatives (Bahnsen et al., 2023). However, supervised approaches face significant challenges in real-world fraud detection. High-quality labeled data are often scarce, delayed, or noisy, particularly in public-sector domains such as tax or welfare fraud, where investigations may take months to conclude. Moreover, models trained on historical data may degrade over time due to concept drift, as fraud strategies evolve in response to deployed detection systems (Dal Pozzolo et al., 2022). These limitations have motivated the exploration of alternative learning paradigms that reduce reliance on fully labeled datasets.

3.2. Unsupervised and Semi-Supervised Learning

Unsupervised learning approaches aim to detect fraud by identifying deviations from normal behavior without relying on labeled fraud examples. These methods are particularly attractive in domains where labeled data are unavailable or unreliable. Common techniques include clustering algorithms, One-Class Support Vector Machines, and Isolation Forests, which model the distribution of legitimate transactions and flag outliers as potential fraud (Pumsirirat & Yan, 2022). Deep autoencoder models have gained prominence in recent years due to their ability to learn compact representations of normal behavior from high-dimensional data. Fraudulent transactions are then identified based on reconstruction error. Fiore et al. (2023) demonstrated that deep autoencoders can outperform traditional anomaly detection methods in highly imbalanced financial datasets, particularly when fraud patterns differ substantially from normal behavior. Nonetheless, unsupervised methods may suffer from high false-positive rates, as rare but legitimate behaviors can be incorrectly flagged as anomalies. Semi-supervised learning techniques attempt to bridge the gap between supervised and unsupervised approaches by leveraging limited labeled fraud cases alongside large volumes of unlabeled data. Positive-unlabeled (PU) learning, self-training, and label propagation methods have been explored in this context (Carcillo et al., 2022). While these approaches improve robustness to label scarcity, their effectiveness depends heavily on assumptions about the unlabeled data and may be sensitive to noise, especially in heterogeneous, multi-sector environments.

3.3. Deep Learning for Sequential and Behavioral Modeling

Deep learning has enabled significant advances in fraud detection by capturing complex temporal and behavioral patterns that are difficult to model using traditional feature engineering. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) and gated recurrent unit (GRU) architectures, have been widely applied to model sequences of transactions associated with individual accounts or users. These models can capture temporal dependencies and behavioral evolution, which are critical for detecting account takeover and behavioral fraud (Fiore et al., 2023). Convolutional neural networks (CNNs) have also been explored for fraud detection, particularly for feature extraction from structured or semi-structured data. More recently, transformer-based architectures have gained attention due to their ability to model long-range dependencies and process sequences in parallel using attention mechanisms. Yang et al. (2023) reported that transformer models outperform RNN-based approaches on large-scale transaction datasets, especially when long transaction histories are available.

Despite their strong performance, deep learning models present challenges related to interpretability, computational complexity, and deployment. In regulated financial and public-sector environments, the lack of transparency associated with deep neural networks can hinder adoption. Furthermore, deep models typically require substantial amounts of data and computational resources, which may not be available in smaller institutions or public agencies. These limitations have spurred interest in hybrid approaches that combine deep learning with interpretable models or rule-based systems.

3.4. Graph-Based Learning and Network Analysis

Graph-based learning has emerged as a particularly promising direction for fraud detection, especially for identifying organized and collusive fraud schemes. In financial contexts, entities such as users, accounts, devices, merchants, and transactions can be represented as nodes, with edges encoding relationships such as transactions, shared identifiers, or temporal proximity. Early graph-based approaches relied on handcrafted network features, including degree centrality, clustering coefficients, and community detection algorithms. Recent advances in graph representation learning have significantly expanded the capabilities of network-based fraud detection. Node embedding techniques, such as DeepWalk and Node2Vec, learn low-dimensional representations of entities based on graph structure and have been shown to improve classification performance when combined with traditional ML models (Weber et al., 2022). More recently, graph neural networks (GNNs), including Graph Convolutional Networks and Graph Attention Networks, have been applied to fraud detection tasks with promising results (Cheng et al., 2023). GNN-based models are particularly well suited for multi-sector fraud detection, as they can capture relational patterns that span multiple institutions or domains. However, scalability remains a key challenge, as real-world financial graphs may contain millions of nodes and edges. Additionally, privacy concerns arise when constructing graphs that integrate data across organizational boundaries, motivating research into federated and privacy-preserving graph learning techniques (Yang et al., 2023).

3.5. Privacy-Preserving and Collaborative Learning

The increasing emphasis on data privacy and regulatory compliance has driven interest in privacy-preserving ML techniques for fraud detection. Federated learning enables multiple institutions to collaboratively train models without sharing raw data, instead exchanging model updates or gradients (Kairouz et al., 2021). Recent studies have explored federated learning for banking and payment fraud detection, demonstrating that collaborative models can outperform institution-specific models while respecting data privacy constraints (Yang et al., 2023). Secure multi-party computation and differential privacy have also been investigated as complementary techniques to enhance privacy guarantees. While these approaches offer strong theoretical protections, they often introduce computational overhead and may degrade model performance. Balancing privacy, utility, and scalability remains an open challenge, particularly in multi-sector settings involving both public and private stakeholders.

The reviewed literature demonstrates that no single ML approach is universally optimal for fraud detection across all financial sectors. Supervised models excel when labeled data are abundant, while unsupervised and semi-supervised methods address label scarcity. Deep learning and graph-based approaches capture complex behavioral and relational patterns but raise concerns regarding interpretability, scalability, and privacy. Notably, most existing studies focus on single-sector applications, with limited attention to cross-sector fraud dynamics.

These gaps highlight the need for integrated frameworks that combine multiple learning paradigms, support privacy-preserving collaboration, and account for sector-specific constraints. The next sections build on this review by examining evaluation practices, operational challenges, and future research directions for multi-sector fraud detection.

4. Data Characteristics, Evaluation Metrics, and Practical Challenges

Machine learning–based fraud detection differs substantially from conventional predictive modeling tasks due to the unique characteristics of financial data, the adversarial nature of fraud, and the operational constraints under which detection systems must function. These challenges are amplified in multi-sector environments, where data originate from heterogeneous public and private systems and must comply with strict regulatory and ethical requirements. This section reviews the key data-related challenges, evaluation practices, and practical considerations that shape the design and deployment of fraud detection models.

4.1. Data Characteristics in Financial Fraud Detection

One of the most prominent characteristics of fraud detection datasets is extreme class imbalance. In many real-world financial systems, fraudulent transactions represent less than 0.5% of all observations, and in some public-sector datasets, the proportion may be even lower (Dal Pozzolo et al., 2022). This imbalance poses significant difficulties for standard machine learning algorithms, which are typically optimized to maximize overall accuracy and may therefore prioritize the majority class. As a result, models trained without appropriate imbalance-handling techniques may achieve deceptively high accuracy while failing to detect meaningful fraud patterns.

Closely related to class imbalance is the issue of label scarcity and label delay. Fraud labels are often assigned only after extensive investigation, which may take weeks or months, particularly in insurance and public financial systems. During this delay, fraudulent instances may be treated as legitimate, introducing noise into training datasets (Carcillo et al., 2022). Moreover, labels themselves may be uncertain or subject to revision, as legal disputes or appeals can overturn prior determinations. These factors reduce the reliability of supervised learning and motivate the use of semi-supervised, unsupervised, and continual learning approaches.

Another critical challenge is data heterogeneity, especially in multi-sector fraud detection. Private-sector systems, such as banking and payment platforms, typically generate high-frequency, structured transaction data with standardized schemas. In contrast, public-sector systems may rely on low-frequency administrative records, manual reports, or semi-structured documentation, often with missing or inconsistent fields (Hernández et al., 2024). Insurance fraud datasets may additionally include unstructured data such as free-text claims descriptions or multimedia evidence. This heterogeneity complicates feature engineering, model transferability, and cross-sector analysis.

Concept drift further complicates fraud detection. Fraudsters actively adapt their strategies in response to deployed detection mechanisms, leading to changes in data distributions over time. Additionally, external factors such as regulatory changes, economic shocks, or shifts in user behavior can alter the statistical properties of legitimate transactions. As a result, models trained on historical data may experience significant performance degradation if not continuously monitored and updated (Dal Pozzolo et al., 2022). Drift detection and adaptive learning mechanisms are therefore essential components of robust fraud detection systems.

4.2. Data Quality, Bias, and Representation Issues

Beyond structural challenges, data quality issues significantly impact fraud detection performance. Missing values, incorrect entries, and inconsistent identifiers are common in large-scale financial datasets, particularly when integrating data from multiple institutions or sectors. In public-sector datasets, incomplete reporting or outdated records may introduce systematic biases that disproportionately affect certain populations (Barredo Arrieta et al., 2022).

Bias in training data can lead to discriminatory outcomes when deployed models disproportionately flag transactions or individuals associated with specific demographic groups. While demographic attributes are often excluded from fraud detection models, proxy variables—such as geographic location or transaction timing—may still encode sensitive information. Addressing these representation and bias issues requires careful dataset analysis, fairness-aware modeling techniques, and ongoing auditing of deployed systems.

4.3. Evaluation Metrics for Fraud Detection Systems

The evaluation of fraud detection models requires careful consideration of both statistical performance and operational impact. Due to extreme class imbalance, traditional metrics such as accuracy are largely uninformative and may obscure poor fraud detection performance. Consequently, the literature emphasizes metrics that focus on the minority (fraud) class.

Precision and recall are the most commonly reported metrics in fraud detection studies. Precision reflects the proportion of detected fraud cases that are truly fraudulent, while recall measures the proportion of actual fraud cases that are successfully detected. In practice, the balance between these metrics depends on the application context. For example, public-sector systems may prioritize high precision to minimize harm to legitimate beneficiaries, whereas private financial institutions may tolerate higher false-positive rates to reduce financial losses (Bahnsen et al., 2023).

The F1-score, which combines precision and recall, provides a single summary measure but may still fail to capture operational priorities. As a result, many studies report the precision–recall area under the curve (PR-AUC), which is more informative than ROC-AUC in highly imbalanced settings (Carcillo et al., 2022). PR-AUC emphasizes model performance on the positive class and facilitates comparison across different detection thresholds.

In real-world deployments, cost-sensitive evaluation is often more relevant than purely statistical metrics. False negatives may result in direct financial loss, while false positives incur investigation costs and potential reputational damage. Cost matrices and decision-theoretic frameworks have therefore been proposed to align model evaluation with institutional objectives (Bahnsen et al., 2023). However, the lack of standardized cost assumptions across studies limits comparability and reproducibility in the literature.

4.4. Temporal Validation and Benchmarking Challenges

Another important consideration in fraud detection evaluation is temporal validation. Random train–test splits may lead to overly optimistic performance estimates by allowing information leakage from future data. Time-aware validation strategies, such as rolling-window or forward-chaining evaluation, are therefore recommended to better reflect real-world deployment scenarios (Dal Pozzolo et al., 2022).

Benchmarking fraud detection models also presents challenges due to the scarcity of publicly available, realistic datasets. Many widely used benchmarks are either outdated, synthetically generated, or limited to single-sector scenarios, reducing their relevance to modern, multi-sector fraud detection problems. This lack of standardized benchmarks complicates fair comparison of methods and slows progress in the field.

4.5. Operational and Deployment Constraints

Beyond data and evaluation issues, operational constraints strongly influence the design of fraud detection systems. Scalability and latency requirements are particularly critical in high-volume transaction environments, where detection decisions must be made in milliseconds. While complex models such as deep neural networks and graph-based approaches may offer improved detection accuracy, they can be difficult to deploy under strict latency and resource constraints (Cheng et al., 2023).

Interpretability and transparency are also essential, especially in regulated industries and public-sector applications. Stakeholders may require clear explanations for automated decisions, both for regulatory compliance and for maintaining public trust. Explainable AI techniques and hybrid human–AI systems are therefore increasingly adopted to balance predictive performance with accountability (Barredo Arrieta et al., 2022).

Finally, privacy and data governance constraints impose additional limitations on fraud detection systems. Regulations such as the General Data Protection Regulation (GDPR) restrict the collection, processing, and sharing of personal data, particularly across institutional boundaries. In multi-sector settings, these constraints often prevent centralized data aggregation, motivating research into privacy-preserving learning paradigms such as federated learning and secure computation (Yang et al., 2023). fraud detection systems operate under a unique combination of data, evaluation, and operational challenges that distinguish them from standard machine learning applications. Extreme class imbalance, label uncertainty, data heterogeneity, bias, and concept drift complicate model development, while evaluation metrics and deployment constraints shape practical decision-making. These challenges are particularly pronounced in multi-sector environments, where public and private systems intersect under diverse regulatory and ethical frameworks. Addressing these issues is critical for advancing the effectiveness, fairness, and trustworthiness of machine learning–based fraud detection systems.

5. Open Challenges and Future Research Directions

Despite significant advances in machine learning–based fraud detection, numerous challenges remain unresolved, particularly in the context of multi-sector financial systems that span public and private domains. Existing approaches often address isolated aspects of the problem, such as predictive accuracy or scalability, while overlooking broader issues related to generalization, fairness, privacy, and long-term adaptability. This section identifies key open challenges in the literature and outlines promising directions for future research.

5.1. Cross-Sector Generalization and Transferability

One of the most significant open challenges in fraud detection research is the limited ability of models to generalize across sectors. Most existing studies focus on single-sector datasets, such as credit card transactions or insurance claims, and develop models tailored to the characteristics of a specific domain. However, fraud increasingly operates across institutional boundaries, exploiting inconsistencies between public and private systems (Hernández et al., 2024).

Future research should explore cross-sector transfer learning and domain adaptation techniques that enable models trained in one sector to be effectively applied to another. Representation learning methods that learn sector-invariant features, as well as meta-learning approaches that adapt quickly to new domains with limited labeled data, represent promising avenues. Developing benchmarks and evaluation protocols for cross-sector generalization remains an important prerequisite for progress in this area.

5.2. Privacy-Preserving and Collaborative Fraud Detection

Data sharing across institutions is often restricted by legal, ethical, and competitive concerns, particularly when public-sector data are involved. While federated learning has emerged as a promising solution for collaborative model training without raw data exchange, current approaches face limitations related to communication overhead, statistical heterogeneity, and vulnerability to privacy leakage through model updates (Yang et al., 2023).

Future work should focus on scalable and robust federated learning frameworks tailored to fraud detection, including techniques for handling non-identically distributed data across participants. Privacy-enhancing technologies such as differential privacy and secure multi-party computation require further investigation to balance privacy guarantees with model utility and computational efficiency. Extending these techniques to graph-based fraud detection, where relational data pose additional privacy risks, remains an open and underexplored challenge.

5.3. Scalability of Graph-Based and Hybrid Models

Graph-based learning has demonstrated strong potential for detecting organized and collusive fraud; however, scalability remains a major bottleneck. Real-world financial graphs may contain millions of nodes and edges, with dynamic updates occurring in near real time. Many existing graph neural network (GNN) models struggle to scale to such settings without sacrificing performance or incurring prohibitive computational costs (Cheng et al., 2023).

Future research should investigate efficient graph representation learning, including sampling strategies, streaming graph models, and distributed graph processing frameworks. Hybrid architectures that combine lightweight tabular models for real-time screening with deeper graph-based analysis for post hoc investigation may offer practical trade-offs between accuracy and latency. Additionally, methods for incremental graph learning that update models without full retraining are critical for deployment in evolving fraud environments.

5.4. Continual Learning and Concept Drift Adaptation

Fraud detection systems operate in adversarial environments where fraud strategies evolve in response to detection mechanisms. Concept drift is therefore a persistent and unavoidable challenge. While some studies have explored drift detection and periodic model retraining, these approaches are often reactive and may fail to adapt quickly to emerging fraud patterns (Dal Pozzolo et al., 2022).

Continual and lifelong learning frameworks represent a promising direction for addressing concept drift in fraud detection. Such approaches aim to update models incrementally as new data arrive, while avoiding catastrophic forgetting of previously learned patterns. However, designing continual learning systems that are stable, interpretable, and compliant with regulatory requirements remains an open research problem, particularly in public-sector applications where model changes must be carefully audited.

5.5. Explainability, Fairness, and Human-Centered Design

As fraud detection systems increasingly influence high-stakes decisions, ensuring explainability and fairness has become a central concern. While explainable AI (XAI) techniques such as SHAP and rule extraction are widely used, their effectiveness in complex models—particularly deep learning and graph-based systems—remains limited (Barredo Arrieta et al., 2022).

Future research should move beyond post hoc explanations toward intrinsically interpretable models and human-centered design frameworks that integrate domain expertise into model development and decision-making. Additionally, fairness-aware learning techniques tailored to fraud detection are needed to mitigate disparate impacts on specific demographic groups, especially in public financial systems. Evaluating fairness in the presence of incomplete or sensitive demographic data remains a significant methodological challenge.

5.6. Benchmarking, Reproducibility, and Data Availability

The lack of standardized, realistic benchmarks continues to hinder progress in fraud detection research. Many publicly available datasets are outdated, heavily sanitized, or limited to narrow use cases, reducing their relevance to modern, multi-sector fraud scenarios. Furthermore, inconsistent evaluation protocols and cost assumptions limit the comparability and reproducibility of reported results (Carcillo et al., 2022).

Future efforts should prioritize the development of shared benchmarking frameworks, including synthetic but realistic datasets that capture cross-sector fraud dynamics while preserving privacy. Open-source implementations and standardized evaluation pipelines would further enhance reproducibility and accelerate innovation in the field.

5.7. Toward Integrated and Responsible Fraud Detection Systems

Looking forward, an important research direction lies in the development of integrated fraud detection systems that combine multiple learning paradigms, data modalities, and human expertise. Rather than relying on a single model, future systems are likely to employ ensembles of supervised, unsupervised, and graph-based components, complemented by human analysts in the decision loop.

Such systems must also be designed with responsible AI principles in mind, balancing detection performance with transparency, accountability, and societal impact. In public-sector contexts, this includes mechanisms for appeal, oversight, and continuous evaluation of system outcomes. Achieving this integration remains a complex challenge that spans technical, organizational, and policy domains. while machine learning has significantly advanced fraud detection capabilities, substantial research challenges remain, particularly in multi-sector financial systems. Addressing issues related to cross-sector generalization, privacy-preserving collaboration, scalability, concept drift, explainability, and benchmarking is essential for the next generation of fraud detection systems. Progress in these areas will require interdisciplinary collaboration between machine learning researchers, domain experts, policymakers, and practitioners, paving the way for more effective, fair, and trustworthy fraud detection solutions.

6. Conclusion

This paper reviewed machine learning approaches for fraud detection in multi-sector financial systems, emphasizing both public and private domains. By analyzing fraud typologies, learning paradigms, data characteristics, and evaluation practices, the review highlights the strengths and limitations of existing methods in real-world deployments. The analysis shows that no single model is sufficient to address the diversity and complexity of modern fraud, particularly when fraud spans institutional and sectoral boundaries. Key challenges remain in cross-sector generalization, scalability, interpretability, privacy preservation, and adaptation to evolving fraud strategies. Addressing these challenges will require integrated frameworks that combine multiple learning paradigms with human oversight and regulatory awareness. This review aims to support future research and development toward more effective, trustworthy, and deployable fraud detection systems.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Kairouz, P., et al. (2021). *Advances and open problems in federated learning*. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

[2] Barredo Arrieta, A., et al. (2022). *Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information Fusion, 58, 82–115.

[3] Carcillo, F., Dal Pozzolo, A., & Bontempi, G. (2022). *Scarff: A scalable framework for streaming credit card fraud detection*. Data Mining and Knowledge Discovery, 36(1), 1–34.

[4] Dal Pozzolo, A., et al. (2022). *Adversarial drift detection for fraud detection*. IEEE Transactions on Neural Networks and Learning Systems, 33(11), 6625–6637.

[5] Pumsirirat, A., & Yan, L. (2022). *Credit card fraud detection using deep learning*. IEEE Access, 10, 21645–21658.

[6] Weber, M., et al. (2022). *Anti-money laundering in transaction networks using graph neural networks*. Proceedings of the ACM KDD Conference.

[7] Bahnsen, A. C., Aouada, D., & Ottersten, B. (2023). *Cost-sensitive learning for fraud detection*. Expert Systems with Applications, 213, 118984.

[8] Cheng, D., Xiang, S., Shang, C., Zhang, Y., & Yang, Y. (2023). *Graph neural networks for fraud detection: A review*. ACM Computing Surveys, 55(8), 1–38.

[9] Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2023). *Using deep learning for fraud detection in financial systems: A survey*. IEEE Transactions on Emerging Topics in Computing, 11(1), 44–62.

[10] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2023). *Federated learning*. ACM Transactions on Intelligent Systems and Technology, 14(1).

[11] Hernández, L., Jiménez, A., & Martín, R. (2024). *Machine learning techniques for financial fraud detection: A cross-sector review*. Humanities and Social Sciences Communications, 11(1), 1–14.