SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I3P118

Volume 13 Issue 3, 213-226, March 2026
© 2026 Seventh Sense Research Group®

Review Article

Fraud Detection in Credit Card Transactions using a Hybrid Approach of GNN+XAI: A Review

Riya Chaudhari[1], Shilpa Pant[2]

[1,2]*Department of Computer Engineering, MKSSS's Cummins College of Engineering for Women, Maharashtra, India.*

[1]*Corresponding Author: shilpa.deogirkar@cumminscollege.in*

Received: 14 December 2025 Revised: 15 January 2026 Accepted: 20 February 2026 Published: 23 March 2026

Abstract - *The digital payment systems have grown in leaps and bounds, resulting in an increased volume of financial transactions, but simultaneously exposing systems to massive new vulnerabilities. Credit card fraud has increasingly become a challenge to detect because financial transaction datasets in reality present an imbalanced nature, a dynamic nature, and intricate relationships among users, devices, merchants, and geographic positions. Conventional machine learning algorithms have provided important baselines but have broadly treated each financial transaction independently without considering the relational behaviour present in collusive financial fraud. Deep learning algorithms have enhanced these baselines by incorporating temporal characteristics and nonlinear dependencies, but have not offered sufficient capabilities to model intricate financial relationships among fraudulent actors. Recently, research has focused on graph models where Graph Neural Networks have modelled financial settings as graph systems, providing opportunities for multiple financial entities to identify relationships and collusive patterns among them. The black box nature of GNN models, however, has spurred debate among financial institutions concerning their transparency requirements and capabilities in adhering to financial rules and regulations. Techniques in Explainable AI have increasingly gained attention to address these issues, but have presently demonstrated limited capabilities in providing transactional-level explanations for large graphs with heterogeneous elements. This review mainly aims to critically analyse literature developments in machine learning, deep learning, graph models, and Explainable AI paradigms, and identify an emerging requirement for hybrid architectures incorporating Graph Neural Networks and Explainable AI for developing financially sound solutions with optimized paths in financial systems.*

Keywords - *Credit Card Fraud Detection, Explainable AI, Graph Neural Networks, Machine Learning, Transaction Networks.*

1. Introduction

The increased usage of digital financial services has led to a revolution in the way people and institutions make financial transactions on a daily basis. Credit card usage, in particular, has become an important factor in facilitating online payments and online shopping. As a result, the usage of credit card financial transactions has significantly increased. Although this has led to increased accessibility, it also provides increased opportunities for people to commit financial fraud. Present-day fraud rarely occurs in isolation; it often involves adaptive and coordinated strategies such as identity manipulation, device sharing, synthetic profiles, and multi-account collusion, all of which make fraud detection a difficult challenge for both researchers and industry practitioners [1, 2].

One major challenge lies in the nature of real-world datasets. The fraudulent transaction accounts for less than a few percent of all other transactions, making it a class imbalance problem with a significantly reduced performance of traditional classifiers being biased towards the majority class [1, 3]. Moreover, scammers keep updating their methods, which demands adaptive fraud detection systems capable of handling dynamic environments rather than fixed rules or predefined features. A third major challenge is that a transaction is not an independent event. Users, devices, merchants, IP addresses, and geolocations exist in a graph representing a network of inter-linked relationships, in which a major part of them have a definitive impact on a transaction being considered suspicious or not [3, 9].

In the last decade (see Figure 1), work on credit card fraud detection research has progressed with different levels of methodologies. While machine learning algorithms such as logistic regression, decision trees, random forests, SVMs, and boosting algorithms established a solid groundwork with their interpretability and efficiency [1, 3], deep learning algorithms which include LSTMs [7], CNNs [3], and autoencoders [11], have improved these capabilities by capturing temporal and nonlinear relationships from the data without relying on manual feature construction. However, both machine learning and deep learning methods tend to focus on individual samples rather than network behaviour.

Graph Neural Networks (GNNs) have emerged as a very promising approach by representing financial environments as graphs consisting of users, cards, devices, merchants, and transactions, as it is possible for Graph Neural Networks to detect both local graph behaviour and graph irregularities which are characteristic of organized and multi-entity fraud attacks [4-6]. A significant amount of research has demonstrated that graph-based approaches outperform classical ML and sequential DL models in environments where relational information is important [4, 9, 15].

Although they have shown promising performance, they pose a challenge in terms of two factors: interpretability and scalability. Their message passing schemes, which are performed in graph neural networks and relational reasoning steps, represent a “black box” application since they can hardly be explained in contexts where transparency is a priority. The financial sector operates under a set of regulations that require clear, auditable explanations for adverse decisions, which brings Explainable AI into focus in contemporary research on fraudulent transaction systems utilizing AI to a great extent [7, 12, 28].

Explainable AI (XAI) gained prominence as a solution to address these issues. Techniques such as GNNExplainer, PGExplainer, and counterfactual graph explanations provide explanations for local decision-making, but these methods struggle with a lack of support for heterogeneous graphs and difficulty in providing a consistent transaction-level explanation [7, 8, 12].

The existing literature also highlights several unresolved issues. Limited research work is available on hybrid architectures combining machine learning, deep learning, and graph-based reasoning. Most XAI techniques have not been critically evaluated on highly imbalanced and dynamic fraud datasets. Graph benchmarks are less common, scalability is a problem when it comes to real-time solutions, and some algorithms trade off accuracy for interpretability or interpretability for accuracy [14, 19, 30]. These limitations emphasize the need for hybrid fraud-detection systems.

This literature review integrates findings on machine learning, deep learning, graph neural networks, and research in Explainable AI, and critically assesses their strengths and weaknesses in practical financial scenarios, in order to highlight essential limitations in existing methods. It further anticipates directions in future research work by pointing out research trends in these technologically developing domains.

The remainder of this paper will provide a systematic literature review of existing work, with Section 2 reviewing classical ML and DL approaches, Section 3 highlighting graph-based modelling techniques and recent GNN architectures, Section 4 focusing on explainability methods for graph data, Section 5 emphasizing key research gaps, and Section 6 presenting concluding insights.

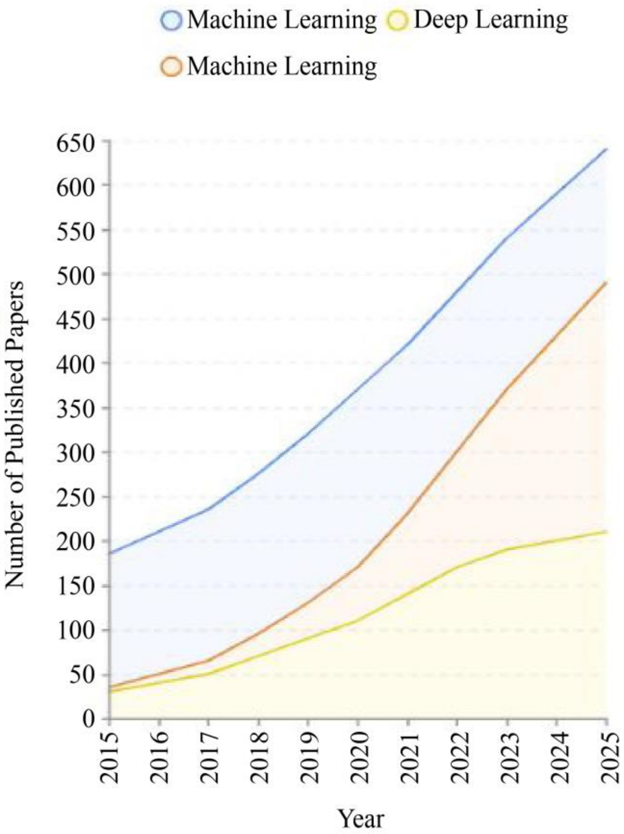

Fig. 1 Research trends in credit card fraud detection (2015–2025)

2. Literature Review

Credit card fraud has become an increasingly puzzling problem for financial organizations, thanks to a rise in e-commerce, a proliferation of devices used for making online payments, and sophisticated tactics used in fraudulent activity. Classical analytics are simply not capable of handling the subtle, dynamically changing behaviour of contemporary fraud, which has encouraged the application of a vastly diverse set of ML, deep learning, and graph solutions to design smarter, more resilient fraud detection systems. This section reviews what has been accomplished, which provides a background for the proposal of the GNN-XAI approach (see Figure 3).

2.1. Machine Learning-based Fraud Detection Models

Early fraud models considered this problem a binary classification problem in a supervised learning setting with a tabular structure. Solutions such as logistic regression, decision trees, random forests, support vector machines, and gradient boosting machines offered a competitive advantage because of their efficiency and interpretability [1, 3].

Furthermore, ensemble methods with feature engineering offered a promising solution. For example, the Neural Network Ensemble with Feature Engineering (IEEE, 2021) [2] illustrated how engineered features at a transaction level can optimize standard machine learning models, while federated learning models such as FedFusion (2023) illustrated that adaptive learning models are efficient in the federated learning process, even with heterogeneous data, with tree-based models still performing comparably [13].

Advantages of ML Techniques

- Rapid training speed. Real-time prediction.
- Easy to audit and compatible with regulatory requirements.
- Good when attributes are descriptive of a domain.
- It can run on relatively low hardware.

Limitations

- It is dependent on human feature engineering.
- Handling highly imbalanced fraud data.
- Frequently overlooks interactions involving users, devices, businesses, and/or transactions.
- The power of generalization to other types of fraud is limited.

2.1.1. Early ML Anomaly Detection & Classical Surveys

In addition, the earlier works demonstrated the limitations of classical machine learning techniques through large-scale empirical evaluations. Bhattacharyya et al. (2011) compared several ML classifiers and concluded that even though tree-based models showed stability under imbalance, the performance of those models deteriorated in identifying fraud patterns with a high degree of overlap [21].

Akoglu et al. (2015) also highlighted the fact that traditional anomaly-detection methods fall short in picking up network-level irregularities, thus relational and structure modelling is required [19]. All of these classical insights are relevant to developing an understanding of why more advanced representations like deep learning and graph-based models became necessary.

2.1.2. XGBoost

In the realm of machine learning algorithms, XGBoost is recognized for its robust predictive capabilities and efficiency with structured data. The role of feature engineering, with the consideration of types, usage, geolocation, and spending, is an essential improvement for fraud detection models. Its ensemble approach is capable of handling nonlinear interconnects of features, thereby ensuring that overfitting is avoided by means of gradient boosting and regularization [1].

The major demerits of XGBoost are that it considers every transaction to be an independent observation, which is highly inappropriate for identifying fraudulent rings, especially when users, merchants, devices, or even networks are interrelated. The interpretability of XGBoost is limited because of its ensemble process.

In order to address these challenges, more interpretable frameworks of boosting, such as iXGB (see Figure 2), have been investigated, which is used to derive human-interpretable rules from XGBoost. Even so, iXGB is still incapable of handling the dependencies on a network level, which is still significant when it comes to sophisticated fraud analysis, thereby contributing to the development of deep learning solutions that are capable of learning more expressive representations [6].

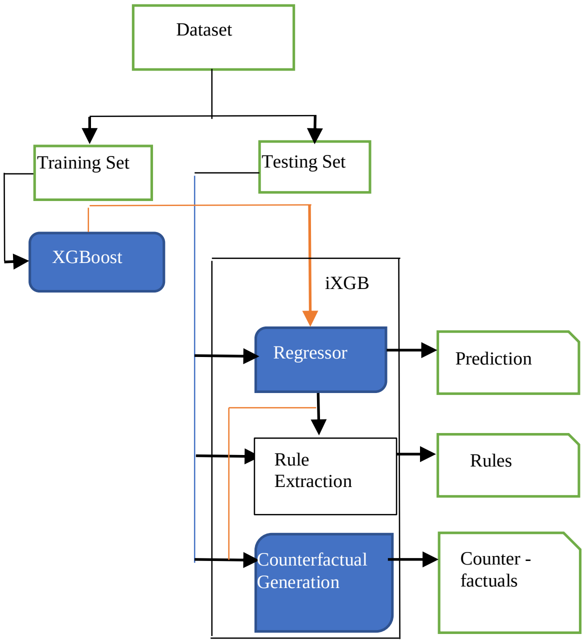

Fig. 2 Overview of the mechanism of existing iXGB [6]

2.1.3. XGBoost Modern Enhancements

Recent research works on explainable trees have also identified other ways boosted models can be made more transparent via global surrogate trees, monotonic constraints, and rule-based compression techniques [16, 29]. These extensions improve acceptability for regulatory needs but do nothing to alleviate the problem of XGBoost's inability to model multi-entity relationships that remain critical in fraud networks.

2.2. Deep Learning Techniques

The increasing amount of digital data led to the popularity of deep learning in fraud detection because deep learning is capable of learning representations on its own, thus eliminating the need for human-intervention-intensive feature engineering. The most common types of deep learning models are autoencoders [11], Recurrent Neural Networks (especially LSTMs), and Convolutional Neural Networks (CNNs) [3], [7].

2.2.1. LSTM Networks

The Long Short-Term Memory network performs efficiently on sequence patterns, such as temporal dependencies, within transaction records, which might not be identified by orthodox models, such as sudden bursts of spending, irregular timing, geolocation shifts, and so on.

RaKShA: A Trusted Explainable LSTM Model to Classify Fraud Patterns on Credit Card Transactions is research indicating that LSTMs can be made more reliable by applying the attribute of explanation, as illustrated in 2023 [7].

Strengths

- Strong skills in time-dependent problems.
- It identifies anomalies in the purchase trend sequence.
- Noise robust, capable of handling nonlinear dynamics.

Limitations

- It models behaviour as a sequence, thus overlooking inter-entity relationships (user-device-merchant).
- Depends on continuous transaction histories; it cannot work with a fragmented list.
- Not capable of detecting cross-user or cross-device fraud, such as an attack.
- Interpretability still has to rely on external XAI solutions

Such problems led to more expressive models for handling multi-relational patterns. Thus, though LSTMs are very efficient in modelling temporal behaviour, they are weak at modelling structural/network-based anomalies, which are characteristic of complicated fraud patterns.

2.2.2. Autoencoders and Anomaly Detection

Autoencoders (AEs), Variational Autoencoders, and Denoising Autoencoders have been extensively researched for unsupervised fraudulent activity detection, learning a compact description of regular activity, and identifying deviations as potential fraudulent activity [11].

New research that combines autoencoders with a graphical structure and real-time analysis (2024–2025) indicates enhanced anomaly detection through the learning of latent transaction embeddings [18].

Strengths

- No need for large labelled fraud datasets.
- Model subtle, nonlinear behaviour anomaly patterns.
- It is resilient against changes in transaction distribution.

More sophisticated autoencoder models have also been explored for financial fraud detection. In this application, deep variational models are demonstrated to be effective at identifying small anomalies in datasets with regard to normal spending patterns in large-scale datasets [18].

Limitations

- High false positives because of noise sensitivity.
- Difficult to tune into dynamic transaction streams.
- Still lacks the incorporation of entity association (user-device association).
- Explanations for anomalies identified by an AE are typically non-intuitive and not very regulator-friendly.

2.2.3. Hybrid Deep Models

In some research, deep learning is used with ensemble learning to obtain improved results. For instance, the combination of LSTM with XGBoost, where XGBoost is used as a classifier with temporal embeddings from LSTMs, has been reported to result in improved detection because it combines the strengths of learning sequences with the power of gradient boosting [8].

However, such hybrids remain mainly at the level of combining features, hence losing the relationally defined structure of the entities, thus not being able to identify clusters of fraud that are only suspicious from a relational perspective.

2.2.4. Deep Learning for Time-Series and Anomaly Detection

Recent deep learning surveys on anomaly detection in sequential financial data have shown that temporal CNNs, sequence autoencoders, and Transformer-based models are more effective in learning long dependencies than the earlier LSTM-only architectures. Choi et al. (2021) and Pang et al. (2021) summarized that while DL models excel in discovering nonlinear deviations, they still struggle when anomalies arise from cross-entity relationships rather than temporal characteristics [20].

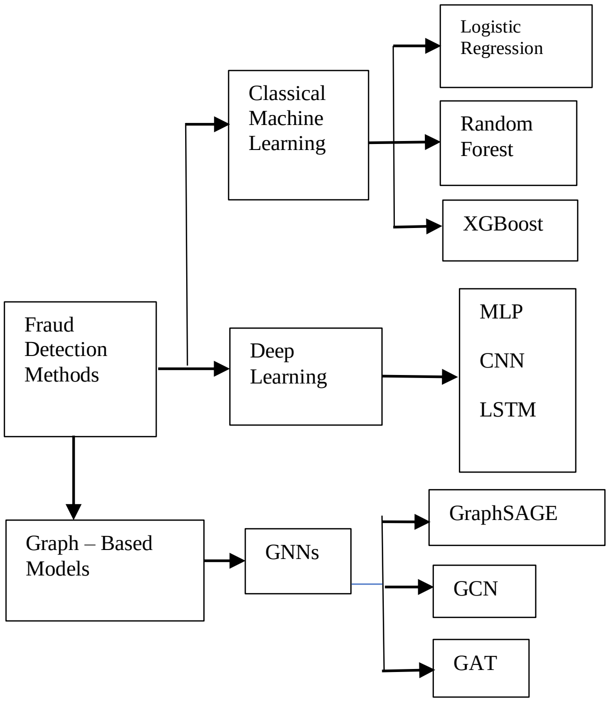

Fig. 3 Traditional fraud detection methods / models

2.2.5. Generative and Synthetic-Data Approaches

Along with discriminative models, there has been a lot of research examining generative architectures to handle extreme imbalance in fraud datasets. GAN-based methods and synthetic data augmentation using Gaussian noise have been applied to enrich minority-class samples [10, 12]. These improve recall but often introduce unrealistic patterns and inflate false positives. The generative models do not incorporate multi-entity dependencies, making their application particularly limited in network-structured fraud scenarios.

2.3. Graph-Based and Hybrid Learning Methods

In financial networks, a common reason for fraudulent activity is the interconnectedness of activity, such as shared accounts, shared devices, and shared transaction paths. Conventional ML/DL approaches consider transaction records as discrete, disregarding such interconnectedness.

The use of Graph Neural Networks (GNNs) has been proposed to interconnect different types of nodes, such as users, merchants, IPs, and devices, modelling interdependencies. It has been observed that GNNs are capable of diffusing information from interconnect nodes, implying that a transaction's probability of being fraudulent can be discerned from its "relation context." Variations such as FraudGNN-RL use reinforcement learning to set thresholds dynamically, enhancing resistance against class imbalance problems [4, 5, 15].

However, GNNs pose certain challenges:

- High computational cost and longer training times.
- Difficult to interpret predictions because the reasoning takes place within a relational latent space.
- Dependence on the quality of the graph, such that poor connectivity affects performance.

The reasons for such vulnerabilities are still unidentified, but some investigators recommend that GNNs be combined with XAI in an effort to retain high performance with interpretability.

2.3.1. Knowledge-Graph + GNN Approaches

A recent work combines GNNs with knowledge-graph structure, enabling models to leverage semantic relationships across financial entities.

Studies in supply-chain finance (2024) showed that combining heterogeneous relational graphs with knowledge-based reasoning significantly improves the detection of multi-hop hidden fraud [17].

This demonstrates that practical fraud scenarios benefit from both statistical learning and domain-driven relational structure.

2.3.2. Large-Scale GNN Training Techniques

To address scalability issues, several large-scale training methods such as FastGCN and ClusterGCN have been proposed [25, 26]. These methods reduce computational overhead through importance sampling and graph partitioning. Although they mainly focus on general graph learning, they provide a way towards real-time application of a GNN-based fraudulent transaction detector in high-volume environments [27].

2.4. Explainable AI (XAI) in Fraud Detection

To be useful, financial AI systems need to be accurate as well as understandable. XAI is a research area that focuses on ensuring that the predictions generated by machine learning models are understandable by auditors and regulatory authorities [7, 8, 12].

The existing XAI techniques are SHAP, LIME, counterfactual explanation, saliency visualization for neural nets, as well as rule extraction platforms such as iXGB [6], which generates understandable rules to interpret XGBoost.

The XAI technique is useful in ensuring that deep learning models overcome the "black box" problem.

But, XAI has challenges such as:

- Most tools are created with table or sequential data, but not graphs.
- Post-hoc hypotheses might not accurately convey the reasoning process of the model.
- High computational cost when scaling.
- They still find it difficult to define relational fraud indicators, such as device-sharing behaviour or clusters of collusion.

Current literature on financial XAI states that financial domains have increased challenges in making an explanation understandable to a human. Hashemi et al. (2023) reported that the majority of financial institutions require explanations that are in accordance with human reasoning used by fraud analysts [28]. These demands are not well met by the current post-hoc graph explanation methods.

This has led to the emergence of more recent work on GNN-specific explanation models, which are described in Section 4.

2.5. Summary of Gaps

The literature review reveals a list of gaps that drive this research:

- Classical machine learning models, such as XGBoost, do not take into consideration the relational ties that exist between different entities.
- The DL models (LSTM, autoencoder), though identifying temporal or feature-level anomalies, are incapable of identifying structural anomalies [3, 7].

- The use of hybrid DL+ML models enhances performance but is targeted at tabular data.
- Although GNNs are relationally powerful, they lack transparency and have high computational costs [4, 5, 27].
- Explainability is still fragmented on the architecture side, sometimes bolted on rather than baked in [7, 12, 28].
- A unified model which incorporates tabular, sequential, and graph-level thinking under a single architecture has not been explored in existing literature [22, 24, 30].
- Dynamic and real-world financial datasets with multiple sources have insufficient benchmarking, which results in limiting reproducibility and comparability across studies [30, 27].

Owing to these challenges, there is a need to shift towards a uniform GNN architecture that has inherent mechanisms, which is discussed in Section 3.

3. Graph-Based Modelling and GNN Architectures for Fraud Detection

The financial transaction is a network phenomenon that is relational by definition. People use the same devices, devices interact with multiple accounts, businesses interact with different sets of customers, while fraudulent behaviour can grow from interactions along these relationships. In the classical approach, each transaction is a separate, isolated event. The use of graphs brings these connections to the forefront, allowing us to identify now fraudulent behaviour that is meaningfully structural. In this section, we will examine how graphs are created, discuss major Graph Neural Network architectures used for fraud modelling, as well as the development and evolution of graph intelligence in finance.

3.1. Graph Representation in Fraud Analysis

3.1.1. Why Graphs?

Fraud rarely occurs on its own [4, 5]. The same device may be used on multiple accounts, the same IP address or location may show up in multiple cases, a single merchant may be used in quick successive thefts, and multiple people may work together on organized crime.

The problem with treating cases in tabular format is that it ignores such shared patterns. This is solved with graphs because it allows us to identify:

- Relational Dependencies (user-device-merchant links)
- Propagation effects (Risk carried through shared nodes),
- Subgraph Anomalies (suspicious communities, hub-shaped structures).
- Multi-hop reasoning (user → device → merchant → other users)

All these complexities cannot be modelled by LSTM [7], XGBoost, and AutoEncoder models [19].

3.2. Types of Graphs in Financial Systems

3.2.1. Homogeneous Graphs

Early fraud research representations used graphs, which considered all nodes to be of the same type (for instance, only users), with links connecting similar properties. This, although computationally efficient, is semantically impoverished, as it is not capable of capturing variations in different types of entities [9].

3.2.2. Heterogeneous Graphs

Contemporary fraud detection relies on heterogeneous graphs with multiple types of nodes and edges, typically including:

- User nodes
- Device nodes
- Transaction Nodes
- Merchant nodes

Edges define actions such as "user makes a transaction" or "device used for transaction." It has been observed that heterogeneity increases fraud detection capabilities as it supports more robust message passing across different entity types [4, 5, 30], which is more similar to a real financial environment.

Other studies have emphasized additional benefits in identifying structural anomalies with heterogeneous graph modelling over homogeneous graph modelling, especially in situations where different device identification or common merchant relationships are used to facilitate fraudulent activities [23, 24].

Recently, research studies have also begun to explore the integration of knowledge graphs [17] with GNNs to enhance the relational context of financial ecosystems. For instance, Feng et al. showed that it is far more effective to combine heterogeneous entities with a knowledge graph topology in identifying multi-hop fraudulent relations in supply chain finance tasks.

3.2.3. Dynamic Graphs

The transactions are temporal, with evolving fraud patterns. The temporal information in dynamic GNNs helps identify temporal collusion or bursty fraud [15]. Such models are more computationally intensive but are closer to actual transaction patterns in financial systems.

3.3. Evolution of GNN Architectures in Fraud Detection

3.3.1. Graph Convolutional Network (GCN)

The neighbours are used by means of spectral or spatial convolutions for gathering information from neighbours. Initially, fraud systems recognized the use of GCNs to propagate labels over neighbouring nodes, thereby enhancing the detection of suspicious clusters [9].

The problem with the use of GCNs arises in dealing with:

- varied types (Heterogeneity) of nodes,
- edge-level classification,
- deep layer over-smoothing.

Despite these limitations, GCN has laid the foundation of message-passing algorithms.

Comparative studies performed among multilayer GCN and attention-based variants prove that although deeper layers capture richer neighbourhood behaviour, they face over-smoothing and increased complexity [22, 24]. As a consequence, a trade-off between representational depth and operational feasibility in fraud settings is created.

3.3.2. Graph Attention Network (GAT)

GATs introduce attention weights on top of the aggregation of neighbours, giving the chance for the model to focus on notable neighbours. This is useful in fraud detection, where some links are more vulnerable, certain nodes are more suspicious, and attention is capable of pointing to anomalous structures.

Still, the computational cost of attention is not scalable enough on large graphs of transactions, thereby restricting its deployment in real-time applications.

3.3.3. GraphSAGE (Sampling and Aggregation)

GraphSAGE introduced inductive learning and neighbourhood sampling, which made it possible to train on large transaction graphs. Generalization to unseen nodes makes it useful for modelling dynamic user and device communities. Recent works in finance have adopted GraphSAGE as it is useful for:

- Processing millions of transactions
- Strong aggregation mechanisms (mean, LSTM, pooling).
- Incorporation of new users/devices without retraining.

These properties make GraphSAGE a useful baseline for relational fraud detection [4, 5].

3.3.4. Heterogeneous Graph Networks (HGNN, HAN, HeteroConv)

Heterogeneous GNNs describe interactions between multiple entities. The Heterogeneous Graph Attention Network (HAN), as well as the HeteroConv implementation within PyTorch Geometry, supports multiple types of convolution per edge type.

It is ideal for capturing nonspecific device behaviour, shared card properties, and unseen connections in distinct transactions.

The FraudGNN-RL [5] (2024) improved this approach with a reinforcement learning component to address sampling context-aware subgraphs.

3.4. Edge-Level Fraud Classification

Fraud detection may concentrate on edge classification (transactions), which are either fraudulent or legitimate, and not on nodes, as in the node-level Fraud Classification. This means that it is necessary to aggregate embeddings from multiple entities:

- User embedding represents historical behaviour and connections.
- Device embedding captures device-level risk.
- Transaction embedding is used for embedding transaction features.
- Edge attributes capture deviations from behavioural norms.

The current models involving GNN aggregate these representations together with the use of an MLP classifier to predict the likelihood that it is a fraud phenomenon [4, 15]. This edge-centric design reflects how actual fraud detection in banking systems operates.

3.5. Strengths and Limitations of GNN-Based Fraud Detection

3.5.1. Strengths

- Captures multi-hop relational fraud patterns.
- Detects coordinated fraud rings.
- Processes expressive, structure-aware embeddings.
- Handles imbalanced datasets by propagating the risk signals.
- Enables inductive learning for new users/devices.

3.5.2. Limitations

- High memory and computational needs
- Sensitive to noisy or incorrectly recorded edges
- Difficult to implement in real-time settings
- Explainability is relatively weaker when compared to traditional ML models [28].
- Depends on the quality of the graph constructed.

Such challenges constitute a reason for integrating methods from the explainable AI frameworks, which are introduced in Section 4.

4. Interpretation Techniques for Graph-Based Fraud Detection

The environments in which fraud detection systems are deployed are high-risk environments where decisions must be transparent, accountable, and defensible in a court of law. Graph Neural Networks (GNNs) have shown excellent performance in the discovery of complex relational fraud patterns, and they also introduce significant translucency. Their message-passing workflows, multi-hop aggregations, and mix of different relationship types make them hard to interpret. This lack of transparency is in sharp contrast to regulatory demands like PSD2, GDPR, and RBI norms that require explainability for adverse financial decisions [7]. Recent conceptual discussions on GNN-based fraud detection have also been presented in preprint literature [32].

This section surveys interpretability methods for graph-structured data, highlights their relevance to fraud detection, and identifies the limitations that push us toward more targeted, transaction-level explanations.

4.1. The Need for XAI in GNN-Based Fraud Detection

While classic models, like Logistic Regression, XGBoost, or Decision Trees, provide human-readable insights such as feature importance, decision paths, or explicit rules, the added complexity of fraud detection shifting to a graph-based frame will increase the complexity of the risk reasoning:

- A user could appear to be legitimate, but due to their ties to known fraudsters, they may look suspicious.
- A device could propagate risk across multiple accounts.
- A group of merchants might be involved in fraudulent collusion.
- The anomalies in subgraphs may point toward a coordinated attack.

These are signals that need contextual interpretation.

Without XAI, financial institutions cannot explain:

- Why a certain transaction was flagged,
- Which are the relational dependencies that influenced the model
- Which nodes or edges have contributed to the risk propagation
- How multi-hop reasoning led to the fraud decision.

Thus, XAI becomes essential not just for trust, but also for making the model usable in operation.

4.2. Categories of Explainability in Graph Models

A consolidated overview of these categories, their core idea, strengths, and limitations is presented in Table 1.

4.2.1. Post-Hoc Local Explanations

Local methods explain individual predictions, such as why a given transaction is labelled fraud. They are necessary because fraud decisions are per-transaction and explanations must be case-specific, human-interpretable, and actionable for analysts. Leading local GNN explainers include GNNExplainer, PGExplainer, SubgraphX, and GraphMask [28, 30]. They all aim to identify the smallest set of nodes, edges, and features that most influenced a prediction.

4.2.2. Global Explanations

Global interpretability describes the general model behaviour, such as:

- How risk propagates through the graph
- which types of nodes are driving the decisions,
- Generally, the importance of user–device versus device–merchant relationships.

While useful for model validation and debugging, global explanations are less useful for auditing individual transactions.

4.2.3. Node-Level vs. Edge-Level Explanations

Most of the explainers in GNNs focus on node classification, while fraud detection is all about classifying edges. Explanations to the following must be highlighted by the explainers:

- For any given transaction edge, which user–device–transaction links influenced risk?
- What other edges, for example, prior use of the device, supported suspicion?
- What multi-hop neighbourhood patterns were most indicative?

This requires edge-centric explanation models.

4.3. Why Classical XAI is Insufficient for Graph Fraud Detection

Despite the value of existing explainers, financial fraud introduces special complexities:

- Heterogeneous, multi-relational graphs-most explainers assume homogeneous graph structures. Real finance, however, involves many different node and edge types.
- Edge-level classification: fraudulent activities sit on the transaction level, not only on nodes; many XAI methods are node-centric.
- Data is highly imbalanced; explanations can therefore be biased toward the majority class.
- Regulatory requirements - explanations must be human-understandable, not just mathematically valid.
- Scaling and real-time requirements: The explainers need to produce fast results on live alerts.

These challenges are highlighted in the recent and emerging studies of explainability for financial GNNs [6, 28, 30],

These gaps motivate the development of transaction-specific, edge-focused strategies, which include

- Personalized subgraphs
- Edge-removal sensitivity analysis,
- Per-transaction behavioural deviation analysis
- Relational anomaly scoring.

This underpins the push to develop heterogeneous PG-style explainers tailored for fraud-detection scenarios.

4.4. Summary

The section highlights how explainability frameworks evolved from simple feature-attribution methods to sophisticated graph-based explainers. Though much more powerful, current methods struggle with the unique demands of the credit-card transaction graph, like heterogeneity, scale, real-time performance, and regulatory expectations.

These limitations indicate that specialized, transaction-level XAI methods need to be developed and integrated directly into GNN-based fraud systems, thereby opening up ways for identifying the research gaps in Section 5.

Table 1. A review of existing GNN explainability methods

GNN Explainability Methods	Core Idea	Strengths	Limitations
GNNExplainer	Identifies a minimal subgraph and key node features that maximize the model's prediction	-Produces a human-readable subgraph. -Effective for small/ medium graphs.	- Requires training in a mask generator network. - Risk of masking critical edges.
PGExplainer	Learns a parametric explainer that maps node embeddings to edge importance scores.	-Better scaling than GNNExplainer. -Works effectively with evolving graphs -Supports multi-hop reasoning.	- Requires separate training from the base GNN - Does not inherently support heterogeneous graphs unless extended [30]. - May select mathematically influential but intuitively confusing edges.
SubgraphX	Uses Shapley-value-based evaluation to identify subgraph contributions to predictions	-Provides theoretically grounded attributions -Can highlight suspicious transaction motifs.	-Highly computationally intensive. - Produces multiple candidate subgraphs → harder interpretation.
GraphMask	Mask edges to identify which relationships the GNN actively uses.	-Much accurate in detecting influential edges. -Performs well in sparse fraud graphs.	- Requires training in a mask generator network. - Risk of masking critical edges → may destabilize predictions

5. Research Gaps and Open Challenges

Despite the evident success in fraud detection via machine learning, deep learning, and graph models, there are still challenges that limit the effectiveness, deployability, and interpretability of these systems in real-world financial settings. The incorporation of the literature cited discusses the essential gaps that exist (see Figure 4).

5.1. Inadequate Modelling of Multi-Entity Relationships

Traditional models such as Logistic Regression, SVMs, RF, and XGBoost (Gradient Boosting) operate on a transaction as a single, isolated instance, with independent features [1, 2]. In reality, fraud is a relational phenomenon of joint behaviour exhibited by users, devices, merchants, and IPs.

Autoencoders and LSTMs identify non-linearities and temporal dependencies, respectively, but still fail to encode cross-entity structures or relational dependencies and disregard inter-entity relationships [3, 7]. Graph Neural Networks(GNNs) mitigate this problem by modelling users, devices, transactions, and merchants as graphical interconnected components [4], but are limited by difficulties in constructing the graphs and scalability issues.

Present studies do not address in depth how knowledge graphs can be integrated with GNNs for financial fraud, although early evidence indicates that they can identify hidden semantic schemes of financial fraud [17].

The consensus is that most existing models are pivoted on a tabular structure, with significant blind spots in relation to fraudulent rings, device-sharing, and collusion.

5.2. Limited Interpretability in Existing Models

High-performing models such as XGBoost [2], DNNs, and GNNs are largely black-box models, which go against the regulations (GDPR, PSD2, RBI guidelines) that require a human-oriented interpretation of fraud detection.

Although models such as iXGB that provide rule-based explanations are available, interpretability is still not adequate for fraud detection models based on GNNs [6].

Despite the availability of models like RaKShA (LSTM-XAI) [7], edge-level, graph-based interpretability remains insufficient, particularly for edge-level fraud decisions where analysts require clear, transaction-specific reasoning.

5.3. Concept Drift and Evolving Fraud Patterns

Fraudsters adapt behaviours to evade detection, making static models less accurate with time. It is reported that changes keep happening quickly for the distribution of features as well as the relationship patterns [3, 4].

Although models such as ensembles and periodic training can mitigate the drift problem, there is a lack of models that address the temporal modelling of user, device, and merchant relationship changes.

Although dynamic Graph Neural Networks (GNNs) are promising, they are very computation-intensive and rarely deployed in production environments. There is a need to research online, adaptive, or temporal GNNs for fraud detection, which are capable of learning the changing behavioural patterns continuously [27].

5.4. Class Imbalance in Fraud Datasets

Fraud cases constitute less than 1% of transactions, resulting in:

- High overall accuracy but low fraud detection.
- Biased decision boundaries.
- Challenges to learn minority-class patterns [1, 10].

Sampling-based approaches, such as SMOTE and GAN-based approaches [10, 12], overcome performance issues but result in noise or overfitting.

Graph models are capable of efficiently diffusing signals from the minority class [4], but there is no consensus on how to train imbalance-aware GNNs. Few works explore class-weighted message passing, minority-oriented attention, or robust representations.

Generative augmentation techniques address imbalance only partially, but they create synthetic anomalies that do not always generalize across banks or transaction channels [10, 12]. There is a critical need for imbalance-aware graph learning mechanisms that avoid majority-class dominance.

5.5. Lack of Real-Time and Scalable Detection

Most such research is offline, based on static inputs [1-4], while finance requires decisions on a millisecond scale for millions of events per day.

The real-time intervals make it even harder to apply:

- Deep learning models that require sequential inference
- Autoencoders require reconstruction for each event
- Multi-hop neighbourhood aggregation in GNNs
- Dynamic updates of the graphs used in temporal GNNs.

Recent studies propose distributed inference for GNNs, a technique that has been considered for a stream, but has been largely untouched on an industrial scale.

Large-scale GNN architectures, such as FastGCN and ClusterGCN, have solutions, but they are yet to be adapted specifically for fraud workflows involving streaming updates [25, 26].

Scalability and real-time inference represent one of the largest gaps between academic research and industrial deployment.

5.6. Limited Research in Graph Explainability (XAI for GNNs)

While classical XAI frameworks (SHAP, LIME, Counterfactuals) provide strong interpretability for tabular data [6], they do not support heterogeneous graph structures.

Graph-specific explainers such as GNNExplainer, PGExplainer, SubgraphX, and GraphMask offer local interpretability [28], but are limited by:

- High computational requirements for processing large graphs
- Limited support for heterogeneous nodes/edges
- Low real-time performance
- Difficulty in aligning with human reasoning, which is used by the fraud analysts.

New developments in Knowledge-graph-based GNNs and encoder–decoder GNNs are highlighting promising leads [17], but explanation mechanisms for these architectures remain early in development.

No existing GNN explainer provides simultaneous local plus global explanations specifically for finance, which is a key regulatory requirement [28, 30].

There is a strong need for edge-level, heterogeneous, and real-time GNN explainers for financial decision systems.

5.7. Additional Emerging Gaps

Recent works on generative fraud detection [10, 13], crypto-based graph fraud [11], and knowledge-graph financial systems [17] show further constraints:

- Generative models lack realism
- Unsupervised models lack generality, especially when generalized over multiple financial sectors.
- Lack of models that realize a learning architecture that integrates table, sequence, and relation learning within a single architecture.

These highlight opportunities for hybrid GNN–ML architectures and cross-domain fraud frameworks. A consolidated overview of these gaps is presented in Table 2.

Table 2. A summary of identified gaps in the classical methods

Gap ID	Research Gap	Impact	Future Direction
G1	Inadequate modelling of multi-entity relationships.	Fails to detect coordinated fraud patterns.	Adoption of graph-based relational learning.
G2	Limited interpretability of deep/ensemble models.	Hinders regulatory compliance and analyst trust.	Integration of XAI with GNNs.
G3	Concept drift and evolving fraud patterns.	Model degradation over time.	Temporal and adaptive GNN frameworks.
G4	Dataset Imbalance.	Low recall and biased decisions.	Imbalance-aware graph attention mechanisms.
G5	Lack of real-time, scalable systems.	Delayed fraud response.	Stream-based and distributed GNN inference.
G6	Limited Graph Explainability Research.	Limited transparency for relational reasoning.	Edge-focused heterogeneous explainers

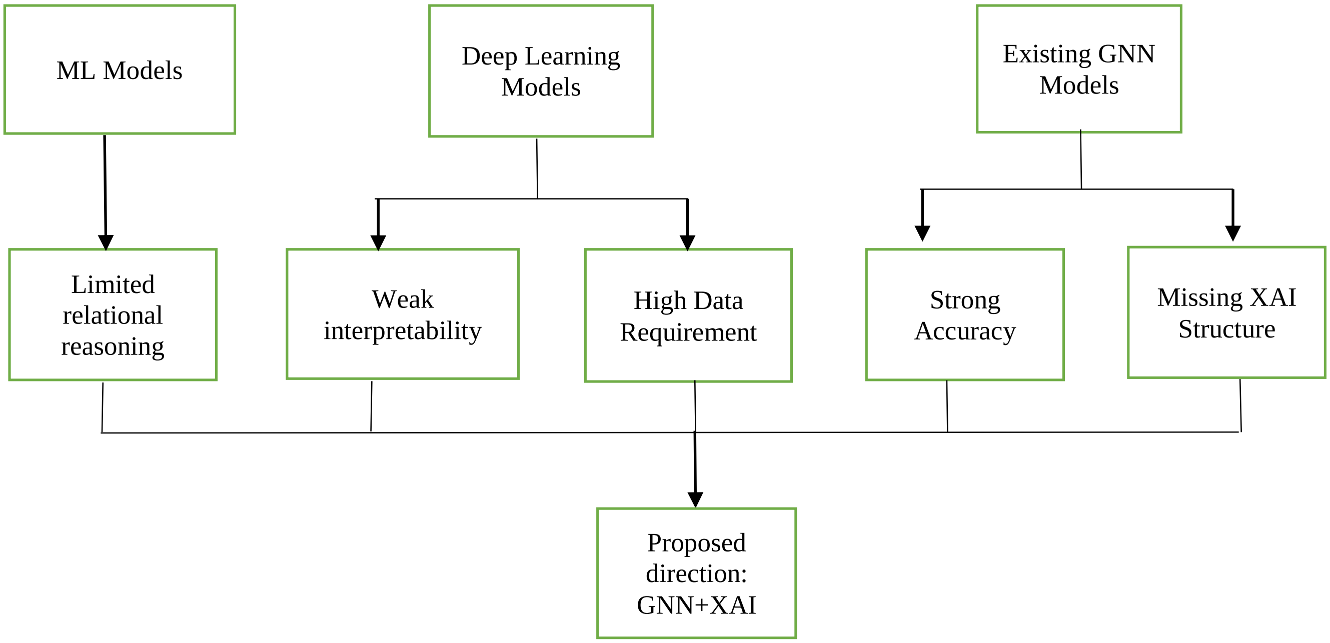

Fig. 4 Current research gaps in fraud detection

5.7.1. Overall Observation

The literature shows a progression from classifiers based on feature learning to deep relational thinking, but existing solutions struggle with explainability, imbalance, and real-time interpretability. However, filling these research and application gaps with graph-based XAI solutions can provide a way forward towards an adaptive and production-ready system for fraud detection.

6. Conclusion

There has been a definite transition in credit card fraud identification solutions from straightforward rule-based methods to more advanced learning algorithms, which are capable of identifying intricate behavioural patterns. This review highlights how classic machine learning methods offer reliable and interpretable baselines while still facing two core limitations: that these methods rely on handcrafted features, and they fail to model interactions across users, devices, and merchants. Deep learning models address several of these limitations by learning richer feature representations and capturing temporal dynamics, but they introduce new challenges in terms of interpretability, data imbalance, and the difficulty of modelling cross-entity relationships.

Graph Neural Networks represent a significant advancement by framing financial transactions as interconnected networks and thus enabling the detection of multi-hop and structural fraud patterns that earlier models frequently overlook. However, despite their promising performance, GNNs face scalability constraints, dependence on the quality of graph construction, and a lack of transparency. Research and studies in explainable AI have begun to address these concerns, but existing XAI tools for graph-based systems remain in their early days of development and are not yet capable of providing transparent, case-specific justifications necessary in critical financial settings.

Put together, the results of this review suggest that no single modelling approach at present meets the combined requirements for performance, adaptability, relational awareness, scalability, and explainability of fraud-detection systems in the real world. The emerging direction points toward hybrid architectures that integrate the structural strengths of GNNs with principled, regulator-aligned explainability mechanisms. This has the potential to deliver both high predictive performance and transparent reasoning, which enables financial institutions to detect fraud more effectively while maintaining trust, accountability, and regulatory compliance.

To conclude, this review underlines the increased demand for fraud-detection frameworks that are not only strong from a technical perspective but also interpretable, scalable, and able to act within the fast-changing conditions in modern financial ecosystems. Moving in this direction, one will be able to develop the next generation of fraud-detection solutions that can satisfy modern operational and regulatory requirements.

6.1. Study Limitations

Though providing an extensive compilation of research on credit card fraud detection by machine learning, graph neural networks, and explainable AI models, this research has some limitations that need to be acknowledged.

To start, the nature of this study, being review-based, inherently poses limitations on its scope. The study does not focus on designing new models for fraudulent act detection. Rather, the comparative analysis and conclusions drawn herein result from reviews of studies already conducted and published by other researchers. The experiment settings and measures used for determining performance trends observed herein, therefore, rely on those adopted by the original authors.

Second, the findings of this review are dependent on the accuracy, completeness, and reporting quality of the existing literature. The fact that there are differences in the studies with regard to the nature of the dataset, the methods used for dealing with the problem of class imbalance, as well as the methods for the evaluation, makes a fair comparison of the techniques difficult. This heterogeneity may influence the generalizability of some observations presented in this work.

Third, credit card fraud detection as well as graph-based machine learning interpretability are rather new research domains. With new architectures, optimization strategies, and methods for interpretability being introduced from time to time, there may be a chance that certain recent developments are not captured in the scope of this review. Hence, the identified trends in this study may be regarded as a manifestation of the available literature at the time of the analysis rather than as definitive.

Furthermore, the evaluation of explainability AI tools in fraud analysis is hindered by the lack of common, universally applicable evaluation indicators. Many reviewed studies rely on qualitative or case-based explanations, which depend on human interpretation and contextual understanding. There is, therefore, a need to establish universal evaluation guidelines with respect to explainability, especially if it is to apply in areas such as financial fraud.

Future work could consider leveraging this research by creating comprehensive benchmarking frameworks that combine tabular, sequential, and graph representation types in one assessment scenario.

Additionally, there is a need for further research into graph neural networks that scale in real-time, coupled with inherent explanation methods.

Finally, future research could be aimed at establishing quantitative measures by which explanation quality could be assessed, ensuring improved trust and compliance in fraud detection models.

6.2. Future Scope

Based on the limitations identified in this review, there exist several encouraging directions for future work. Additionally, based on the research gaps outlined in Section 5, some of the promising directions to push forward credit card fraud detection include narrowing the gap between academic works and what the industry wants, while addressing limitations related to graph construction, scalability, interpretability, and the evolution of fraud patterns.

6.2.1. Development of Scalable Heterogeneous and Temporal GNNs

Most works on GNNs now rely on static or at most partially heterogeneous graphs [4, 9]. Real-world financial systems, however, require models that are capable of handling:

- Numerous entity types, which include users, devices, merchants, and IP addresses
- Evolving transaction histories
- Streaming updates in near real time
- Adaptive changes in relations due to concept drift.

Future work must explore temporally aware architectures, such as Dynamic GNNs, Temporal Graph Attention, and incremental graph updates to capture shifting fraud patterns better [15]. These models are to be made scalable for millions of transactions with efficiency in neighbourhood sampling, graph sparsification, and distributed inference.

6.2.2. Advancement of Graph-Based Explainable AI (XAI)

Explainability is currently the most underdeveloped area in graph-based fraud detection. Traditional XAI tools work great for tabular models, but they do not generalize to multi-relational financial graphs.

For future work, the emphasis should be placed on:

- Edge-level explainers that highlight why a transaction was flagged.
- Counterfactual graph reasoning (e.g., "which connection made this transaction suspicious?")
- Subgraph visual summaries catering to fraud analysts.
- Regulatory-aligned justification reports that support compliance requirements.
- Real-time XAI will be able to support live transaction monitoring [16].

Developing interpretable GNN architectures rather than relying solely on post-hoc explanation represents an important research aspect.

6.2.3. Improved Handling of Extreme Class Imbalance

Fraud datasets are highly imbalanced, with fraud cases less than 1% in general. Methods such as GAN-based augmentation, variational autoencoders, and synthetic data generation offer partial solutions [10, 13], but risk adding noise or distorting rare fraud patterns.

The future work can be along the following lines:

- imbalance-aware message passing in GNNs
- minority-sensitive attention mechanisms
- cost-aware training losses for edge classification
- oversampling at the cluster level to generate realistic subgraph patterns.
- Multi-task learning fusing anomaly detection with other supervised tasks.

These approaches would allow the model to focus on rare but high-impact fraudulent links.

6.2.4. Hybrid Architectures

Studies provide visible evidence of significant returns when LSTMs' temporal embeddings are combined with the decision boundaries of XGBoost or other ensemble methods [8, 14]. Graph-based reasoning offers structural insights not captured by either approach in isolation [4].

The integration of unified user behaviour, device pattern, and relational topology representations may significantly improve performance while preserving explainability.

6.2.5. Real-Time Deployment Pipelines

The decisions are required to come in milliseconds for fraud detection at a production level. Achieving real-time performance with GNNs remains challenging due to multi-hop aggregation, high memory usage, and graph update overhead. The pipelines developed to address these strict latency needs should be streamlined and efficient.

References

[1] Suraya Nurain Kalid et al., "A Multiple Classifiers System for Anomaly Detection in Credit Card Data with Unbalanced and Overlapped Classes," *IEEE Proceedings*, vol. 8, pp. 28210-28221, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ebenezer Esenogho et al., "A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection," *IEEE Access*, vol. 10, pp. 16400-16407, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ibomoiye Domor Mienye, and Nobert Jere, "Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions," *IEEE Access*, vol. 12, pp. 96893-96910, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Asma Cherif et al., "Encoder–Decoder Graph Neural Network for Credit Card Fraud Detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 3, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yiwen Cui et al., "FraudGNN-RL: A Graph Neural Network with Reinforcement Learning for Adaptive Financial Fraud Detection," *IEEE Open Journal of the Computer Society*, vol. 6, pp. 426-437, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Mir Riyanul Islam, Mobyen Uddin Ahmed, and Shahina Begum, "iXGB: Improving the Interpretability of XGBoost Using Decision Rules," *16th International Conference on Agents and Artificial Intelligence*, Rome, Italy, pp. 1345-1353, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Jay Raval et al., "RaKShA: A Trusted Explainable LSTM Model to Classify Fraud Patterns on Credit Card Transactions," *Mathematics*, vol. 11, no. 8, pp. 1-27, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kianeh Kandi, and Antonio García-Dopico, "Enhancing Performance of Credit Card Fraud Detection Models by Utilizing LSTM Networks and XGBoost Algorithms," *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Fawaz Khaled Alarfaj, and Shabnam Shahzadi, "Enhancing Fraud Detection in Banking with Deep Learning: Graph Neural Networks and Autoencoders for Real-Time Credit Card Fraud Prevention," *IEEE Access*, vol. 13, pp. 20633-20646, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Mohammed Tayebi, and Said El Kafhal, “Generative Modelling for Imbalanced Credit Card Fraud Transaction Detection,” *Journal of Cybersecurity and Privacy*, vol. 5, no. 1, pp. 1-36, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Huang Tingfei, Cheng Guangquan, and Huang Kuihua, “Using Variational Auto Encoding in Credit Card Fraud Detection,” *IEEE Access*, vol. 8, pp. 149841-149853, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Fray L. Becerra-Suarez, Halyn Alvarez-Vasquez, and Manuel G. Forero, “Improvement of Bank Fraud Detection through Synthetic Data Generation with Gaussian Noise,” *Technologies*, vol. 13, no. 4, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Nahid Ferdous Aurna et al., “FedFusion: Adaptive Model Fusion for Addressing Feature Discrepancies in Federated Credit Card Fraud Detection,” *IEEE Access*, vol. 12, pp. 136962-136978, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Asma Cherif et al., “Credit Card Fraud Detection in the Era of Disruptive Technologies: A Systematic Review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 1, pp. 145-174, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ahmad Asiri, and K. Somasundaram, “Graph Convolution Network for Fraud Detection in Bitcoin Transactions,” *Scientific Reports*, vol. 15, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[16] Scott M. Lundberg et al., “From Local Explanations to Global Understanding with Explainable AI for Trees,” *Nature Machine Intelligence*, vol. 2, pp. 56-67, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Wenying Xie et al., “Supply Chain Financial Fraud Detection Based on Graph Neural Network and Knowledge Graph,” *Technical Gazette, Directory of Open Access Journals (DOAJ)*, vol. 31, no. 6, pp. 2055-2063, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] L. Selvam et al., “A Deep Learning Model for Investment Scam Prevention in Financial Systems: Autoencoder-based Anomaly Detection,” *2025 3rd International Conference on Inventive Computing and Informatics(ICICI)*, Bangalore, India, pp. 868-875, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Leman Akoglu, Hanghang Tong, and Danai Koutra, “Graph based Anomaly Detection and Description: A Survey,” *Data Mining and Knowledge Discovery*, vol. 29, pp. 626-688, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[20] Guansong Pang et al., “Deep Learning for Anomaly Detection: A Review,” *ACM Computing Surveys*, vol. 54, no. 2, pp. 1-38, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Siddhartha Bhattacharyya et al., “Data Mining for Credit Card Fraud: A Comparative Study,” *Decision Support Systems*, vol. 50, no. 3, pp. 602-613, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[22] Guoxiang Tong, and Jieyu Shen, “Financial Transaction Fraud Detector Based on Imbalance Learning and Graph Neural Network,” *Applied Soft Computing*, vol. 149, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Feifen Shi, and Chuanjun Zhao, “Enhancing Financial Fraud Detection with Hierarchical Graph Attention Networks: A Study on Integrating Local and Extensive Structural Information,” *Finance Research Letters*, vol. 58, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Uzair Aslam Bhatti et al., “Deep Learning with Graph Convolutional Networks: An Overview and Latest Applications in Computational Intelligence,” *International Journal of Intelligent Systems*, vol. 2023, pp. 1-28, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Jie Chen, Tengfei Ma, and Cao Xiao, “FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling,” *ICLR Proceedings*, pp. 1-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] Wei-Lin Chiang et al., “Cluster-GCN: An Efficient Algorithm for Training Deep and Large Graph Convolutional Networks,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA, pp. 257-266, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Zhao Li et al., “A Graph-Powered Large-Scale Fraud Detection System,” *International Journal of Machine Learning and Cybernetics*, vol. 15, pp. 115-128, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Rahul Kapale et al., “Explainable AI for Fraud Detection: Enhancing Transparency and Trust in Financial Decision-Making,” *Proceedings of the 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)*, Wardha, India, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Christoph Molnar, *Interpretable Machine Learning: A Guide for Making Black-Box Models Explainable*, pp. 1-320, 2020. [Google Scholar] [Publisher Link]

[30] Soroor Motie, and Bijan Raahemi, “Financial Fraud Detection Using Graph Neural Networks: A Systematic Review,” *Expert Systems with Applications*, vol. 240, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Quentin Cappart et al., “Combinatorial Optimization and Reasoning with Graph Neural Networks,” *Journal of Machine Learning Research*, vol. 24, pp. 1-61, 2023. [Google Scholar] [Publisher Link]

[32] Diego Vallarino, “AI-Powered Fraud Detection in Financial Services: GNN, Compliance Challenges and Risk Mitigation,” *SSRN Working Paper*, pp. 1-34, 2025. [CrossRef] [Google Scholar] [Publisher Link]