International Journal of Data Informatics and Intelligent Computing (IJDIIC)
Vol. 5, No. 3, September 2026, pp. 78~113
DOI: 10.59461/ijdiic.v5i3.306

Adversarially Robust and Explainable Graph Neural Networks for Fraud Detection in Dynamic Blockchain Networks: A Problem-Based Systematic Literature Review

Oluwaseun Adeniyi Ojerinde[1], Ramatu Abubakar[1], Ajagun Abimbola Susan[1,2], Enesi Femi Aminu[1]
[1]Computer Science Department, Federal University of Technology, Minna, Niger State, 920101, Nigeria
[2]Department of Energy and Electrical Engineering, Hohai University, Nanjing, 210098, China

Article Info

Article history:

Received July 24, 2026
Revised September 10, 2026
Accepted September 19, 2026

Keywords:

Cryptocurrency Analytics
Graph Representation Learning
Adversarial Resilience
Financial Cybercrime
Temporal Transaction Modelling

ABSTRACT

Graph Neural Networks (GNNs) have emerged as an effective approach for blockchain fraud detection because they can model complex relationships among transactions and participating entities. However, existing GNN-based approaches remain limited by vulnerability to adversarial manipulation, limited explainability, continuously evolving transaction graphs, benchmark-data constraints, and the absence of standardized evaluation practices. This study presents a Problem-Based Systematic Literature Review (PBSLR) of adversarially robust and explainable GNNs for fraud detection in dynamic blockchain networks. The review covers studies published between 2021 and 2026 and searches eight academic databases: IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Wiley Online Library, Scopus, Web of Science, and Google Scholar. A total of 412 records were initially identified, 77 full-text studies were assessed for eligibility, and 73 studies were included in the final systematic synthesis. The review analyses the literature across blockchain fraud characteristics, GNN-based detection, adversarial threats, explainability, dynamic graph learning, benchmark datasets, evaluation practices, scalability, and deployment challenges. The synthesis shows that existing research remains fragmented, with limited integration of adversarial robustness, explainability, and dynamic graph learning within a unified fraud detection framework. The review further identifies persistent limitations involving dynamic and labelled benchmark datasets, adversarial evaluation, cross-platform generalisation, standardized robustness and explainability metrics, scalability, and reproducibility. Based on these findings, a research roadmap is proposed to guide the development of robust, interpretable, adaptive, scalable, and deployable GNN-based fraud detection systems for dynamic blockchain environments.

Corresponding Author: Oluwaseun Adeniyi Ojerinde (e-mail: o.ojerinde@futminna.edu.ng)

1. INTRODUCTION

Blockchain technology has revolutionized digital financial ecosystems by providing a way for transactions to take place without centralized intermediaries, with transparency, and cannot be tampered with [1]. The adoption of blockchain in the world of cryptocurrencies and decentralized finance (DeFi) has brought attention to blockchain's versatility in other sectors such as supply chain management, healthcare, digital identity management, and smart contracts. Although blockchain networks offer inherent security features, it also faces vulnerabilities to numerous forms of fraudulent activity as transactions are pseudonymous (not anonymous), which makes it challenging to detect fraudulent actions while still maintaining user privacy [2]. With the surge of Bitcoin and blockchain usage, criminals are developing more complex and sophisticated means to launder money, conduct phishing attacks, perpetrate Ponzi schemes, pay ransom, manipulate markets, and facilitate unfair fund transfers [3]. While many traditional methods of fraud detection have proven effective in systems without blockchain components, such as rule-based systems and

conventional machine learning algorithms, these methods are not suitable for blockchain because they rely on analyzing a single user transaction as well as handcrafted attributes, which provide less effectiveness in blockchain [4]. In blockchain transactions, a naturally highly interconnected graph structure is formed where the wallets are nodes and the transactions are edges of the graph [5]. Traditional approaches are often inaccurate in detection, lack scalability, and can't pick up on complex fraud cycles with multiple addresses and multiple transactions. To address these challenges, one of the most promising deep learning frameworks for the blockchain domain to detect fraud is Graph Neural Networks (GNNs). Blockchain networks are of immense size, and models like Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), GraphSAGE, Graph Isomorphism Networks (GINs), and temporal GNN models have proved to be very effective at identifying illegal activities, suspicious transactions, and unusual behaviors across large-scale networks [6]. Being able to model the higher-order dependencies allows for more effective detection of fraud patterns that are undetectable by traditional classifiers. Blockchain graphs are constantly updated, with new addresses, transactions, and interactions with smart contracts being added at a rate of one per second [7]. Also, fraud schemes continually change, leading to concept drift, and mitigating models introduced over time become less effective with newly released data. Several recent works have espoused the following directions as important: temporal graph learning, adaptive representation learning, and dynamic graph modeling [8].

Another important concern of GNNs is the susceptibility of Graph Neural Networks (GNNs) to adversarial attacks. Small changes in the graph structure, node features, or the connections of the edges between nodes can vastly affect the performance of the model and be hard to recognize [9]. Malicious nodes can be intentionally added, links to transactions can be altered, a fraudulent account can be disguised within an acceptable community, or fraudulent entries can be added to training sets to go undetected [10]. This type of manipulation poses big security concerns since it would affect the ability of fraud detection efforts in the real world. Besides robustness, another challenge is the lack of explainability for implementing a GNN-based fraud detection system [11]. This uninterpretability makes users less confident in its results, makes forensic investigations more challenging, and adds difficulties to regulatory compliance, especially in anti-money laundering (AML) applications, where financial institutions have to explain automated decisions [12]. In order to increase transparency, some Explainable Artificial Intelligence (XAI) techniques, such as the GNNExplainer, PGExplainer, attention-based explanations, and subgraph interpretation methods, have been introduced, but their integration with adversarially robust GNN frameworks is still insufficient. Encouraged by these research gaps, this study aims to tackle the problem of Adversarially Robust and Explainable Graph Neural Networks for Fraud Detection on Dynamic Blockchain Networks by carrying out a problem-based systematic literature review. In this paper, the authors do not organize the previous works by GNN architectures but rather summarize the most important technical, security, explainability, and deployment challenges reported in the period 2021–2026. The current review outlines the gaps in the existing research, provides analysis of existing limitations, and pinpoints potential research avenues for the future for advancement towards an implementation of trustworthy (robust, interpretable, and adaptive) GNN-based fraud detection systems to provide security in next-generation blockchain ecosystems.

1.1. Novelty and Contributions of This Review

Recent review studies focused on blockchain fraud detection, Graph Neural Networks (GNNs), adversarial machine learning, explainable artificial intelligence (XAI), and dynamic graph learning. While these studies have focused on various aspects such as algorithm development, model performance, or specific application domains, most studies have not extensively explored the broader challenges that stem from the dynamic nature of blockchain networks and that must be addressed to ensure the deployment of trustworthy fraud detection systems [13][14]. This study uses a Problem-based Systematic Literature Review (PBSLR) methodology to critically synthesize the techniques, security, explainability, dataset, evaluation, and regulation issues of adversarially robust and explainable GNNs for blockchain fraud detection, which is a novel approach to previous surveys. This review does not examine specific algorithms but rather summarizes general limitations, research trends, unsolved problems, and gaps in knowledge reported in studies of algorithms published in the last 5 years (2021–2026) [15]. Moreover, this paper combines five related research areas, which are researched separately and closely related to blockchain fraud detection: Graph Neural Networks, adversarial robustness, explainable artificial intelligence, dynamic graph learning, and blockchain fraud detection. The review illustrates how, when combined, these factors impact the trustworthiness, transparency, and usability of blockchain fraud detection systems [16].

The major contributions of this review are summarised as follows:

- This review identifies and systematically analyses the key technical, security, explainability, dataset, evaluation, and regulatory problems involved in applying GNN to blockchain fraud detection.

- Robustness and explainability from a unified perspective: The review includes the two phenomena of robustness and explainability in a single analytical framework and shows that the two phenomena are complementary to build trustworthy blockchain fraud detection systems.
- Critical literature synthesis: This review not only summarises the existing research but also critically analyses the existing approaches by highlighting recurring shortcomings, compromises, developing research trends, and unmet research needs.
- Dynamic blockchain environments are analyzed comprehensively: Challenges in temporal graph evolution, concept drift, scalability, real-time fraud detection, and others, which are not sufficiently covered in the existing review literature, are addressed.
- Research direction for future development: Based on identified research gaps, a research direction for future development is proposed to develop a robust, explainable, scalable, and regulation-compliant GNN-based fraud detection framework for dynamic blockchain networks.

Table 1. Comparison of Existing Review Articles and the Present Review

Review Article	Blockchain Fraud Detection	Graph Neural Networks	Adversarial Robustness	Explainability (XAI)	Dynamic Blockchain Networks	Problem-Based Analysis	Research Roadmap
[13]	✗	✓	✗	✓	✗	✗	✓
[17]	✗	✓	✗	✓	✗	✗	✓
[14]	✗	✓	✗	✗	✓	✗	✓
[18]	✗	✓	✓	✗	✗	✗	✓
[16]	✗	✓	✗	✓	✗	✗	✓
[19]	✓	✓	Limited	Limited	Limited	✗	Limited
[15]	✓	✓	Limited	✗	Limited	✗	Limited
This Review (2026)	✓	✓	✓	✓	✓	✓	✓

Table 1 shows that the current review articles are mostly specific to their respective research areas like Graph Neural Networks, Adversarial Robustness, Explainability, Temporal Graph Learning, or Blockchain Fraud Detection. Although these surveys give abundant information about the fields they cover, each field lacks an integrated problem-based framework considering the following areas: blockchain fraud detection, adversarial robustness, explainability, and dynamic graph learning. The focus of the present review is instead on the practical, security, explainability, and regulatory challenges, dataset issues, and evaluation challenges encountered by trustworthy GNN-based blockchain fraud detection, adopting a Problem-Based Systematic Literature Review (PBSLR) approach. Also, it puts forward an integrated research roadmap that can inform future research toward robust, explainable, scalable, and practical blockchain fraud detection systems.

2. REVIEW METHODOLOGY

2.1. Review Objective and Analytical Scope

The objective of this systematic literature review is to critically examine the challenges affecting the development and deployment of robust and explainable Graph Neural Networks (GNNs) for fraud detection in dynamic blockchain networks. The review focuses on five interconnected analytical dimensions: blockchain fraud characteristics, limitations of graph-based fraud detection, adversarial threats to GNN models, explainability requirements, and challenges associated with dynamic graph learning, benchmark datasets, evaluation, and deployment.

These dimensions provide the analytical structure for the literature search, study selection, data extraction, and synthesis. Rather than organizing the review around individual research questions, the literature is synthesized according to the major problems that influence the trustworthiness, effectiveness, and practical deployment of GNN-based blockchain fraud detection systems.

2.2. Search Strategy

A comprehensive search strategy was designed to identify relevant studies published between 2021 and 2026. Multiple scientific databases were selected to ensure broad coverage of research in blockchain technology, graph machine learning, cybersecurity, artificial intelligence, and fraud detection. The selected databases included IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Wiley Online Library, Scopus, Web of Science, and Google Scholar. These databases are widely recognized as authoritative sources of high-quality peer-reviewed research and have been extensively used in previous systematic reviews on blockchain and artificial intelligence [1][20]. To maximize retrieval effectiveness, Boolean operators and keyword combinations were employed. The search strings were developed based on the key concepts underlying the study, namely blockchain fraud detection, graph neural networks, adversarial robustness, explainable artificial intelligence, and dynamic graph learning.

The principal search string used was shown in Table 2 for the search combination: ("Blockchain Fraud Detection" OR "Cryptocurrency Fraud") AND ("Graph Neural Network" OR "Graph Learning") AND ("Adversarial Robustness" OR "Adversarial Attack") AND ("Explainable Artificial Intelligence" OR "Explainable GNN") AND ("Dynamic Blockchain Network" OR "Temporal Graph Learning"). Additional search combinations were executed to capture relevant studies that may have used alternative terminologies. The search process was conducted between 2025 and 2026 to ensure the inclusion of the most recent developments in the field, as shown in Figure 1 and Figure 2.

Figure 1. Word Cloud analysis regarding Fraud Detection in Dynamic Blockchain Networks

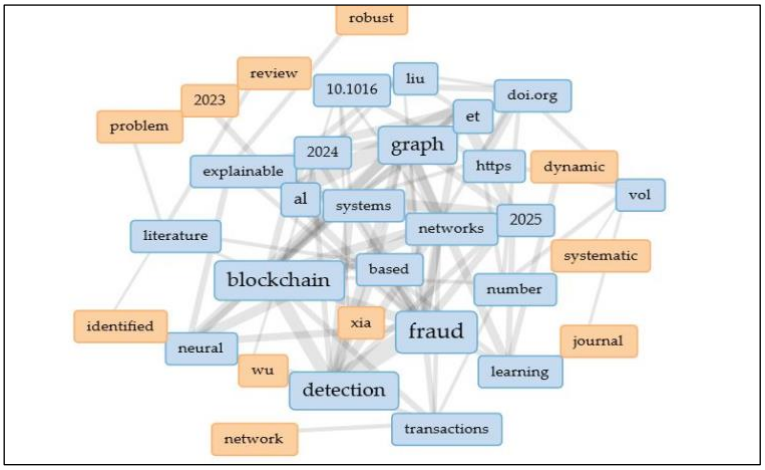

Figure 2. Text analysis visualization of search strings generated using the VOYANT tools

Table 2. Search Strings Used Across the Selected Databases

Search set	Search combination
S1	(“blockchain fraud” OR “cryptocurrency fraud”) AND (“graph neural network” OR “graph learning”)
S2	(“blockchain fraud detection” OR “cryptocurrency fraud detection”) AND (“adversarial attack” OR “adversarial robustness” OR “adversarial machine learning”)
S3	(“blockchain fraud” OR “financial fraud”) AND (“explainable AI” OR “explainable GNN” OR “GNN explanation”)
S4	(“blockchain” OR “cryptocurrency”) AND (“temporal graph” OR “dynamic graph learning” OR “temporal GNN”) AND fraud
S5	(“blockchain fraud detection”) AND (“robustness” OR “explainability” OR “dynamic graph”)

2.3. Inclusion and Exclusion Criteria

To ensure that the selected studies were relevant to the objective and analytical scope of the review, predefined inclusion and exclusion criteria were applied during the study-selection process. Establishing the criteria before full-text assessment helped to reduce selection bias and improve the transparency and reproducibility of the review process [24]. The criteria considered publication period, publication type, research relevance, methodological contribution, language, and availability of sufficient technical information, as shown in Table 3.

Table 3. Inclusion and Exclusion Criteria

Criterion	Inclusion Criteria	Exclusion Criteria
Publication period	Studies published between 2021 and 2026	Studies published outside the 2021–2026 period
Publication type	Peer-reviewed journal articles, conference papers, and reputable scholarly publications	Editorials, book reviews, theses, dissertations, white papers, and non-peer-reviewed publications
Research domain	Studies addressing blockchain or cryptocurrency fraud detection, GNNs, graph learning, adversarial robustness, XAI, or dynamic/temporal graph learning	Studies unrelated to blockchain fraud detection or graph-based learning
Research contribution	Empirical, methodological, or theoretical contributions relevant to the review objective and analytical scope	Studies without a meaningful technical or methodological contribution to the review topic
Technical relevance	Studies addressing fraud detection, graph learning, adversarial attacks/defence, explainability, or dynamic graph learning	Studies that did not address any of the core analytical dimensions of the review
Language	Studies written in English	Studies written in languages other than English
Information availability	Studies providing sufficient methodological, experimental, or technical information for analysis	Studies with insufficient methodological or technical information
Full-text availability	Full text accessible for eligibility and quality assessment	Full text unavailable

After the initial screening, studies that satisfied the inclusion criteria and did not meet any exclusion criterion were retained for full-text assessment. The eligible studies were subsequently subjected to quality assessment before inclusion in the final synthesis.

2.4. Study Selection Process

The study-selection process was conducted using the PRISMA framework to provide a transparent account of study identification, screening, eligibility assessment, and final inclusion. The initial database

search identified 412 records from the selected academic sources. During the identification stage, 98 duplicate records were removed, leaving 314 records for title and abstract screening. Following screening, 237 records were excluded because they were outside the scope of the review, including studies that were not relevant to blockchain or graph-based fraud detection, adversarial robustness, or explainability, as well as non-peer-reviewed publications. Consequently, 77 articles proceeded to full-text eligibility assessment. The 77 full-text articles were assessed against the predefined inclusion and exclusion criteria and examined for methodological adequacy and relevance to the analytical scope of the review. At this stage, four studies were excluded: two because of insufficient methodological detail, one because it did not adequately align with the scope of the review, and one because the full text was inaccessible. This resulted in 73 studies being retained for the final systematic synthesis. The complete selection process is presented in Figure 3, illustrating the progression from the 412 initially identified records to the 73 studies included in the final review.

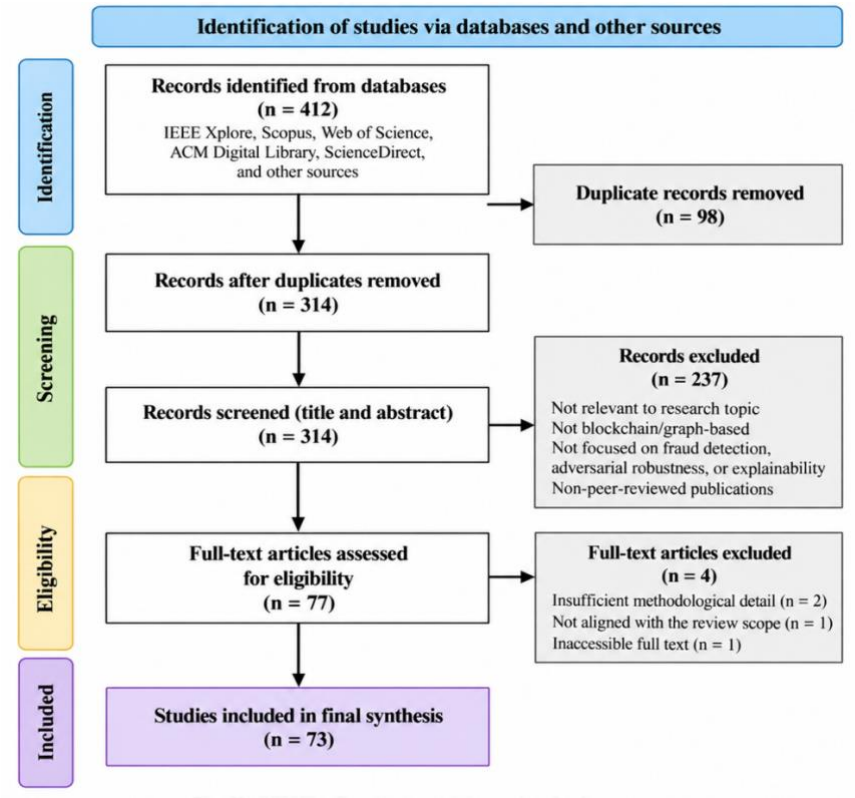

Figure 3. PRISMA flow diagram of the study-selection process

2.5. Quality Assessment

Following the eligibility assessment, the methodological quality and relevance of the selected studies were evaluated using a structured quality-assessment procedure as shown in Table 4. The purpose of the assessment was to ensure that the final synthesis was based on studies that provided sufficiently clear objectives, methodological information, experimental evidence, and evaluation details [20][24].

Table 4. Quality Assessment Criteria

Code	Quality Assessment Criterion	Score
QA1	Are the research objectives clearly stated?	0 or 1
QA2	Is the methodology adequately described?	0 or 1
QA3	Are the datasets and experimental settings clearly specified?	0 or 1
QA4	Are the evaluation metrics appropriately defined?	0 or 1
QA5	Are the reported findings supported by empirical evidence?	0 or 1
QA6	Are the limitations and future research directions discussed?	0 or 1
Maximum score		6

Each criterion was scored using a binary scheme, where 1 indicated that the criterion was satisfied and 0 indicated that it was not satisfied. Consequently, each study could obtain a maximum quality score of 6. The quality scores were used to assess the methodological strength of the eligible studies and to determine their suitability for inclusion in the final synthesis.

Table 5. Distribution of Quality Assessment Scores

Quality Score	Interpretation	Number of Studies	Percentage (%)
6	High quality	28	36.4
5	High quality	27	35.1
4	Acceptable quality	18	23.4
3	Moderate quality	3	3.9
0–2	Low quality	1	1.3
Total assessed		77	100.0

In Table 5, a minimum quality score of 4 out of 6 was adopted as the inclusion threshold. Of the 77 full-text studies subjected to quality assessment, 73 studies achieved scores between 4 and 6 and were retained for the final systematic synthesis, while four studies scoring below the threshold were excluded. As shown in Table 5, 28 studies achieved the maximum score of 6, 27 studies scored 5, and 18 studies scored 4, indicating that the majority of the included studies demonstrated satisfactory methodological quality.

2.6. Data Extraction and Synthesis

A structured data-extraction process was used to collect relevant and comparable information from each study included in the final review. A standardized extraction framework was applied to ensure consistency across the selected publications. The extracted information covered bibliographic characteristics, blockchain platforms and datasets, fraud categories, GNN architectures, adversarial threats and defense mechanisms, explainability techniques, evaluation methods, major findings, limitations, and future research recommendations.

Table 6. Data Extraction Framework

Data Category	Information Extracted
Bibliographic information	Author(s), publication year, and publication venue
Study objective	Main aim or purpose of the study
Blockchain platform	Bitcoin, Ethereum, another blockchain/network, or financial graph
Dataset	Dataset name, source, and characteristics where reported
Fraud category	Fraud, illicit transaction, phishing, money laundering, scam, Ponzi scheme, or other investigated activity
GNN architecture	GCN, GAT, GraphSAGE, temporal GNN, heterogeneous GNN, or other graph-learning method
Dynamic learning	Static, temporal, dynamic, streaming, incremental, or continual graph learning
Adversarial component	Attack type, threat model, defence strategy, or robustness mechanism
Explainability	XAI technique, explanation method, or interpretable architecture
Evaluation metrics	Accuracy, precision, recall, F1-score, ROC-AUC, robustness, explainability, efficiency, or scalability metrics
Major findings	Main results reported by the study
Limitations	Limitations identified by the authors or through review analysis
Future research	Recommendations and unresolved research problems

Following data extraction, the evidence was analyzed using a problem-based thematic synthesis as shown in Table 6. Rather than organizing the literature solely according to algorithms or datasets, the studies

were grouped according to the recurring problems affecting trustworthy GNN-based blockchain fraud detection [25][26]. The principal themes identified included blockchain fraud characteristics, complex and dynamic graph structures, limited fraud-detection performance, adversarial attacks, insufficient adversarial robustness, explainability limitations, regulatory and compliance requirements, benchmark dataset limitations, evaluation inconsistencies, scalability, and deployment constraints [27].

The thematic synthesis was subsequently used to compare the reviewed studies, identify areas of agreement and disagreement, and determine the major unresolved research gaps. Particular attention was given to the extent to which existing approaches jointly address dynamic graph learning, adversarial robustness, and explainability, because these dimensions form the central focus of the present review in Table 7.

Table 7. Characteristics of the Included Studies

Ref	Dataset / Blockchain	Fraud Type / Focus	GNN Method	Dynamic / Temporal	Attack / Defense	XAI	Main Finding
[21]	Ethereum	Phishing	GNN/PDGNN	No	None	None	Graph transaction relationships improve Ethereum phishing detection.
[22]	Ethereum	Phishing	TransWalk + GNN	Limited	None	None	Multi-scale graph features effectively distinguish phishing accounts.
[23]	Ethereum	Phishing	imGraph2Vec + XGBoost	Temporal attributes included	None	Feature-based	Transaction subgraphs provide discriminative phishing representations.
[24]	Ethereum	Phishing	PDTGA/TGAT	Yes	None	Attention-based interpretation	Temporal graph attention captures changing Ethereum transaction behaviour; reported AUC 94.78%.
[25]	Cryptocurrency transaction graph	Phishing	CT-GCN+	No	None	None	Improved transaction-graph convolution supports phishing-node classification.
[26]	Elliptic2 / Bitcoin	Money laundering	Subgraph GNN	Transaction structure	None	Subgraph-level patterns	Introduces Elliptic2 with 122K labelled subgraphs in a 49M-node-cluster Bitcoin graph.
[27]	Elliptic Bitcoin	Illicit transactions	GCN	Dataset has temporal steps; model largely static	None	None	GCN achieved 98.5% accuracy and AUC 0.9444 on Elliptic.
[28]	Bitcoin/Elliptic	AML	Global-local graph attention	Limited	Robust pseudo-label learning	Attention	Combines global/local representations and pseudo-label learning for Bitcoin AML.
[29]	Bitcoin/Ethereum transaction networks	Fraud/anomaly	Enhanced TGN	Yes	Abnormality-aware modelling	Attention	Integrates temporal transaction evolution with graph representation learning.
[30]	Blockchain transaction	Illicit activity	Graph representati	Graph-based	None	None	Shows graph structures are effective for

	graph		on / GNN				identifying illicit blockchain transactions.
[31]	Ethereum	Phishing	PGEA graph embedding	Temporal/repetitive patterns	None	None	Pheromone-inspired sampling captures repetitive and temporal transaction patterns.
[32]	Blockchain transactions	Phishing	Dynamic feature-fusion GNN	Yes/graph-aware	None	None	Combines global topology with local semantics for phishing detection.
[33]	Ethereum	Phishing	GAT + multi-graph message passing	Transaction direction	None	Attention weights	Direction-aware graphs achieved 97.21% accuracy and AUC 0.9721.
[34]	Ethereum	Phishing	Temporal graph-sequence model	Yes	None	None	Dynamically ordered transaction-subgraph sequences improve phishing detection.
[35]	Two Ethereum datasets	Phishing	Graph contrastive learning	No explicit TGN	Self-supervised robustness to label scarcity	None	Self-supervised node–subgraph contrastive learning achieved 97% precision and 95% F1.
[36]	Blockchain transaction graph	Phishing	DeepPhishDetect/GAT	Transaction dependencies	None	Attention/ relationship interpretation	Models' node representations and label dependencies to uncover suspicious fraudsters.
[37]	Ethereum	Fraud	Cluster-GAT	Partition-based	Semi-supervised	None	Addresses labelled-data scarcity and Ethereum graph scale.
[38]	Ethereum	Fraud	ML/XAI benchmark	Near-real-time	None	SHAP	Demonstrates interpretable Ethereum fraud screening; useful XAI comparison evidence.
[39]	Ethereum	Phishing	Temporal-relational attention	Yes	None	Attention	Jointly models temporal and relational transaction patterns.
[40]	Ethereum	Phishing	Multi-stream graph/feature fusion	Transaction-aware	None	Feature contribution	Fuses complementary transaction information for phishing identification.
[41]	Ethereum	Phishing	Heterogeneous Graph Transformer	Heterogeneous transaction context	None	PDHGExplainer	Integrates heterogeneous transaction modeling with an explicit graph explainer.
[42]	Blockchain transaction	Fraud	CoSemiGNN	Yes	Semi-supervised	None	Uses dynamic graph learning and semi-

	graph				resilience		supervised association for evolving fraud.
[43]	Three blockchain phishing datasets	Phishing	GraphFlow Gen + GAT/Transformer	Yes	RL-based graph refinement	Attention	Generates realistic dynamic subgraphs to alleviate severe class imbalance.
[44]	Ethereum + four real-world temporal graphs	Phishing	EFD-AW/Transformer	Yes	None	Attention	Models' anonymity and long-term temporal evolution in continuous transaction graphs.
[45]	Ethereum	Phishing	Contrastive graph encoder	Graph transaction structure	Label-efficient learning	Explanation analysis	Targets severe label scarcity and analyst-review constraints.
[46]	Ethereum	Phishing	Personalized heterophilic GNN	Transaction semantics	None	Semantic interpretation	Motif-based sampling and semantic GNN learning outperform baselines.
[47]	Blockchain networks	Anomalous/ illicit nodes	GNN	No explicit temporal component	None	None	Applies GNN representation learning to blockchain anomalous-node identification.
[48]	Ethereum	Account fraud	Graph-based fraud detector	Transaction history	None	None	Recent account-level Ethereum fraud detection framework.
[49]	Ethereum literature	Phishing review	Multiple	Reviews temporal methods	Reviews threats	Reviews explainability	Identifies persistent data, scalability, temporal, and evaluation gaps.
[50]	Ethereum smart contracts	Vulnerability/fraud-related security	Multiple ML/GNN	Some temporals	Reviews attacks	Reviews interpretable methods	Supports broader blockchain security and benchmark discussion.
[51]	Financial fraud studies	Multiple fraud types	Multiple GNNs	Static/dynamic/temporal	Reviews security gaps	Reviews XAI	Reviewed 33 studies and highlighted limited unsupervised, temporal, and edge/graph-level research.
[52]	Financial systems	Multiple	Multiple	Reviews dynamic graphs	Reviews robustness	Reviews explainability	Provides a unified framework for GNN financial-fraud research.
[53]	YelpChi, Amazon	Review/fraud	PC-GNN	No	Imbalance-aware sampling	None	Label-balanced sampling improves detection under severe fraud-class imbalance.
[54]	eBay transaction networks	Transaction fraud	Heterogeneous GNN	Production/streaming setting	None	Integrated explainer	Combines scalable heterogeneous GNN detection with human-readable explanations.
[55]	Public fraud datasets	Fraud	H²-FDetector	No	Camouflage-	None	Explicitly models both homophilic

					resistant aggregation		and heterophilic fraud relationships.
[56]	Large e-commerce transaction graph	Payment fraud	Lambda Neural Network	Real-time/dynamic	Leakage-safe directed topology	None	Improves precision while reducing P99 inference latency by >75%.
[57]	YelpChi, Amazon, industrial dataset	Fraud	GAGA Transformer	No	Low-homophily/camouflage resilience	None	Group aggregation improves performance in low-homophily fraud graphs.
[58]	Two financial datasets	Transaction fraud	HHLN-GNN/GTS	No	Imbalance handling	Attention	Subgraph sampling and synthetic minority-node generation improve minority fraud detection.
[59]	Telecom transaction/ call networks	Collaborative fraud	CORE-DGNN	Yes	Collaborative-fraud resilience	Self-attention	Dynamic collaboration modeling improves gang-fraud identification.
[60]	Financial fraud graphs	Fraud	SCN_GNN	No	Camouflage/graph sparsity	None	Joint node and topology learning helps sparse multi-relation fraud detection.
[61]	Financial transaction graph	Financial fraud	Dual-channel GAT	No	None	Attention	Node and semantic attention jointly capture complex fraud relations.
[62]	Real online-payment platform	Payment fraud	TGN	Yes	None	None	Event-based temporal graphs improve detection of evolving fraudulent events.
[63]	Five supply-chain finance datasets	Financial fraud	Heterogeneous GNN	Multi-view graph	None	Integrated explanations	Demonstrates joint heterogeneous fraud detection and explanation.
[12]	Transaction fraud graphs	Transaction fraud	ASA-GNN	No	Adaptive sampling	None	Adaptive sampling reduces noisy/irrelevant neighbourhood information.
[64]	Ethereum/heterogeneous transaction networks	Fraud	Heterogeneous GNN	No	Camouflage-resistant neighbour selection	None	Similarity-aware filtering suppresses misleading neighbouring accounts.
[65]	Telecom networks	Collaborative fraud	Multi-network GNN	Dynamic behaviour	Heterophily-aware	Attention	Multi-network modelling captures coordinated fraud behaviour missed by single graphs.
[66]	Fraud benchmark graphs	Fraud	GE-GNN	No	Camouflage-resistant edge modelling	None	Integrating rich edge information improves detection of camouflaged fraudsters.

[7]	Heterogeneous financial networks	Financial fraud	Metapath-guided GNN	No	None	Metapath attention	Semantic relation paths improve heterogeneous financial fraud representation.
[67]	SIFT synthetic transaction graph	Financial fraud	LayerWeighted-GCN	Pattern-aware	Robust layer weighting	Layer weights	Adaptive layer weighting improves detection across diverse fraud patterns.
[68]	Real Latin-American payment platform	Payment fraud	TGN	Yes	None	None	Combining event graphs improves fraudulent-event detection.
[69]	Financial transaction graph	Fraud	Multi-layer GNN	Adaptive/evolving setting	None	None	Node and edge information improves transaction fraud representation.
[70]	Industrial financial datasets	Transaction/credit fraud	C2GAT	Yes	None	Temporal attention	Streaming dynamic graphs improve accuracy while satisfying real-time latency requirements.
[71]	Bank transaction networks	Cash-out fraud	Dense-subgraph modelling	Transaction networks	None	Structural interpretation	Demonstrates importance of coordinated subgraph patterns in bank fraud.
[72]	Large fraud graphs	Fraud	R-GCN/GNN workflow	Repeated model updating	None	None	Shows scalability and accelerated graph processing for fraud detection.
[73]	Benchmark graphs	Adversarial robustness	GNN + spectral topology refinement	No	Structural attack defence	None	Improves adversarial robustness while scaling to million-node graphs.
[74]	Three benchmark graphs	Robustness	ERGCN	No	Structure/feature defence	None	Graph structure enhancement and self-training improve robustness.
[75]	Benchmark graphs	Robustness	C²oG	No	Multi-view adversarial defence	None	Co-training feature and structural views improves attack resistance without sacrificing clean accuracy.
[76]	Benchmark graphs	Adversarial attack	GCN/GAT/GraphSAGE/GIN targets	No	Node/feature/topology attack	None	Shows that perturbing a single node can mislead several GNN architectures.
[77]	Benchmark graphs	Robustness	Graph transformation + GCN	No	Lightweight structural defence	None	Obtains comparable defence with substantially lower computational cost.
[78]	Benchmark graphs	Robustness	Similarity-based robust GNN	No	Structural defence	None	Similarity filtering mitigates adversarial graph

							perturbations.
[79]	Standard GNN benchmarks	Robust architecture	Robust-NAS GNN	No	Adaptive adversarial defence	None	Searches GNN architectures specifically optimized for adversarial robustness.
[80]	Benchmark graphs	Robustness	Dual robust GNN	No	Structural adversarial defence	None	Preserves original node similarity while suppressing malicious topology changes.
[81]	Benchmark graphs	Attack/defence analysis	Multiple GNNs	No	Adversarial attack/defence	None	Provides theoretical and empirical insight into uneven vulnerability across graph nodes.
[16]	Review	GNN security	Multiple	Static/dynamic discussion	Attack and defence taxonomy	Limited	Synthesizes adversarial, privacy and defensive mechanisms relevant to trustworthy GNN deployment.
[18]	Review	Trustworthy GNNs	Multiple	Multiple	Robustness taxonomy	XAI taxonomy	Connects robustness, privacy, fairness, and explainability as joint trustworthiness dimensions.
[82]	Benchmark graphs	Robustness	Tensor-enhanced GNN	No	Multi-attack defence	None	Low-rank tensor modeling strengthens structural integrity and robustness.
[83]	Benchmark graphs	Robustness analysis	Multiple GNNs	No	Attack/defence analysis	None	Examines why common GNN architectures remain vulnerable across attacks.
[84]	Synthetic + real graph datasets	Explainability	SubgraphX	No	N/A	Subgraph/ Shapley	Uses Monte-Carlo tree search and Shapley values to identify important subgraphs.
[85]	Cora/PubMed + synthetic	Explainability	Model-agnostic GNN explainer	No	N/A	GraphSVX/Shapley	Provides node/feature attribution using graph-specific Shapley values.
[86]	Three GNN explanation datasets	Explainability	Counterfactual explainer	No	Graph perturbation for explanation	CF-GNNExplainer	Generates minimal edge-removal counterfactuals with ≥94% reported accuracy.
[87]	Multiple datasets + eBay fraud case	Explainability evaluation	Multiple GNNs	No	N/A	Multiple explainers	Establishes systematic GNN-XAI evaluation and includes an eBay fraud case study.
[88]	Graph	Explainabili	Model-	No	N/A	GStarX	Incorporates graph

	classification datasets	ty	agnostic				structure directly into cooperative-game explanations.
[89]	Multiple node-classification graphs	Explainability	Model-agnostic GNN explainer	No	N/A	GNNShap	GPU-parallel Shapley estimation improves both explanation fidelity and scalability.

3. PROBLEM-BASED LITERATURE REVIEW

3.1. Blockchain Fraud and Fraud Detection Challenges

Blockchain technology has its origins in creating a secure, transparent, and decentralized system for digital transactions [90]. These characteristics have led to the deployment of cryptocurrencies, decentralized finance (DeFi), and blockchain services, but have also opened new avenues for blockchain-related fraud. The anonymity of blockchain transactions, along with the decentralized and borderless nature of blockchain networks, makes it hard for the typical financial monitoring systems to effectively detect suspicious activity [9]. This has made blockchain networks a popular target for a variety of financial crimes, such as cryptocurrency theft, investment fraud, money laundering, ransomware payments, illicit financial transactions, and other high-tech cybercrime scams. According to recent reports, blockchain crimes are becoming more sophisticated and extensive, with financial institutions, regulatory bodies, law enforcement agencies, and blockchain analytics firms facing serious difficulties in their attempts to combat them, resulting in billions of dollars in financial losses each year [6][12]. With the rapid development of fraudulent techniques, the complexity of pattern detection in suspicious transactions has increased, which inspires researchers to find more intelligent and adaptive fraud detection methods from the perspective of graph learning and artificial intelligence. Cryptocurrency theft and investment scams are among the most common types of blockchain fraud. They take advantage of weaknesses in the security of cryptocurrency exchanges, smart contracts, decentralized applications, or user security practices to steal digital assets or trick investors into fraudulent investment schemes [11]. These can range from fake Initial Coin Offerings (ICOs), giveaway scams, and fraudulent trading platforms to impersonation attacks and deceptive investment opportunities. As reported by [10], cryptocurrency scams in 2023 resulted in billions of dollars of stolen money, and investment fraud was one of the main causes of financial losses experienced in cryptocurrencies. Social engineering and phishing are becoming more sophisticated thanks to the increasing use of artificial intelligence. Moreover, the highly interconnected and complex transaction structures within blockchain networks make it increasingly difficult for traditional fraud detection methods to identify coordinated fraudulent activities [70][91]. Another significant obstacle in blockchain ecosystems is money laundering. Cryptocurrencies have seen a rise in their use by criminal organizations to obfuscate illegally obtained money using transaction layering, cryptocurrency mixing services, tumblers, cross-chain transfers, decentralized exchanges, and privacy-preserving cryptocurrencies. The decentralized blockchain structure enables users to send money quickly between jurisdictions without needing to traverse numerous conventional financial tracking systems, making illicit financial transactions easier [92]. Elliptic says billions of dollars' worth of cryptocurrency transactions are laundered every year. These transactions are especially hard to detect as money laundering is done by mimicking "legitimate" financial transactions. Moreover, the advent of Decentralized Finance (DeFi) protocols has created more wash trades, making current blockchain analytics and fraud detection systems more complex.

There has also been a significant rise in Ponzi and pyramid schemes in blockchain ecosystems, notably in the decentralized finance (DeFi) sector and investment platforms using tokens as an entry point. The schemes usually offer unrealistic returns with minimal risk of investment; investment by new investors is used to pay earlier investors [93]. More recently, fraudulent investment schemes have morphed into what has been dubbed a 'rug pull attack', in which the developers of blockchain projects simply pull the rug from under investors after raising a lot of money. The findings have revealed that many cryptocurrency fraud losses involve fraudulent DeFi tokens and rug pull schemes [94][95]. These fraudulent projects are difficult to detect, as a large number of them appear to be similar to real investment opportunities in the early stages of development, and therefore cannot be distinguished using standard fraud detection methods [21][96]. Another significant cybersecurity issue is ransomware cryptocurrency transactions. Ransomware attacks involve the attackers encrypting victims' data and asking for a ransom, usually in cryptocurrencies, to unlock it. Cryptocurrencies are highly appealing for ransomware payments because of their decentralization, anonymity, and speed of transactions [29]. Blockchain payment channels are being increasingly used by ransomware groups to receive and distribute ransom funds, according to Europol (2024) and the United Nations Office on Drugs and Crime (UNODC, 2024). The use of intermediary wallets, cryptocurrency

mixing services, and privacy-focused cryptocurrencies also makes it difficult to track ransomware payments. However, a recent study indicated that ransomware transactions tend to have unique graph structures and transaction patterns, which can enhance ransomware transaction detection models based on graphs, but accurately determining the transactions remains a challenging research problem [8][58].

In addition to fraud, blockchain networks are also used to enable widespread illicit financial flows (IFFs) such as tax evasion, dark web trading, underground market activities, sanctions evasion, and financing of terrorism. Both the Financial Action Task Force (FATF) and Europol (2024) have recently raised concerns about the use of decentralized financial (DeFi) infrastructures for transnational criminal activities. Regulatory enforcement and International financial investigations are very complex due to the cross-border and decentralized nature of blockchain transactions. Advanced graph analytics and machine learning algorithms are needed to model the multi-hop transaction networks that typically include many intermediaries, and have been shown to be applicable to the task of illicit financial flows detection [30]. The amount and sophistication of financial crimes using blockchain technology are growing, and there is a pressing need for more sophisticated, explainable, and adaptive fraud detection methods that can detect more complex fraudulent patterns in a changing blockchain environment. Overall, the literature highlights that the continuous evolution and diversification of blockchain-related fraud have identified major shortcomings in traditional fraud detection methods. The complexity, scale, and dynamic nature of blockchain transaction networks require smart graph-based learning models with the ability to capture complex transaction relationships and possess good robustness, scalability, and interpretability. The main driving force behind adversarially robust and explainable Graph Neural Networks for the purpose of accurate and trustworthy fraud detection in a dynamic blockchain environment is these persistent challenges shown in Table 8.

Table 8. Major Fraud Types in Blockchain Networks and Associated Challenges

Fraud Type	Primary Objective	Detection Challenge
Cryptocurrency Theft and Scams	Asset theft and investor deception	Hidden transaction relationships
Money Laundering	Concealing illicit fund origins	Transaction obfuscation and mixing
Ponzi and Pyramid Schemes	Fraudulent investment schemes	Similarity to legitimate investments
Ransomware Transactions	Extortion payments	Use of intermediary wallets
Illicit Financial Flows	Criminal financial operations	Cross-border and multi-layer transactions

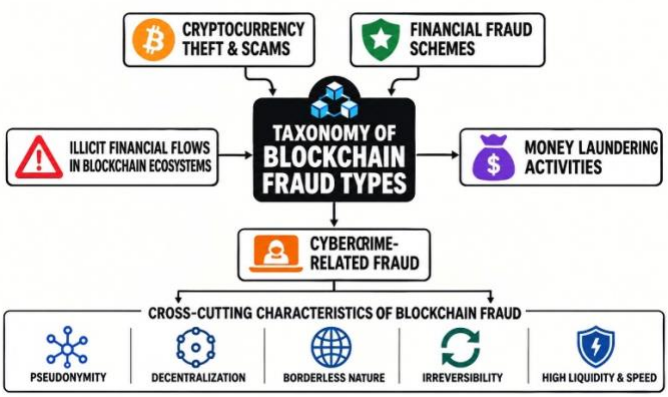

Figure 4. Taxonomy of Blockchain Fraud Types

Figure 4 classifies significant fraud mechanisms identified in the blockchain ecosystems based on the nature of the tech fraud, such as cryptocurrency theft and scams, financial fraud schemes, fraud involving cybercrime, money laundering, ransomware-related transactions, and frauds under other categories. Based on recent blockchain crime reports and literature on the subject of fraud detection (Chainalysis, 2024; Europol, 2024; FATF, 2023; UNODC, 2024).

3.2. Problem of Complex Graph Structures in Blockchain Networks

The blockchain transaction network is naturally modeled as a graph, as blockchain wallets, smart contracts, cryptocurrency exchanges, miners, decentralized applications, and other entities interact with each

other continually through millions of transactions and blocks [30]. In the blockchain domain, the complexity of the graphs poses serious challenges for blockchain fraud detection systems [58][97][98] though Graph Neural Networks (GNNs) have shown great promise in learning from such relational data. Conventional tabular datasets are contrasted by the highly connected, high-dimensional, sparse, and heterogeneous nature of blockchain networks, which constantly change over time. This special property makes representation learning on graphs far more challenging and diminishes the impact of a lot of previous fraud detection methods. The complexity of the graph is one of the most significant issues that prevent accurate blockchain fraud detection [97][98]. The main issue is the extensive size of blockchain transaction graphs. The transaction graph on public blockchain platforms, like Bitcoin and Ethereum, expands in size and complexity with the millions of transactions and hundreds of millions of wallet addresses [27]. When the number of nodes and edges grows large, graph learning models start to experience computational bottlenecks in neighborhood aggregation, message passing, and graph convolution operations. According to [42] there are several popular GNN-based algorithms that suffer from low efficiency in large-scale blockchain networks due to memory constraints and high computational demands. In addition, fraudulent activities have long-range dependencies and complex interaction patterns across various transactions, and cannot be fully captured by shallow graph neural network architectures. Given the increasing size of blockchain ecosystems, efficient processing of large-scale transaction graphs is still a key research challenge, as shown in Figure 5.

High-dimensional node and edge features of blockchain transaction networks [32] is another significant challenge. In contrast to nodes, edge attributes are often transaction values, timestamps, gas fees, transaction directions, interaction frequencies, and transaction behavior [99] Node attributes can include wallet balances, transaction frequency, account age, smart contract interactions, token holdings, and transaction behavior. These are very rich features and useful for detecting fraudulent behavior but can also be very high-dimensional learning problems that make model development difficult. High-dimensional feature spaces lead to high computation costs, higher risk of overfitting, and lower interpretability. Several recent studies have shown that many blockchain fraud detection models are limited in their ability to select the most informative features from thousands of transaction features, as noted by [19][93]. Besides, because of the features that have redundant information, large variations in node attributes, and high correlations between features, meaningful and discriminative representations are hard to learn for nodes in the GNNs. Therefore, feature selection and dimensionality reduction still have a long way to go as open problems in the field of blockchain graph learning. The lack of data density and noise in blockchain transactions also makes it difficult to detect fraud. Many blockchain networks have a significant number of wallet addresses that have sparse neighborhood information and weak graph connectivity, as most of them are engaged in few transactions [100]. This isolation is exacerbated by the fact that bad actors often use temporary wallets and intentionally spread fraudulent actions over many wallets to evade detection. There is also a lot of noise in blockchain datasets, related to exchange wallets, smart contract interactions, inconsistent labels, missing data, and silent addresses. Proved that sparse graph structures make graph representation learning less effective since less contextual information is available for the classification of nodes [101]. Likewise, [8] have found that noisy transaction data causes classification uncertainty and results in incorrect fraud prediction. Even in the presence of noise and sparseness, however, enhancing the quality of the graph representation is still one of the core challenges for graph-based fraud detection systems.

Blockchain ecosystems are made up of highly heterogeneous entities, each with different behavioral properties, apart from their graph scale and data sparsity [9]. Individual users, cryptocurrency exchanges, decentralized finance protocols, smart contracts, liquidity pools, mining pools, token issuers, and regulatory monitoring services [21]. There are different types of transaction patterns, behavioral patterns, and risk profiles among the entities. Most of the existing graph learning methods, however, assume that blockchain graphs are structurally homogeneous, which doesn't account for the complex semantic relationships between various blockchain actors. The model Heterogeneous Graph Neural Networks (HGNNs) was introduced by [6], which models different node and edge types in a single graph framework, thus achieving better fraud detection performance. However, there are still some issues to be addressed, including scalability, consistency in representation, computational efficiency, and explainability [102][103]. Thus, modeling of the heterogeneous blockchain ecosystems in an accurate manner remains an active field of research and is still considered a key requirement for reliable fraud detection in the real-world blockchain ecosystem. In general, the literature shows that the structure of blockchain transaction networks presents significant challenges to graph learning models: very large, feature high-dimensional data, are sparse and noisy, and contain various heterogeneous entities. Each of these challenges hinders the scalability, representation capability, and detection accuracy of existing fraud detection systems based on GNNs. All of these challenges impede scalability, representational capability, and detection accuracy of the existing fraud detection systems based on GNNs. To overcome these challenges, new graph learning methods need to be created that can be scaled, are robust, and offer explainability to successfully model the complexity of the dynamic blockchain

ecosystem. A comprehensive summary of the major challenges, their causes, impacts on detection performance, and corresponding research gaps is presented in Table 9.

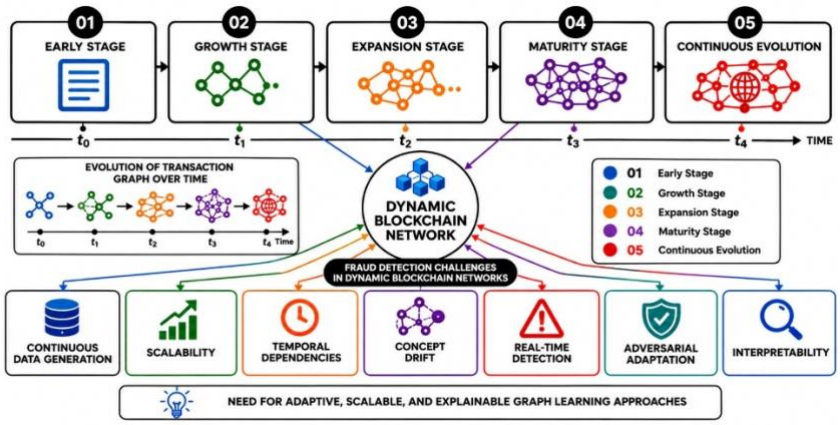

Figure 5. Evolution of Dynamic Blockchain Networks and Fraud Detection Challenges

Table 9. Summary of Major Challenges Affecting Blockchain Fraud Detection Accuracy

Problem	Cause	Impact on Detection Performance	Research Gap
Large-Scale Transaction Graphs	Massive number of nodes, edges, and continuous transaction growth	Increased computational cost, memory consumption, and training complexity	Need for scalable GNN architectures capable of handling billion-scale transaction networks
High-Dimensional Features	Numerous node and edge attributes (transaction amount, timestamps, balances, smart contracts, etc.)	Feature redundancy, overfitting, reduced interpretability, and increased complexity	Development of efficient feature selection and representation learning techniques
Sparse and Noisy Data	Limited interactions among addresses, incomplete labels, inactive wallets, and noisy transaction records	Reduced learning effectiveness and inaccurate node embeddings	Robust learning methods capable of handling sparse and noisy blockchain data
Heterogeneous Blockchain Entities	Diverse node types such as wallets, exchanges, miners, smart contracts, and DeFi platforms	Difficulty in modelling complex semantic relationships	Advanced heterogeneous GNN frameworks with improved explainability
Dynamic Network Evolution	Continuous generation of transactions, addresses, and smart contracts	Outdated node representations and declining model performance over time	Dynamic and temporal graph learning models capable of real-time adaptation
Concept Drift	Evolution of fraud strategies and attacker behaviours	Reduced detection accuracy and poor generalization to emerging fraud patterns	Continual learning and adaptive fraud detection mechanisms
Class Imbalance	Fraudulent transactions constitute a very small proportion of blockchain activities	Bias toward legitimate transactions and low fraud recall rates	Effective imbalance-handling techniques and fraud-aware learning strategies
False Positives	Similarity between legitimate and suspicious transaction behaviours	Incorrect flagging of legitimate users and reduced trust in detection	More precise and explainable classification mechanisms

		systems	
False Negatives	Sophisticated fraud concealment techniques and hidden transaction patterns	Fraudulent activities remain undetected, causing financial losses	Improved detection of subtle and evolving fraud behaviors
Cross-Platform Generalization	Differences in blockchain architectures, consensus mechanisms, and transaction patterns	Poor performance when models are transferred across blockchain platforms	Development of transferable and platform-independent fraud detection models
Scalability Constraints	Computationally intensive message passing and graph aggregation operations	Limited deployment in real-world blockchain environments	Lightweight and distributed graph learning frameworks
Adversarial Attacks	Node injection, edge manipulation, feature perturbation, poisoning, and evasion attacks	Significant degradation of fraud detection accuracy and robustness	Unified adversarial defense frameworks for blockchain GNNs
Low Adversarial Robustness	Sensitivity of GNN models to small graph perturbations	Unreliable predictions under adversarial conditions	Robust GNN architectures with certified security guarantees
Lack of Explainability	Black-box nature of deep graph learning models	Reduced transparency, trust, and regulatory acceptance	Explainable GNN frameworks tailored to blockchain fraud analysis
Dataset Limitations	Scarcity of labelled fraud data and lack of dynamic benchmark datasets	Restricted model training and unrealistic evaluation outcomes	Public benchmark datasets containing temporal and adversarial scenarios
Evaluation Challenges	Absence of standard robustness and explainability metrics	Difficult comparison of competing approaches and reproducibility issues	Standardized evaluation frameworks and benchmarking protocols
Regulatory Compliance Requirements	Increasing demands for transparency, accountability, and AML compliance	Constraints on deployment of opaque AI systems	Integration of explainability, auditability, and compliance mechanisms into GNN-based fraud detection systems

3.3. Problem of Dynamic and Evolving Blockchain Networks

Blockchain transaction networks are dynamic systems that continuously evolve with each passing transaction, wallet address, smart contract, and decentralized application (DApps) created [9]. In contrast to the usual graph datasets employed in machine learning, blockchain networks are dynamic, with continually changing architectures and fraud patterns, which makes fraud detection a much more difficult task [22]. Blockchain graphs grow over time, and what is learned at one moment might quickly become outdated, limiting the ability of fraud detection models [97]. Thus, temporal dependencies and reaction to the changing tactics of fraud have emerged as one of the primary challenges of graph-based blockchain fraud detection [8]. One of the primary reasons causing this problem is that constant blockchain transactions are being generated. [104] Public blockchain networks like Bitcoin and Ethereum handle millions of transactions daily, leading to the growth of transaction graphs in the network. This ongoing process necessitates models for fraud detection to adapt to node representations as they change over time in accordance with the emergence of new interactions. Most of the traditional GNN architectures, however, are static and fail to provide flexible training when there are significant changes in the graph structure, which makes them unsuitable for large-scale real-time blockchain applications. Efficiently updating graph representations is essential for maintaining accurate embeddings and ensuring the detection system performs well [29].

One more important challenge is modeling temporal dependencies. Ransomware payments, money laundering, and other fraudulent activities are often not single transactions but series of transactions over a long duration of time. Thus, learning both structural relationships and temporal interactions between blockchain entities is necessary to accurately detect these activities [29]. However, many conventional GNN models are mainly based on graph topology and fail to take temporal information into account, which

restricts their ability to detect fraudulent behaviors over time. Recently, Temporal Graph Neural Networks (TGNNs) have appeared to overcome this problem, but long-term temporal dependency modeling is still an open research problem [105]. Another challenge with blockchain fraud is concept drift, which happens when fraudsters change their tactics and techniques. Over time, fraudsters constantly evolve their tactics and techniques, rendering old frauds useless and creating new fraud methods [70]. This inevitably harms the performance of machine learning models trained from past blockchain data, especially in varying settings. [106] showed how concept drift significantly affects the trustworthiness of blockchain fraud detection models and the importance of using adaptive learning techniques to keep up-to-date with the evolving fraud patterns in various blockchain platforms.

Moreover, in today's blockchain environment, timely fraud detection is essential to reduce financial losses and enable applications like cryptocurrency exchanges, decentralized finance (DeFi) platforms, and Anti-Money Laundering (AML) solutions [107]. Real-time detection is still challenging, as the learning algorithms frequently involve neighborhood aggregation, embedding updates, and message passing, which are all computationally costly; achieving a balance between detection accuracy and computational efficiency is one of the major challenges in the development of practically deployed blockchain fraud detection systems [108]. Thus, dynamic graph learning, streaming graph analytics, and online graph representation learning have become crucial research directions to develop scalable and adaptive blockchain fraud detection frameworks that can function in the real world [79]. In summary, the literature shows that current GNN-based fraud detection systems face significant challenges in light of the ongoing development of blockchain networks. Traditional static graph learning models have their limitations due to several factors, including continuous transaction generation, temporal dependencies, concept drift, and real-time processing needs. These major challenges, their impacts on fraud detection, and the corresponding research gaps are summarized in Table 10. The solution to these problems will demand dynamic, adaptive, and computationally efficient graph learning procedures that can keep up with the rapidly changing blockchain-based systems and stay accurate in fraud detection.

Table 10. Summary of Dynamic and Evolving Blockchain Network Challenges

Challenge	Description	Impact on Fraud Detection	Research Gap
Continuous transaction generation	Blockchain graphs continuously expand with new transactions, wallet addresses, and smart contracts.	Static GNN models become outdated and require expensive retraining.	Efficient online and incremental graph learning methods.
Temporal dependency modelling	Fraud often occurs through transaction sequences over time rather than isolated events.	Conventional GNNs fail to capture temporal relationships and evolving behaviors.	Advanced temporal and dynamic graph neural networks for fraud detection.
Concept drift	Fraud strategies continuously evolve as attackers adapt to existing detection methods.	Detection accuracy decreases when models rely on historical fraud patterns.	Adaptive learning models capable of detecting and responding to concept drift.
Real-time fraud detection	Modern blockchain platforms require immediate identification of suspicious transactions.	Computational complexity limits real-time deployment of graph learning models.	Scalable streaming graph analytics and low-latency GNN architectures.

3.4. Problem of Limited Fraud Detection Accuracy

Although Graph Neural Networks (GNNs) have proven to be valuable tools for detecting fraud in the blockchain domain, there are still a number of limitations in their current models that hinder their accuracy, generalization, and applicability [32]. Class imbalance, classification errors, poor cross-platform adaptability, and scalability limitations are recurring problems reported by previous studies, which are still considered significant challenges in reliable fraud detection [12]. Together, these difficulties limit the utility of graph learning models in the context of dynamic blockchain settings, as criminal patterns become more intricate and more difficult to detect. The biggest issue limiting detection accuracy is the extremely skewed nature of the blockchain fraud datasets. In fact, the number of fraudulent transactions is a minor proportion of the total number of transactions on the blockchain, and thus, the class distributions are highly skewed [9]. This results in an overall high accuracy of the machine learning models and the failure to correctly classify

numerous fraudulent transactions because they belong to a smaller class that is not the majority class. As a consequence, machine learning models have high overall accuracy but poor performance in detecting many fraudulent transactions because they are in a small class, which is not the most prevalent class. [21] showed that, in fact, recall and sensitivity to fraudulent activities are strongly affected if there are differences in the number of positive and negative samples; many fraudulent transactions might go unseen even when the classification performance seems to be high.

Additionally, fraud detection systems on the blockchain rely on classification errors, which reduce their trustworthiness. False positives are false alarms that occur when a legitimate user is flagged as a fraud user, interfering with financial transactions, undermining the trust of users, and increasing investigation expenses. On the other hand, false negatives enable fraudulent activities to go undetected, money to be lost, and security breaches to occur. The challenge of finding a good balance between sensitivity and specificity lies in the fact that fraudulent transaction patterns are very complex and frequently appear similar to legitimate financial transactions [15][42]. Therefore, improving false positive and false negative rates remains a key goal of blockchain fraud detection research. Poor generalization of the existing fraud detection models across different blockchain platforms is another big constraint. Most of the GNN-based models are trained and tested with data from a single blockchain ecosystem, e.g., Bitcoin or Ethereum. The transaction structures, user behaviors, consensus mechanisms, and fraud characteristics vary significantly among blockchain platforms, however [19]. Therefore, the models created for one blockchain are not likely to have similar performance on another blockchain. However, the issue of cross-chain generalization remains largely unsolved, and current blockchain fraud detection systems have not been widely used in practice, as reported by [102]. Scalability is another significant issue with the growing networks of blockchain transactions. In spite of their impressive detection accuracy in the experimental setting, many GNN models suffer from scalability issues when applied to real-world blockchain networks with millions of nodes and transactions [91]. As the size of the graph grows significantly, the memory usage, computation, training time, and inference latency also increase, making real-time deployment of the graph more difficult. Thus, it highlights the importance of scalable architectures for graph learning that can ensure high detection accuracy while efficiently handling next-generation blockchain transaction networks [8]. In conclusion, the literature shows that class imbalance, classification errors, limited cross-platform generalization, and scalability issues all contribute to the lack of reliability of current fraud detection systems based on GNNs, as summarized in Table 11. To overcome these limitations, strong graph learning models that are capable of handling highly imbalanced datasets, reducing classification errors, generalizing over heterogeneous blockchain platforms, and having computational efficiency in large blockchain environments are required.

Table 11. Summary of Challenges Affecting Blockchain Fraud Detection Accuracy

Challenge	Description	Impact on Detection Performance	Research Gap
Imbalanced fraud datasets	Fraudulent transactions represent only a small proportion of blockchain data.	Low recall and poor detection of minority fraud classes.	Robust learning techniques for highly imbalanced graph datasets.
False positive and false negative errors	Legitimate users may be misclassified while fraudulent transactions remain undetected.	Reduced trust, financial losses, and decreased detection reliability.	Cost-sensitive and balanced classification approaches.
Cross-platform generalisation	Models trained on one blockchain often perform poorly on others.	Limited transferability and practical deployment.	Cross-chain graph learning and domain adaptation methods.
Scalability constraints	Large transaction graphs increase computational and memory requirements.	High latency and limited real-time performance.	Scalable graph learning architectures for large blockchain networks.

3.5. Problem of Adversarial Attacks Against GNN-Based Fraud Detection

Graph Neural Networks have substantially improved blockchain fraud detection by effectively modeling complex transaction relationships. Nevertheless, recent studies have shown that these models are highly vulnerable to adversarial attacks designed to manipulate graph structures or node features in ways that mislead the learning process [6]. Unlike conventional cyberattacks that exploit software vulnerabilities, adversarial attacks deliberately introduce carefully crafted perturbations into graph data while remaining

difficult to detect. In blockchain environments, these attacks can enable fraudulent transactions, money laundering operations, and malicious wallet addresses to evade fraud detection systems, making adversarial manipulation one of the most critical security challenges confronting modern graph learning models [22]. One common attack strategy is node injection, where attackers introduce malicious wallet addresses into blockchain transaction graphs to manipulate neighborhood aggregation and graph representations. These injected nodes interact with legitimate accounts, influencing message passing and producing misleading node embeddings that reduce fraud detection accuracy [94][107] demonstrated that node injection attacks can significantly degrade node classification performance because malicious wallets are often designed to resemble legitimate users. The permissionless nature of public blockchain networks further facilitates the creation of fraudulent wallet addresses, making node injection an ongoing challenge for graph-based fraud detection systems.

Another important threat involves edge manipulation attacks, in which attackers modify graph connectivity by adding, removing, or altering transaction links. Since GNNs rely heavily on graph topology during message passing, even minor structural perturbations can substantially influence node representations and classification outcomes. In blockchain fraud detection, adversaries may create fake transaction links between legitimate and fraudulent wallets or remove suspicious connections to conceal illicit financial relationships. Demonstrated that carefully designed edge perturbations can significantly reduce fraud detection accuracy, particularly because malicious structural modifications are often difficult to distinguish from the natural evolution of blockchain transaction graphs [13][98]. Feature perturbation attacks present another significant challenge by modifying node or edge attributes without changing graph topology. Attackers may manipulate transaction values, frequencies, behavioral patterns, or account characteristics so that malicious activities resemble legitimate behavior [94]. Although these modifications are often subtle, recent studies have shown that small feature perturbations can substantially degrade GNN performance, highlighting the vulnerability of feature-based graph learning mechanisms [109]. As blockchain transaction features become increasingly complex, detecting adversarial feature manipulation becomes progressively more difficult.

Adversarial evasion attacks are performed during model deployment rather than training. Instead of modifying training data, attackers adapt transaction behaviors, distribute activities across multiple wallet addresses, and manipulate transaction sequences to avoid triggering fraud detection mechanisms [98][108]. Demonstrated that adversaries can exploit weaknesses in GNN decision boundaries to bypass detection while continuing malicious activities. The rapidly evolving nature of blockchain fraud further increases the effectiveness of evasion attacks because attackers continuously adapt their strategies in response to newly deployed detection systems. Poisoning attacks represent an additional threat during the model training phase. In these attacks, malicious or incorrectly labeled samples are deliberately inserted into training datasets to corrupt the learning process and produce unreliable fraud detection models. Within blockchain environments, attackers may inject fraudulent transaction records, fabricated wallet addresses, or manipulated graph structures into training datasets [32][110] demonstrated that poisoning attacks can substantially reduce model accuracy and increase vulnerability during deployment. Since blockchain datasets frequently contain noisy labels and imperfect transaction records, distinguishing poisoned samples from legitimate data remains an important challenge for secure graph learning.

Overall, the literature indicates that node injection, edge manipulation, feature perturbation, evasion, and poisoning attacks collectively expose fundamental vulnerabilities in existing GNN-based blockchain fraud detection systems. These attacks significantly reduce detection accuracy, compromise model reliability, and threaten the practical deployment of graph learning technologies in financial security applications. Consequently, developing adversarially robust graph neural networks has become a critical research priority for securing blockchain fraud detection systems against increasingly sophisticated attack strategies.

Table 12. Comparative Analysis of Adversarial Attacks Against GNN-Based Fraud Detection

Adversarial attack	**Primary target**	**Effect on GNN-based detection**	**Relevance to blockchain fraud detection**	**Main research concern**
Node injection	Graph nodes and their attributes	Introduces misleading entities and alters neighborhood representations	Fraudulent addresses or intermediary entities may be introduced to obscure suspicious relationships	Limited evaluation on realistic blockchain transaction graphs
Edge manipulation	Graph connectivity	Adds, removes, or modifies relationships between nodes and changes message-	Transaction relationships can be manipulated to conceal connections between fraudulent and	Robustness to structural manipulation remains

		passing patterns	legitimate addresses	insufficiently studied
Feature perturbation	Node or edge features	Changes transaction or entity attributes used by the GNN	Small changes to transaction-related features may affect classification decisions	Difficulty determining realistic and constrained perturbations
Evasion attack	Model input during inference	Attempts to cause a trained model to misclassify fraudulent activity	Attackers may modify observable transaction behavior to avoid detection	Limited assessment under changing real-world attack behavior
Poisoning attack	Training data or graph structure	Corrupts the learning process and reduces model reliability	Malicious or manipulated transactions may influence the training graph and future predictions	Few studies jointly consider poisoning, dynamic graphs, and explainability

The Table 12 comparison shows that adversarial attacks can affect GNN-based blockchain fraud detection at different stages of the learning process. Structural attacks, such as node injection and edge manipulation, directly alter the relationships represented in the transaction graph, whereas feature perturbation modifies the information associated with existing entities or transactions. Evasion attacks primarily target the prediction stage, while poisoning attacks can influence the model during training by introducing misleading information into the learning process. Although these attacks differ in their mechanisms, they share a common objective of causing the detection model to produce unreliable predictions.

The literature in Figure 6 also indicates that the threat is particularly challenging in blockchain environments because transaction graphs are continuously updated and fraudulent actors can change their transaction behavior over time. Consequently, robustness evaluated on a static graph or under a single attack type may not adequately represent the conditions encountered in operational blockchain networks. Another important limitation is that adversarial robustness is often examined separately from explainability and dynamic graph learning. This separation makes it difficult to determine whether a model can remain accurate under manipulation while still providing reliable explanations for its decisions. Therefore, future GNN-based blockchain fraud detection frameworks need to consider structural and feature-based attacks together with temporal graph changes and explainability requirements.

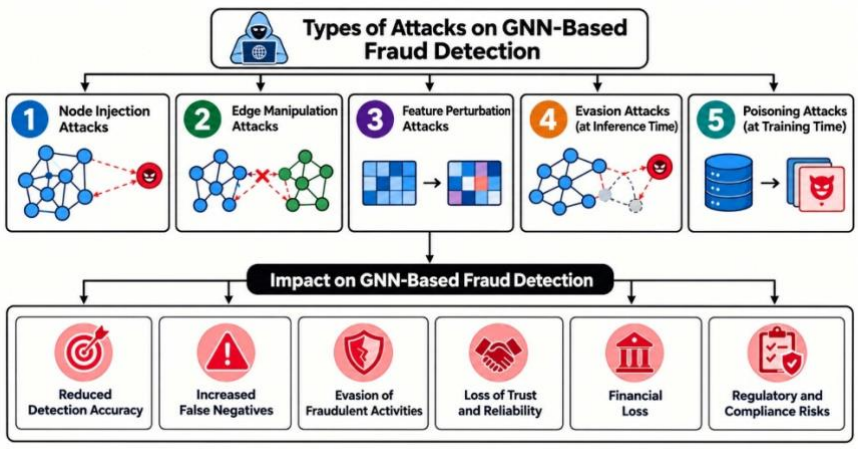

Figure 6. Adversarial Attacks against GNN-Based Blockchain Fraud Detection Taxonomy

3.6. Problem of Low Adversarial Robustness in Existing GNN Models

Graph Neural Networks (GNNs) have demonstrated strong potential for blockchain fraud detection; however, existing architectures remain vulnerable to adversarial perturbations. Changes to graph structures or node attributes can significantly influence model predictions, reducing the reliability and trustworthiness of fraud detection systems in real-world blockchain environments [56]. This limitation is particularly important because blockchain fraud detection operates in increasingly adversarial environments in which malicious

actors may deliberately modify transaction behaviors or graph relationships to evade detection. Consequently, improving adversarial robustness has become an important challenge in graph machine learning and trustworthy artificial intelligence [57].

Existing GNN architectures, including Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), are particularly affected because their learning mechanisms depend on neighborhood aggregation and message passing. In GCNs, malicious information introduced into neighboring nodes can propagate through the graph and influence node representations, potentially resulting in incorrect classifications even when relatively small perturbations are introduced [25][58]. Although GATs use attention mechanisms to improve the representation of important neighborhood information, studies have also demonstrated that adversarial manipulation of graph neighborhoods can substantially reduce their performance under attack [7][26]. These vulnerabilities demonstrate that improving representation learning alone is insufficient to guarantee reliable fraud detection in adversarial blockchain environments.

Another important challenge is the transferability of adversarial attacks across different GNN architectures. An adversarial example designed to mislead one model may also affect other architectures, even when the attacker does not have detailed knowledge of the target model. Studies have reported that attacks developed against one architecture can transfer to other graph learning models, including GraphSAGE and GAT [35][28]. This transferability suggests that several GNN architectures share common weaknesses, making it more difficult to develop defense mechanisms that remain effective across different models. In the context of blockchain fraud detection, the problem is further complicated by the continuous evolution of fraudulent behaviors and transaction structures, which may create additional opportunities for attackers to circumvent detection mechanisms.

Developing effective defense mechanisms also introduces a trade-off between adversarial robustness and predictive performance. Techniques such as adversarial training, graph purification, robust aggregation, and certified defense can improve resistance to adversarial manipulation, but they may increase computational requirements or reduce predictive performance under non-adversarial conditions [59]. This trade-off is particularly relevant to blockchain fraud detection because practical systems must simultaneously maintain high detection performance, withstand adversarial manipulation, adapt to changing transaction structures, and operate efficiently at scale.

Overall, the literature indicates that adversarial robustness remains insufficiently addressed in existing GNN-based blockchain fraud detection approaches. Current solutions often evaluate robustness against selected attack scenarios without fully considering the combined effects of adversarial manipulation, evolving graph structures, computational constraints, and model interpretability. The major limitations related to attack coverage, graph dynamics, dataset diversity, robustness metrics, explainability, and scalability are summarized in Table 13. This creates a gap for the development of GNN-based fraud detection approaches that can maintain reliable predictive performance under both normal and adversarial conditions while remaining suitable for dynamic blockchain environments.

Table 13. Low Adversarial Robustness Challenges in Existing GNN Models

Aspect	Existing limitation	Implication for blockchain fraud detection
Attack coverage	Many studies evaluate only selected attack types	Robustness against combined or unseen attacks remains uncertain
Graph dynamics	Robustness is often tested on static graph structures	Performance may decline as transaction relationships evolve
Dataset diversity	Evaluation is frequently limited to individual datasets	Cross-platform generalization is difficult to establish
Robustness metrics	No consistent set of robustness measures	Results are difficult to compare across studies
Explainability	Robustness and XAI are often evaluated separately	Reliability of explanations under attack remains unclear
Scalability	Computational cost is not consistently reported	Practical deployment on large transaction graphs remains challenging

3.7. Problem of Lack of Explainability in GNN-Based Fraud Detection

Graph Neural Networks (GNNs) have been found to be very good at modeling the complex network of transactions in a blockchain system, but their decision-making processes are opaque. Most GNN architectures were black-box, generating fraud predictions without offering a clear explanation of how specific transaction patterns, wallet behaviors, and/or graph structure affected the prediction. This opacity is a

major hurdle for investigators, auditors, regulators, and financial institutions that rely on clear and credible proof to pursue enforcement efforts [48]. The explainability problem stems from the fact that deep graph learning models use complex message-passing and neighborhood aggregation mechanisms. These mechanisms improve the predictive performance, but they do not make it easy to know which node features, neighboring transactions, or subgraph structures are most relevant for fraud predictions [109]. This results in users being provided with only one of the two classes and with little justification, which makes it not very useful for forensic investigations or regulatory decision-making [2].

One of the other challenges is the interpretation of decision items at the node level and at the subgraph level. Often, blockchain fraud does not happen in isolation through a single wallet address, but rather through the interplay and manipulation of several wallets, smart contracts, and transaction chains. These relationships are complex and need to be explained by locating influential nodes, suspicious transaction paths, and fraudulent communities in large transaction graphs, as illustrated in Figure 7. But in many cases, the existing explanation methods fail to consistently or stably yield explanations, and subgraph interpretation is one of the least-studied topics in explainable graph learning [13][111]. These restrictions limit the confidence in automated fraud detection systems and slow down their use in high-risk financial applications. Moreover, explainability has become a key ingredient in trust, accountability, and responsible AI. Financial institutions and regulatory agencies demand high predictive accuracy for fraud detection, but they also need to be able to give clear, auditable, and human-readable explanations to help investigate fraud and ensure regulatory compliance [4]. Enhancing the explainability of GNNs has emerged as one of the most crucial research challenges to developing trustworthy blockchain fraud detection systems. Thus, the explainability of a GNN is one of the most significant research areas for trustworthy blockchain fraud detection systems. The major explainability challenges, their impacts on blockchain fraud detection, and the corresponding research gaps are summarized in Table 14.

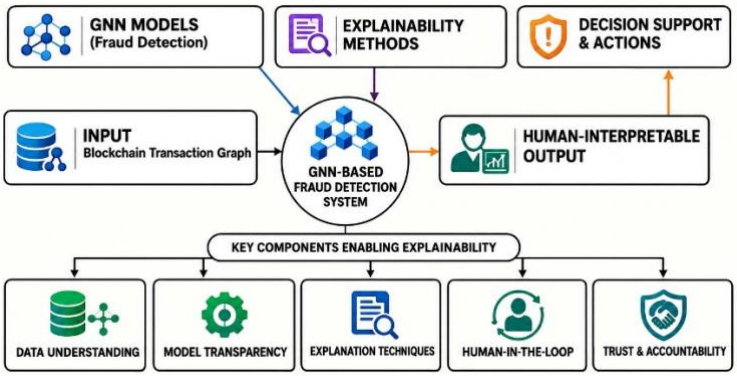

Figure 7. Explainability Framework for Graph Neural Network-Based Blockchain Fraud Detection

Table 14. Summary of Explainability Challenges in GNN-Based Blockchain Fraud Detection

Challenge	Description	Impact on Fraud Detection	Research Gap
Black-box GNN models	Complex graph representations make model decisions difficult to interpret.	Reduces user trust and adoption in financial applications.	Development of inherently interpretable GNN architectures.
Limited transparency	Existing models provide fraud predictions without sufficient evidence or justification.	Difficult for investigators and auditors to validate decisions.	Explainable decision-support mechanisms for fraud investigation.
Node-level explanation	Individual wallet classifications are difficult to explain because of neighborhood aggregation.	Low confidence in node classification results.	Stable and consistent node explanation methods.
Subgraph-level explanation	Fraud often involves multiple connected wallets and transaction paths.	Difficult to identify coordinated criminal activities.	Effective subgraph interpretation techniques.

Trust and accountability	Lack of understandable explanations limits regulatory acceptance.	Reduced confidence in AI-assisted fraud detection.	Human-centered explainability and accountable AI frameworks.

3.8. Problem of Regulatory and Compliance Requirements

In addition to performance, blockchain fraud detection systems need to meet ever-rising legal, ethical, and regulatory standards. With the increasing presence of AI in financial crime investigations, regulators are calling for more transparency, accountability, fairness, and auditability of automated decision-making systems [2]. The adoption of Graph Neural Networks for blockchain fraud detection is not just about prediction accuracy but also about meeting regulatory expectations. The explainability requirement from emerging AI governance frameworks is one of the key challenges. Some financial AI applications are considered to be high-risk systems under the European Union Artificial Intelligence Act, meaning that the law imposes obligations on the organization to make meaningful explanations about the decisions it can make using an automated system [112] (European Commission, 2024). Many GNN models are still very opaque, making such regulatory requirements hard to meet. Likewise, the process of fraud detection should be traceable and provide sufficient evidence for any regulatory review and/or legal investigations. Financial institutions should keep clear written minutes that will allow auditors to trace a company's thinking behind fraud classifications. Deep graph learning architectures can be complex, however, which often makes decisions hard to trace back and leads to low confidence in automated systems [111]. Anti-Money Laundering (AML) regulations add to the challenges of compliance. Decentralized finance platforms, cross-chain transactions, cryptocurrency mixers, and privacy-preserving technologies are all popular features used by criminals to hide their financial operations, making it difficult for regulators to monitor them (FATF, 2023; Europol, 2024). In addition, ethical issues such as fairness, privacy, accountability, and possible discrimination against legitimate users arise with automated fraud detection systems [94]. The regulatory and ethical issues suggest the need for future blockchain fraud detection systems to incorporate the concepts of explainability, auditability, fairness, and privacy along with predictive functionality as summarized in Table 15.

Table 15. Summary of Regulatory and Compliance Challenges

Challenge	Description	Impact	Research Gap
Explainability requirements	Financial regulations require interpretable AI decisions.	Black-box GNNs struggle to satisfy compliance requirements.	Regulation-aware explainable GNN frameworks.
Auditability	Fraud detection decisions must be traceable and verifiable.	Difficult auditing of deep graph models.	Auditable graph learning architectures.
Anti-Money Laundering compliance	Blockchain fraud detection should support Anti-Money Laundering investigations.	New laundering techniques reduce monitoring effectiveness.	Graph learning models aligned with AML regulations.
Ethical concerns	Automated decisions may introduce bias, privacy risks, and unfair classifications.	Legal and ethical concerns limit deployment.	Fair, privacy-preserving, and trustworthy AI models.

3.9. Problem of Benchmark Dataset Limitations

Benchmark datasets are essential for training and evaluating GNN-based blockchain fraud detection models; however, existing datasets remain limited by class imbalance, incomplete labels, restricted fraud coverage, and limited representativeness. The Elliptic Bitcoin dataset, one of the most widely used benchmarks, contains 203,769 transaction nodes and 234,355 directed edges across 49 temporal steps, but only 4,545 transactions are labeled illicit and 42,019 licit, while most remain unlabelled [63]. Elliptic++ extends this benchmark by incorporating transaction- and address-level information, providing a richer heterogeneous graph representation, although the large proportion of unknown labels remains a challenge for supervised learning [64]. More recently, Elliptic2 introduced approximately 49 million node clusters, 196 million transaction edges, and about 122,000 labeled subgraphs for money-laundering analysis; however, its emphasis on anti-money laundering limits its coverage of other fraud categories such as phishing, ransomware, Ponzi schemes, and smart-contract fraud [65]. Similar limitations occur in Ethereum fraud research, where datasets are frequently constructed independently from blockchain transaction records, resulting in differences in observation periods, labeling procedures, feature extraction, and graph construction that make cross-study comparison difficult [66]. Although some blockchain datasets provide timestamps or predefined temporal steps, they generally represent fixed historical observations rather than continuously

evolving transaction streams, limiting realistic evaluation of concept drift and continual learning [32]. Furthermore, few benchmark datasets provide standardized adversarial settings for evaluating node injection, edge manipulation, feature perturbation, poisoning, and evasion attacks [31][59]. Consequently, the literature demonstrates a need for larger and more diverse multi-platform benchmarks containing reliable fraud labels, temporal transaction evolution, multiple fraud categories, and standardized protocols for evaluating predictive performance, adversarial robustness, and explainability. The characteristics, strengths, and major limitations of representative blockchain fraud detection datasets are summarized in Table 16.

Table 16. Characteristics and Limitations of Representative Blockchain Fraud Detection Datasets

Dataset	Blockchain / Domain	Graph Characteristics	Temporal Information	Fraud / Label Information	Major Strength	Main Limitation
Elliptic	Bitcoin	203,769 transaction nodes; 234,355 directed transaction edges	49 temporal steps	4,545 illicit; 42,019 licit; majority unlabelled	Widely used public benchmark for illicit Bitcoin transaction detection	Severe label imbalance, many unknown labels, and primarily transaction-level AML evaluation
Elliptic++	Bitcoin	Transaction and wallet-address graphs; heterogeneous transaction–address relationships	49 temporal steps	Licit, illicit, and large unknown classes	Supports both transaction-level and address-level graph analysis	Large proportion of unknown labels and limited fraud-category diversity
Elliptic2	Bitcoin	Approx. 49M node clusters; 196M transaction edges; 122K labelled subgraphs	Historical blockchain graph	Suspicious and licit AML-related subgraphs	Large-scale subgraph benchmark suitable for money-laundering pattern analysis	Primarily AML-oriented and does not cover the full range of blockchain fraud types
Ethereum phishing datasets	Ethereum	Account/address transaction graphs constructed from Etherscan records	Timestamps available; study-specific observation windows	Phishing and legitimate addresses; labels vary by study	Useful for account-level phishing analysis and recent Ethereum activity	No universally standardized dataset; sampling, filtering, and labelling differ across studies
Study-specific blockchain datasets	Bitcoin, Ethereum, WAX and other networks	Varies by study	Varies	Fraud labels differ substantially	Enables investigation of specialised fraud scenarios	Limited reproducibility and difficult cross-study comparison

3.10. Problem of Evaluating Robust and Explainable GNN Models

The evaluation of trustworthy Graph Neural Networks is getting increasingly complicated, as the classic performance measures (such as accuracy, precision, recall, F-1 score, ROC-AUC) reflect merely a part of the model performance [98]. Several other criteria like adversarial robustness, explainability, scalability, fairness, and reproductivity of a blockchain fraud detection system have to be taken into account to assess their merit. But, at present, there is no universal acceptance of a framework for the evaluation of the

different approaches, which makes them hard to compare in objective terms [111]. There is a lack of standard measures of robustness. Previous research uses various metrics, such as attack success rate, robustness accuracy, perturbation budgets, certified robustness, and robustness degradation, which limits the ability to make cross-study comparisons [83]. Such limitations also apply to the evaluation of explainability, in which measures like fidelity, sparsity, stability, completeness, and human interpretability are used in varying ways across different studies [2][3]. The additional challenge of comparative evaluation is that the benchmark datasets vary, as do the way they are generated, plotted, and manipulated to create graphs, the feature engineering approaches, the methods used to generate attacks, and the experimental protocols used [113]. Furthermore, many published works do not provide access to source code or implementation details and/or benchmark datasets, resulting in large reproducibility hurdles and independent verification of the reported results [19]. These evaluation limitations need to be addressed to develop the trustworthy standards of benchmarking and fast-track progress towards a trustworthy blockchain fraud detection system. The major evaluation challenges, their impacts, and the corresponding research gaps are summarized in Table 17.

Table 17. Summary of Evaluation Challenges for Robust and Explainable GNNs

Challenge	Description	Impact	Research Gap
Lack of robustness metrics	No unified metric for evaluating adversarial robustness.	Difficult comparison across studies.	Standard robustness evaluation framework.
Lack of explainability metrics	Explanation quality is measured inconsistently.	Subjective evaluation of explainability methods.	Standard explainability metrics for GNNs.
Comparative evaluation	Different datasets and protocols prevent fair comparisons.	Conflicting performance claims.	Unified benchmarking protocols.
Reproducibility	Limited code, datasets, and implementation details.	Poor experimental reproducibility.	Open-source benchmarks and transparent reporting.

4. FINDINGS AND DISCUSSION

The previous sections present a systematic review of the major obstacles faced in pursuing adversarially robust and explainable Graph Neural Networks (GNNs) for fraud detection in dynamic blockchain networks. This section will organize the results of the analysis by critically examining the relations between the challenges identified, identifying research trends and gaps, analyzing the maturity of the approaches identified, and looking at the main challenges that remain and still hinder the real implementation of the trustworthy blockchain fraud detection system. This section differs from the previous ones, where individual issues were discussed in detail because here the research findings from the literature are combined so that one can obtain a general overview of what currently exists in the field.

4.1. Major Problems Identified in the Literature

Numerous blockchain security studies have established that detecting fraud within a blockchain has emerged from a classical classification problem to a multi-disciplinary research task that includes graph representation learning, cybersecurity, explainable artificial intelligence, adversarial machine learning, and financial regulation [14][114]. Although GNN-based fraud detection has made many strides, previous research shows that there are various recurring issues that still are not sufficiently addressed. In a technical sense, blockchain transaction networks can be characterized as large-scale graphs with heterogeneous entities, high-dimensional node and edge attributes, sparse interactions, and dynamically evolving topologies. In actual blockchain networks, these factors add significant complexity to calculations and affect their expandability of the current persistent security challenges identified as adversarial attacks were also found in the review. Surprisingly, although high predictive accuracy is desirable, existing GNN models are not robust against malicious manipulation, even being vulnerable to node injection, edge manipulation, feature perturbation, poisoning attacks, and evasion attacks, as can be seen in [4]. This is because blockchain fraudsters are constantly tweaking their methods to bypass detection systems, and ensuring that blockchain systems become more adversarial-robust by adapting their methods to evade detection is necessary for practical deployment. Explainable graph learning models are another important aspect that we found lacking in existing models. While effective classification performance of many GNN designs, the black-box nature of the process hinders transparency and limits their use in financial settings where regulatory compliance, auditability, and trust are critical [10][11]. The existing Explainability Methods are still very experimental

and have failed to deliver consistent explanations for real-world financial investigations. Another key limitation, in agreement with the review, was the use of benchmark datasets, which is still problematic in blockchain fraud detection research. Previously collected datasets tend to be static, very unequal in distribution, and have low counts of confirmed fraudulent transactions. Furthermore, there are very few benchmark datasets available to use for evaluating adversarial robustness and/or dynamic graph learning, which makes it challenging to compare the performance of graph learning algorithms objectively [38][47]. All of these results suggest that research has come a long way, making great strides in enhancing the accuracy of fraud detection but has received comparatively less focus on considering robustness, explainability, dynamic adaptability, and practical deployment requirements. As a result, numerous methods are still not viable with real-life blockchain networks.

4.2. Research Trends in GNN-Based Blockchain Fraud Detection

The reviewed literature published between 2021 and 2026 shows a clear shift from conventional machine-learning approaches based on manually engineered transaction features toward graph-based learning methods that can capture relationships among blockchain entities. GNNs have become increasingly important because they enable end-to-end learning from transaction graphs and can represent complex fraud patterns that are difficult to identify using traditional classifiers [16]. More recent studies increasingly focus on dynamic and temporal graph learning, adversarial robustness, and explainability. This indicates a broader transition from accuracy-oriented fraud detection toward trustworthy AI systems that are expected to combine predictive performance with robustness, transparency, scalability, and regulatory compliance. The trend therefore suggests that future blockchain fraud detection research will increasingly favor integrated frameworks rather than isolated improvements in individual model components.

4.3. Trade-Offs Between Detection Accuracy, Robustness and Explainability

The reviewed studies indicate important trade-offs among detection accuracy, adversarial robustness, explainability, and computational efficiency. More complex GNN architectures may improve predictive performance, but they often increase training cost, inference time, and deployment complexity. Similarly, adversarial defence methods such as adversarial training and graph purification can improve resistance to manipulation, but may reduce clean-data performance or require additional computational resources [69][70]. Explainability introduces another challenge because explanations must be sufficiently detailed to support investigation and regulatory requirements without oversimplifying complex graph relationships. Therefore, future research should treat accuracy, robustness, explainability, and efficiency as interconnected objectives rather than independent optimization problems.

4.4. Maturity of Existing Research

The review shows that different areas of GNN-based blockchain fraud detection have reached different levels of maturity. Graph representation learning and GNN-based fraud classification are comparatively well established, with numerous studies demonstrating their effectiveness in modelling complex transaction relationships [21]. Dynamic and temporal graph learning has also advanced considerably, although continual learning, concept drift, and streaming analysis remain open challenges. In contrast, adversarial robustness and explainability are less mature. Robust GNN methods are often evaluated under limited attack settings and rarely on realistic blockchain datasets, while explainability approaches remain largely focused on local explanations that may not fully support financial investigations [11]. Overall, the literature suggests that the most important unresolved direction is the integration of dynamic learning, adversarial robustness, and explainability within a unified and practically deployable GNN framework.

4.5. Relationship Between Robustness and Explainability

A key finding of this review is that adversarial robustness and explainability should not be considered independent research objectives. Instead, they're considered as supplementary traits of reliable AI systems. Robustness allows to ensure that fraud detection models are immune to adversarial manipulation while explainability allows investigators, financial institutions, and regulatory agencies to understand and verify model decisions [10]. If an explanation is not robust, it could describe a false prediction; if a prediction is robust but not explainable, either within or beyond the high-stakes financial domain, it may be hard to justify. Many recent studies have indicated that combining robustness and explainability in a single graph learning framework could lead to better system reliability and gain stakeholder’s trust. But there are very few studies which explicitly explore this relationship, creating a great opportunity for future study.

4.6. Open Research Challenges

Significant strides have been made, however, there are key issues that remain to be solved. These encompass the construction of scalable GNN architectures, that could process large dynamic graphs, with

provable robustness against adaptive adversarial attacks, benchmark datasets to ground truth the robustness of these models, unified robust evaluation protocols, and explainability frameworks. Moreover, the level of demand for transparency, auditability, fairness, and compliance keeps rising, which imposes even more demands on AI-driven fraud detection solutions [12]. This does not only need industry collaboration and partnerships between researchers in graph machine learning, cyber security, blockchain technology, XAI and financial regulation, it will also need a deeper interdisciplinary integration of these domains. Based on the results of this review, the key research recommendations are the need for further research into integrated, accurate and trustworthy GNN frameworks that can meet regulatory compliance, scalability, explainability, robustness and accuracy requirements simultaneously and deployed easily in practice.

5. FUTURE RESEARCH DIRECTIONS

The findings of this review indicate that although GNN-based approaches have significantly improved blockchain fraud detection, several challenges remain before these models can be considered trustworthy and suitable for real-world deployment. In particular, the literature reveals limitations in adversarial robustness, explainability, adaptation to evolving transaction graphs, benchmark datasets, scalability, and standardized evaluation. Based on these identified gaps, the following research directions are proposed as illustrated in Figure 8.

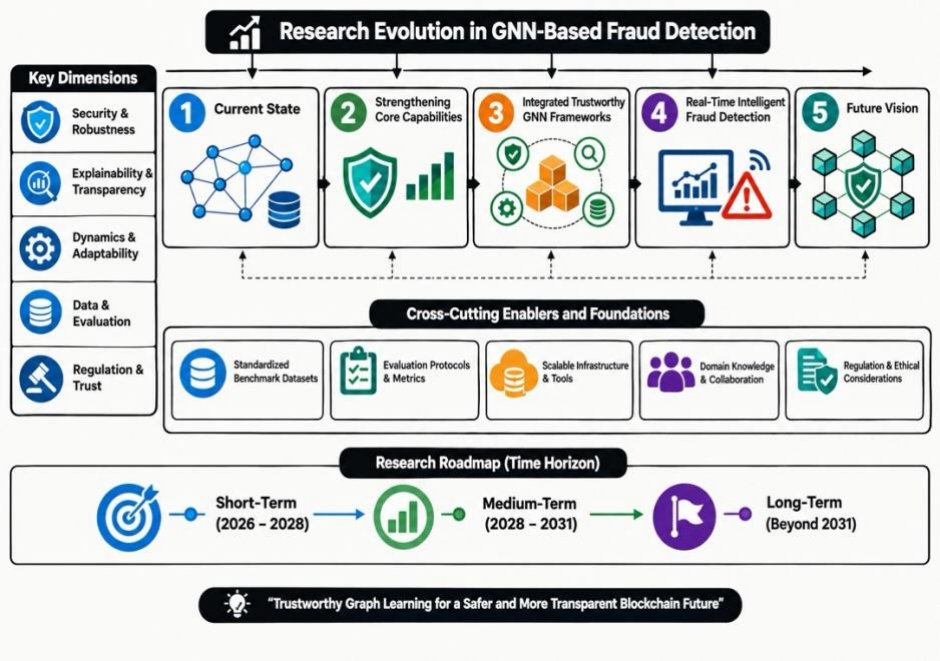

Figure 8. Research Roadmap for Adversarially Robust and Explainable GNN-Based Blockchain Fraud Detection

5.1. Development of Trustworthy Graph Learning Frameworks

Existing GNN-based fraud detection approaches primarily emphasize predictive performance, while robustness, explainability, fairness, uncertainty, and regulatory requirements are often considered separately. Future research should develop trustworthy graph-learning frameworks that jointly optimize these requirements rather than treating them as independent components. Such frameworks should integrate robust learning, interpretable decision-making, uncertainty estimation, fairness-aware modeling, and adaptive learning to support reliable deployment in real-world blockchain environments.

5.2. Next-Generation Adversarially Robust Graph Neural Networks

The review shows that existing GNNs remain vulnerable to structural and feature-based adversarial manipulation, while many defence mechanisms are evaluated against limited attack scenarios. Future studies should investigate adaptive adversarial training, robust message-passing mechanisms, graph purification, certified robustness, and self-supervised defence strategies. More importantly, these methods should be evaluated against multiple and adaptive attacks using realistic blockchain transaction graphs rather than relying primarily on static or synthetic benchmark settings.

5.3. Explainable and Human-Centred Graph Intelligence

Explainability remains insufficiently developed for practical blockchain fraud investigation. Existing approaches often provide local feature- or subgraph-level explanations that may not adequately communicate why a transaction, wallet, or group of entities has been classified as suspicious. Future explainable GNNs should provide interpretable transaction paths, influential entities, temporal fraud patterns, and prediction confidence while incorporating domain knowledge from financial investigators. Human-centred evaluation is also necessary to determine whether explanations are understandable and useful for investigation, auditing, and regulatory decision-making.

5.4. Dynamic and Continual Learning for Blockchain Networks

Blockchain transaction networks continuously evolve as new transactions, wallets, smart contracts, and decentralized applications emerge. Consequently, models trained on historical graph snapshots may lose effectiveness as transaction patterns and fraud strategies change. Future research should investigate temporal GNNs, continual and incremental learning, concept-drift adaptation, online graph learning, and streaming graph analytics. These approaches should enable fraud detection systems to adapt to emerging transaction behaviors without requiring complete model retraining.

5.5. Standardized Benchmark Datasets and Evaluation Protocols

The lack of representative and standardized benchmark datasets remains a major limitation in comparing blockchain fraud detection approaches. Future benchmark datasets should incorporate multiple blockchain platforms, diverse fraud categories, temporal transaction evolution, reliable labels, and realistic adversarial scenarios. Standardized evaluation protocols are also required to assess predictive performance, adversarial robustness, explanation quality, scalability, computational efficiency, and reproducibility under comparable experimental conditions. Such benchmarks would improve the reliability and comparability of future studies.

5.6. Towards Real-Time Autonomous Blockchain Fraud Detection Systems

Future research should move beyond offline transaction classification towards intelligent systems capable of continuously monitoring blockchain networks and detecting suspicious activities in real time. Such systems could integrate dynamic GNNs, adversarial defence, explainability, streaming graph processing, and decision-support mechanisms within a unified architecture. Important capabilities include early fraud detection, continual adaptation, interpretable alerts, scalable processing, and support for regulatory auditing. Achieving this objective will require collaboration across graph machine learning, blockchain security, cybersecurity, explainable AI, distributed systems, and financial regulation.

5.7. Proposed Research Roadmap

Based on the research gaps identified in this review, the development of GNN-based blockchain fraud detection can be viewed as a gradual transition from performance-oriented models towards trustworthy, adaptive, and deployable fraud intelligence systems. In the short term (2026–2028), research should prioritize robust GNN architectures, improved graph explainability, dynamic benchmark datasets, and standardized robustness and explainability metrics. In the medium term (2028–2031), attention should shift towards integrating robustness, explainability, and continual learning within unified frameworks, while improving scalability and cross-blockchain generalization. In the long term (beyond 2031), these developments could support autonomous real-time fraud intelligence systems capable of continuous adaptation, adversarial resilience, interpretable decision-making, and human–AI collaborative investigation. A consolidated summary of these research gaps, proposed directions, and expected outcomes is presented in Table 18.

Table 18. Research Gaps and Proposed Future Research Directions

Research gap identified in the review	Proposed research direction	Expected outcome
Robustness and explainability are often addressed separately	Trustworthy integrated GNN frameworks	Models that jointly support accuracy, robustness and interpretability
Vulnerability to structural and feature-based attacks	Adaptive and certified adversarial defences	Improved resilience under realistic and evolving attacks
Limited practical interpretability of GNN predictions	Human-centred graph explainability	Explanations useful for investigators, auditors and regulators
Static modelling of evolving	Dynamic, continual and	Adaptation to changing transaction

blockchain networks	concept-drift-aware learning	and fraud patterns
Limited and inconsistent benchmark datasets	Standardized multi-platform dynamic benchmarks	More reliable and reproducible model comparison
Limited real-time deployment and scalability	Streaming and autonomous fraud detection	Scalable real-time monitoring and early fraud identification

The proposed research directions emphasize that future progress should not be measured solely by improvements in fraud classification accuracy. The development of trustworthy blockchain fraud detection requires the coordinated advancement of dynamic learning, adversarial robustness, explainability, scalability, and standardized evaluation. Addressing these gaps will provide a stronger foundation for reliable GNN-based fraud detection in evolving blockchain environments.

6. CONCLUSION

Fraud on blockchain has grown in scale, complexity, and sophistication to pose great difficulty for current fraud detection programs. This systematic literature review (SLR) using a problem-centric approach systematically discusses the primary challenges of using Adversarially Robust and Explainable Graph Neural Networks (GNNs) for fraud detection in dynamic blockchain networks. The review pointed to significant problems in the growing number of blockchain frauds, dynamic and complex graph structures, low accuracy of blockchain detection, vulnerabilities to adversaries, lack of robustness in blockchain models, model explainability, regulatory and compliance needs, problems inherent in the available benchmarks, and the absence of a standard assessment methodology. These are still significant hurdles in the pursuit of trustworthy and viable blockchain fraud detection and prevention systems. The review also highlighted that these challenges are deeply interrelated and cannot be dealt with separately. Strengthening adversarial robustness while decreasing the transparency of the model can weaken the interpretability of the model, and vice versa. The increase in adversarial robustness will come at the expense of model transparency, and vice versa, as the improvement of the model's transparency is not sufficient to make it resistant to advanced attacks. Similarly, the difficulty of accessing realistic benchmark datasets, standard evaluation protocols, and scalable learning frameworks remains a significant challenge to fairly evaluate and practically apply current GNN-based methods. The fact that all these were found means there is a need for integrated solutions that tackle these aspects of robustness, explainability, scalability, adaptability, and regulatory compliance, all at once. In general, this review article offers a detailed problem-oriented survey of the new things learned in a blob of literature and sets an adequate groundwork for further research on trustworthy blockchain fraud detection. The study reveals some challenges that still need to be addressed and presents critical research gaps in this context that help guide future research and stakeholders' efforts in developing new types of GNN-based fraud detection systems that can run in secure and effective ways in a dynamic blockchain environment.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] P. Liu, Y. Zhang, and F. Biljecki, "Explainable spatially explicit geospatial artificial intelligence in urban analytics," *Environ. Plan. B Urban Anal. City Sci.*, vol. 51, no. 5, 2024, doi: 10.1177/23998083231204689.

[2] A. Lavecchia, "Explainable Artificial Intelligence in Drug Discovery: Bridging Predictive Power and Mechanistic Insight," 2025. doi: 10.1002/wcms.70049.

[3] J. Rao, S. Zheng, Y. Lu, and Y. Yang, "Quantitative evaluation of explainable graph neural networks for molecular property prediction," *Patterns*, vol. 3, no. 12, 2022, doi: 10.1016/j.patter.2022.100628.

[4] M. Nandan, S. Mitra, and D. De, "GraphXAI: a survey of graph neural networks (GNNs) for explainable AI (XAI)," 2025. doi: 10.1007/s00521-025-11054-3.

[5] Y. Cao, L. Huang, Y. Li, K. Jermsittiparsert, H. Ahmadi-Nezamabad, and S. Nojavan, "Optimal scheduling of electric vehicles aggregator under market price uncertainty using robust optimization technique," *International Journal of Electrical Power and Energy Systems*, vol. 117, 2020, doi: 10.1016/j.ijepes.2019.105628.

[6] V. Sivakumar, A. K. Verma, P. Kumar, N. Saw, and V. H. Krishnan, "Fraud Detection in Financial Transaction Using GNN," in *Proceedings of International Conference on Visual Analytics and Data Visualization, ICVADV 2025*, 2025. doi: 10.1109/ICVADV63329.2025.10960835.

[7] J. Qian and G. Tong, "Metapath-guided graph neural networks for financial fraud detection," *Computers and Electrical Engineering*, vol. 126, 2025, doi: 10.1016/j.compeleceng.2025.110428.

[8] N. R. S. Kumar, A. K. Verma, R. Sourabh, G. M. Babu, A. Kumar, and G. Sivagnanasekar, "Fraud Detection In Financial Transaction Using GNN With XGBoost," in *2025 4th OPJU International Technology Conference on Smart Computing for Innovation and Advancement in Industry 5.0, OTCON 2025*, 2025. doi: 10.1109/OTCON65728.2025.11070687.

[9] S. Verma and A. Sheel, "Blockchain for government organizations: past, present and future," 2022. doi: 10.1108/JGOSS-08-2021-0063.

[10] S. Khosravi, M. Kargari, B. Teimourpour, and M. Talebi, "Transaction fraud detection via attentional spatial–temporal GNN," *Journal of Supercomputing*, vol. 81, no. 4, 2025, doi: 10.1007/s11227-025-06983-8.

[11] A. B. Osmond and D. Burhanuddin, "GNN Feature Engineering for Credit Card Fraud Detection: A Comprehensive Research Framework," *Jurnal Sistem Cerdas*, vol. 8, no. 2, 2025, doi: 10.37396/jsc.v8i2.546.

[12] Y. Tian, G. Liu, J. Wang, and M. Zhou, "ASA-GNN: Adaptive Sampling and Aggregation-Based Graph Neural Network for Transaction Fraud Detection," *IEEE Trans. Comput. Soc. Syst.*, vol. 11, no. 3, 2024, doi: 10.1109/TCSS.2023.3335485.

[13] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, 2021, doi: 10.1109/TNNLS.2020.2978386.

[14] Y. Zhang *et al.*, "A Survey on Privacy in Graph Neural Networks: Attacks, Preservation, and Applications," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 12, 2024, doi: 10.1109/TKDE.2024.3454328.

[15] J. Chen, M. Ma, H. Ma, and H. Zheng, "A Survey of Privacy Security for Graph Neural Networks," *Journal of Cyber Security*, vol. 10, no. 3, 2025, doi: 10.19363/J.cnki.cn10-1380/tn.2025.05.08.

[16] F. Guan, T. Zhu, W. Zhou, and K. K. R. Choo, "Graph neural networks: a survey on the links between privacy and security," *Artif. Intell. Rev.*, vol. 57, no. 2, 2024, doi: 10.1007/s10462-023-10656-4.

[17] H. Yuan, H. Yu, S. Gui, and S. Ji, "Explainability in Graph Neural Networks: A Taxonomic Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, 2023, doi: 10.1109/TPAMI.2022.3204236.

[18] E. Dai *et al.*, "A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability," 2024. doi: 10.1007/s11633-024-1510-8.

[19] L. Hu, "GNN-Augmented RL for Fraud Detection in Decentralized Finance," *Applied and Computational Engineering*, vol. 152, no. 1, 2025, doi: 10.54254/2755-2721/2025.22856.

[20] G. Jayabalasamy, C. Pujol, and K. Latha Bhaskaran, "Application of Graph Theory for Blockchain Technologies," 2024. doi: 10.3390/math12081133.

[21] P. Li, Y. Xie, X. Xu, J. Zhou, and Q. Xuan, "Phishing Fraud Detection on Ethereum Using Graph Neural Network," in *Communications in Computer and Information Science*, 2022. doi: 10.1007/978-981-19-8043-5_26.

[22] A. Xiong *et al.*, "Ethereum phishing detection based on graph neural networks," *IET Blockchain*, 2023, doi: 10.1049/blc2.12031.

[23] H. Lv and Y. Ding, "Phishing detection on Ethereum via transaction subgraphs embedding," *IET Blockchain*, vol. 3, no. 4, 2023, doi: 10.1049/blc2.12034.

[24] L. Wang, M. Xu, and H. Cheng, "Phishing scams detection via temporal graph attention network in Ethereum," *Inf. Process. Manag.*, vol. 60, no. 4, 2023, doi: 10.1016/j.ipm.2023.103412.

[25] B. Fu, Y. Wang, and T. Feng, "CT-GCN+: a high-performance cryptocurrency transaction graph convolutional model for phishing node classification," *Cybersecurity*, vol. 7, no. 1, 2024, doi: 10.1186/s42400-023-00194-5.

[26] C. Bellei *et al.*, "The Shape of Money Laundering: Subgraph Representation Learning on the Blockchain with the Elliptic2 Dataset," Jul. 2024, http://arxiv.org/abs/2404.19109

[27] A. Asiri and K. Somasundaram, "Graph convolution network for fraud detection in bitcoin transactions," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-95672-w.

[28] M. Li, L. Jia, and X. Q. Su, "Global-local graph attention with cyclic pseudo-labels for bitcoin anti-money laundering detection," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-08365-9.

[29] J. Gao, B. Ansu, Q. Xia, C. Akwaboah, I. A. Obiri, and G. M. Ntuala, "Enhanced Temporal Graph Networks for Fraud Detection in Transactional Blockchain Networks," in *Proceedings of the International Joint Conference on Neural Networks*, 2025. doi: 10.1109/IJCNN64981.2025.11228825.

[30] T. Adam and F. Babič, "Identifying Illicit Activities in Blockchain Transaction Graph Networks," *Electronics (Switzerland)*, vol. 14, no. 23, 2025, doi: 10.3390/electronics14234599.

[31] S. Xiao, L. Zhang, Z. Tian, S. Su, J. Qiu, and R. Guo, "Pheromone-based graph embedding algorithm for Ethereum phishing detection," *Computer Networks*, vol. 260, 2025, doi: 10.1016/j.comnet.2025.111123.

[32] Z. Sheng, L. Song, and Y. Wang, "Dynamic Feature Fusion: Combining Global Graph Structures and Local Semantics for Blockchain Phishing Detection," *IEEE Transactions on Network and Service Management*, vol. 22, no. 5, 2025, doi: 10.1109/TNSM.2025.3576130.

[33] M. Ye *et al.*, "GraphSeqGuard: Detecting Ethereum phishing scams via temporally evolving graph sequences," *Knowl. Based. Syst.*, vol. 332, 2025, doi: 10.1016/j.knosys.2025.114895.

[34] H. Sui, J. Zhang, B. Chen, D. Wu, X. Sun, and S. Palaiahnakote, "EPAD: Ethereum phishing scam detection via graph contrastive learning," *Expert Syst. Appl.*, vol. 288, 2025, doi: 10.1016/j.eswa.2025.128227.

[35] W. Du, Q. Huang, and R. Xu, "Follow the vine to get the melon: A deep framework for blockchain phishing fraud detection," *Decis. Support Syst.*, vol. 199, 2025, doi: 10.1016/j.dss.2025.114555.

[36] H. Zhang, Y. Li, T. Zhu, C. Han, S. Wen, and Y. Xiang, "A scalable and semi-supervised approach for Ethereum fraud detection using Graph Neural Networks and K-Nearest Neighbors clustering," *Journal of Information Security and Applications*, vol. 95, 2025, doi: 10.1016/j.jisa.2025.104255.

[37] F. Ertam, "Near Real-Time Ethereum Fraud Detection Using Explainable AI in Blockchain Networks," *Applied Sciences*, vol. 15, no. 19, 2025.

[38] M. Ghosh, R. Halder, and J. Chandra, "TREAT: Temporal and Relational Attention-Based Tensor Representation Learning for Ethereum Phishing Users," *IEEE Trans. Serv. Comput.*, vol. 18, no. 3, 2025, doi: 10.1109/TSC.2025.3568277.

[39] W. Hou, B. Cui, Y. Chen, R. Li, and W. Song, "TSFF: A Triple-Stream Feature Fusion Method for Ethereum Phishing Scam Detection," *IEEE Internet Things J.*, vol. 12, no. 3, 2025, doi: 10.1109/JIOT.2024.3473771.

[40] L. Wang, Y. Mi, Y. Zhang, and J. Zhang, "PDHG: An Ethereum phishing detection approach via heterogeneous graph transformer," *Expert Syst. Appl.*, vol. 298, 2026, doi: 10.1016/j.eswa.2025.129919.

[41] Y. Wang, Q. Zheng, X. Li, L. Wang, and L. Lin, "CoSemiGNN: Blockchain fraud detection with dynamic graph neural networks based on co-association of semi-supervised," *Expert Syst. Appl.*, vol. 298, p. 129853, Mar. 2026, doi: 10.1016/J.ESWA.2025.129853.

[42] S. Xiao *et al.*, "A Deep Dynamic Graph Generative Framework for Blockchain Phishing Detection," *IEEE Trans. Dependable Secure Comput.*, 2026, doi: 10.1109/TDSC.2026.3674523.

[43] C. Li, Y. Zhang, R. Liu, Z. Zhao, H. Duan, and Q. Zeng, "Ethereum phishing fraud detection via dynamic graph neural network with anonymous walks," *Pattern Recognit.*, vol. 180, 2026, doi: 10.1016/j.patcog.2026.114343.

[44] H. Kang and S. G. Lee, "TRACE: Transaction representation learning for large-scale Ethereum phishing detection under limited labels," *Decis. Support Syst.*, vol. 210, 2026, doi: 10.1016/j.dss.2026.114753.

[45] W. Song, J. Tan, H. Wang, S. Han, M. Lai, and W. Yang, "GTSF : A novel ethereum phishing scams detection method based on gaining transaction semantics features," *Expert Syst. Appl.*, vol. 307, 2026, doi: 10.1016/j.eswa.2025.130950.

[46] Z. Chang, Y. Cai, X. F. Liu, Z. Xie, Y. Liu, and Q. Zhan, "Anomalous Node Detection in Blockchain Networks Based on Graph Neural Networks," *Sensors*, vol. 25, no. 1, Jan. 2025, doi: 10.3390/s25010001.

[47] Y. Li, W. Duan, X. Jin, and L. Ma, "The Design and Implementation of an Ethereum Account Fraud Detection Scheme," in *Communications in Computer and Information Science*, 2026. doi: 10.1007/978-981-95-4142-3_33.

[48] M. Ghosh, R. Halder, and J. Chandra, "A Systematic Review on Ethereum Phishing Scam Detection: Challenges, Empirical Insights, and Future Directions," *Blockchain: Research and Applications*, 2025, doi: 10.1016/j.bcra.2025.100424.

[49] J. Crisostomo, F. Bacao, and V. Lobo, "Machine learning methods for detecting smart contracts vulnerabilities within Ethereum blockchain − A review," 2025. doi: 10.1016/j.eswa.2024.126353.

[50] S. Motie and B. Raahemi, "Financial fraud detection using graph neural networks: A systematic review," *Expert Syst. Appl.*, vol. 240, p. 122156, Apr. 2024, doi: 10.1016/J.ESWA.2023.122156.

[51] Vivek Kale and Soumen Chakraborty, "Graph Neural Networks for Financial Fraud Detection: A Detailed Literature Review," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 12, no. 2, pp. 346–353, Apr. 2026, doi: 10.32628/CSEIT26121352.

[52] Y. Liu *et al.*, "Pick and choose: A GNN-based imbalanced learning approach for fraud detection," in *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, 2021. doi: 10.1145/3442381.3449989.

[53] S. X. Rao *et al.*, "xFraud: Explainable Fraud Transaction Detection," in *Proceedings of the VLDB Endowment*, 2021. doi: 10.14778/3494124.3494128.

[54] F. Shi, Y. Cao, Y. Shang, Y. Zhou, C. Zhou, and J. Wu, "H2-FDetector: A GNN-based Fraud Detector with Homophilic and Heterophilic Connections," in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, 2022. doi: 10.1145/3485447.3512195.

[55] M. Lu *et al.*, "BRIGHT - Graph Neural Networks in Real-time Fraud Detection," in *International Conference on Information and Knowledge Management, Proceedings*, 2022. doi: 10.1145/3511808.3557136.

[56] Y. Wang *et al.*, "Label Information Enhanced Fraud Detection against Low Homophily in Graphs," in *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, 2023. doi: 10.1145/3543507.3583373.

[57] G. Tong and J. Shen, "Financial transaction fraud detector based on imbalance learning and graph neural network," *Appl. Soft Comput.*, vol. 149, p. 110984, Dec. 2023, doi: 10.1016/J.ASOC.2023.110984.

[58] L. Ren *et al.*, "Dynamic graph neural network-based fraud detectors against collaborative fraudsters," *Knowl. Based. Syst.*, vol. 278, 2023, doi: 10.1016/j.knosys.2023.110888.

[59] J. Chen, Q. Chen, F. Jiang, X. Guo, K. Sha, and Y. Wang, "SCN_GNN: A GNN-based fraud detection algorithm combining strong node and graph topology information," *Expert Syst. Appl.*, vol. 237, 2024, doi: 10.1016/j.eswa.2023.121643.

[60] S. Wei and S. Lee, "Financial Anti-Fraud Based on Dual-Channel Graph Attention Network," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, no. 1, 2024, doi: 10.3390/jtaer19010016.

[61] D. Saldaña-Ulloa, G. De Ita Luna, and J. R. Marcial-Romero, "A Temporal Graph Network Algorithm for Detecting Fraudulent Transactions on Online Payment Platforms," *Algorithms*, vol. 17, no. 12, 2024, doi: 10.3390/a17120552.

[62] B. Wu, K. M. Chao, and Y. Li, "Heterogeneous graph neural networks for fraud detection and explanation in supply chain finance," *Inf. Syst.*, vol. 121, p. 102335, Mar. 2024, doi: 10.1016/J.IS.2023.102335.

[63] Z. Liu, Y. Wang, S. Wang, X. Zhao, H. Wang, and H. Yin, "Heterogeneous graphs neural networks based on neighbor relationship filtering," *Expert Syst. Appl.*, vol. 239, 2024, doi: 10.1016/j.eswa.2023.122489.

[64] L. Ren *et al.*, "Do not ignore heterogeneity and heterophily: Multi-network collaborative telecom fraud detection," *Expert Syst. Appl.*, vol. 257, 2024, doi: 10.1016/j.eswa.2024.124974.

[65] W. Zhang and C. Luo, "GE-GNN: Gated Edge-Augmented Graph Neural Network for Fraud Detection," *IEEE Trans. Big Data*, vol. 11, no. 4, 2025, doi: 10.1109/TBDATA.2025.3562486.

[66] S. Islam, G. Raj Gupta, A. Chakraborty, S. Singh, A. Soni, and C. Patle, "Detecting Fraudulent Transactions for Different Patterns in Financial Networks Using Layer Weigthed GCN," *Human-Centric Intelligent Systems*, vol. 5, no. 2, 2025, doi: 10.1007/s44230-025-00097-3.

[67] D. Saldana-Ulloa and G. De Ita Luna, "Fraudulent Event Detection via Temporal Graph Networks," in *Proceedings - 2025 IEEE Conference on Artificial Intelligence, CAI 2025*, 2025. doi: 10.1109/CAI64502.2025.00168.

[68] A. Kesharwani and P. Shukla, "FFDM - GNN:A Financial Fraud Detection Model using Graph Neural Network," in *2024 International Conference on Computing, Sciences and Communications, ICCSC 2024*, 2024. doi: 10.1109/ICCSC62048.2024.10830438.

[69] J. Chen and Y. Yang, "Real-time dynamic graph learning with temporal attention for financial fraud detection," *Front. Artif. Intell.*, vol. 9, 2026, doi: 10.3389/frai.2026.1774013.

[70] Y. Ji, Z. Zhang, X. Tang, J. Shen, X. Zhang, and G. Yang, "Detecting Cash-out Users via Dense Subgraphs," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022. doi: 10.1145/3534678.3539252.

[71] B. Rees, X. Wang, J. Eaton, O. Yilmaz, R. Ratzel, and D. Lasalle, "Accelerated GNN Training with DGL and RAPIDS cuGraph in a Fraud Detection Workflow," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022. doi: 10.1145/3534678.3542603.

[72] C. Deng, X. Li, Z. Feng, and Z. Zhang, "GARNET: Reduced-Rank Topology Learning for Robust and Scalable Graph Neural Networks," in *Proceedings of Machine Learning Research*, 2022.

[73] T. Wu *et al.*, "ERGCN: Data enhancement-based robust graph convolutional network against adversarial attacks," *Inf. Sci. (N. Y).*, vol. 617, 2022, doi: 10.1016/j.ins.2022.10.115.

[74] X. G. Wu, H. J. Wu, X. Zhou, X. Zhao, and K. Lu, "Towards Defense Against Adversarial Attacks on Graph Neural Networks via Calibrated Co-Training," *J. Comput. Sci. Technol.*, vol. 37, no. 5, 2022, doi: 10.1007/s11390-022-2129-2.

[75] B. Finkelshtein, C. Baskin, E. Zheltonozhskii, and U. Alon, "Single-node attacks for fooling graph neural networks," *Neurocomputing*, vol. 513, 2022, doi: 10.1016/j.neucom.2022.09.115.

[76] Z. Qiao, Z. Wu, J. Chen, P. Ren, and Z. Yu, "A Lightweight Method for Defense Graph Neural Networks Adversarial Attacks," *Entropy*, vol. 25, no. 1, 2023, doi: 10.3390/e25010039.

[77] M. Yao, H. Yu, and H. Bian, "Defending against adversarial attacks on graph neural networks via similarity property," *AI Communications*, vol. 36, no. 1, 2023, doi: 10.3233/AIC-220120.

[78] B. Xie *et al.*, "Adversarially Robust Neural Architecture Search for Graph Neural Networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2023. doi: 10.1109/CVPR52729.2023.00787.

[79] Q. Tao, J. Liao, E. Zhang, and L. Li, "A Dual Robust Graph Neural Network Against Graph Adversarial Attacks," *Neural Networks*, vol. 175, 2024, doi: 10.1016/j.neunet.2024.106276.

[80] H. Zhan and X. Pei, "Dealing with the unevenness: deeper insights in graph-based attack and defense," *Mach. Learn.*, vol. 113, no. 5, 2024, doi: 10.1007/s10994-022-06234-4.

[81] J. Zhang, Y. Hong, D. Cheng, L. Zhang, and Q. Zhao, "Defending adversarial attacks in Graph Neural Networks via tensor enhancement," *Pattern Recognit.*, vol. 158, 2025, doi: 10.1016/j.patcog.2024.110954.

[82] T. Wu *et al.*, "Understanding the robustness of graph neural networks against adversarial attacks," *Knowl. Based. Syst.*, vol. 323, 2025, doi: 10.1016/j.knosys.2025.113714.

[83] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, "On Explainability of Graph Neural Networks via Subgraph Explorations," in *Proceedings of Machine Learning Research*, 2021.

[84] A. Duval and F. D. Malliaros, "GraphSVX: Shapley Value Explanations for Graph Neural Networks," in *Lecture Notes in Computer Science*, 2021. doi: 10.1007/978-3-030-86520-7_19.

[85] A. Lucic, M. ter Hoeve, G. Tolomei, M. de Rijke, and F. Silvestri, "CF-GNNExplainer: Counterfactual Explanations for Graph Neural Networks," in *Proceedings of Machine Learning Research*, 2022.

[86] K. Amara *et al.*, "GraphFramEx: Towards Systematic Evaluation of Explainability Methods for Graph Neural Networks," in *Proceedings of Machine Learning Research*, 2022.

[87] S. Zhang, Y. Liu, N. Shah, and Y. Sun, "GStarX: Explaining Graph Neural Networks with Structure-Aware Cooperative Games," in *Advances in Neural Information Processing Systems*, 2022. doi: 10.52202/068431-1440.

[88] S. Akkas and A. Azad, "GNNShap: Scalable and Accurate GNN Explanation using Shapley Values," in *WWW 2024 - Proceedings of the ACM Web Conference*, 2024. doi: 10.1145/3589334.3645599.

[89] M. U. Khan, M. T. Shahid, M. Ashraf, M. Riaz, U. Rahman, and A. Iqbal, "Financial Fraud Detection with AI: A Machine Learning-Based Approach for Securing Digital Transactions," *Global Research Journal of Natural Science and Technology*, 2025, doi: 10.53762/grjnst.03.03.09.

[90] Q. Wang, Y. Shen, and H. Dong, "Hypergraph-based contrastive learning for enhanced fraud detection," *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1703135.

[91] L. Xiao, H. Liu, Z. Lv, Y. Chen, Z. Lin, and Y. Du, "Reinforcement-Learning-Based APT Defense for Large-Scale Smart Grids," *IEEE Internet Things J.*, vol. 12, no. 9, 2025, doi: 10.1109/JIOT.2024.3519134.
[92] G. M. Ntuala *et al.*, "TIEBN: An Eigenvalue-Driven Blockchain Network for Anomaly Detection," in *Lecture Notes in Computer Science*, 2026. doi: 10.1007/978-981-95-3055-7_13.
[93] J. Xia, L. Wu, J. Chen, B. Hu, and S. Z. Li, "SimGRACE: A Simple Framework for Graph Contrastive Learning without Data Augmentation," in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, 2022. doi: 10.1145/3485447.3512156.
[94] B. Deng, J. Chen, Y. Hu, C. Chen, and T. Zhang, "Mutual GNN-MLP distillation for robust graph adversarial defense," *Neural Networks*, vol. 189, 2025, doi: 10.1016/j.neunet.2025.107513.
[95] D. An *et al.*, "Reinforcement learning-based secure training for adversarial defense in graph neural networks," *Neurocomputing*, vol. 630, 2025, doi: 10.1016/j.neucom.2025.129704.
[96] W. Gu, M. Sun, B. Liu, K. Xu, and M. Sui, "Adaptive Spatio-Temporal Aggregation for Temporal Dynamic Graph-Based Fraud Risk Detection," 2024.
[97] Y. Liu, Z. Sun, and W. Zhang, "Improving fraud detection via hierarchical attention-based Graph Neural Network," *Journal of Information Security and Applications*, vol. 72, 2023, doi: 10.1016/j.jisa.2022.103399.
[98] J. Li, Y. Chang, Y. Wang, and X. Zhu, "Tracking down financial statement fraud by analyzing the supplier-customer relationship network," *Comput. Ind. Eng.*, vol. 178, 2023, doi: 10.1016/j.cie.2023.109118.
[99] M. An, Q. Fan, H. Yu, and H. Zhao, "Blockchain technology research and application: a systematic literature review and future trends," Jun. 2023, doi: 10.47852/bonviewJDSIS32021403.
[100] W. Zhao, S. Alwidian, and Q. H. Mahmoud, "Adversarial Training Methods for Deep Learning: A Systematic Review," 2022. doi: 10.3390/a15080283.
[101] L. Shi and W. Liu, "A Closer Look at Curriculum Adversarial Training: From an Online Perspective," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024. doi: 10.1609/aaai.v38i13.29418.
[102] X. Zhang, Y. Xu, C. Jiang, L. Shen, and X. Liu, "MoleMCL: a multi-level contrastive learning framework for molecular pre-training," *Bioinformatics*, vol. 40, no. 4, 2024, doi: 10.1093/bioinformatics/btae164.
[103] Andrew Harper and Miriam D. Lee, "Scalable Blockchain Fraud Detection Using Spatial-Temporal Graph Neural Networks," *Frontiers in Applied Physics and Mathematics*, vol. 2, no. 1, 2025, doi: 10.71465/fapm141.
[104] F. Ares-Robledo, H. Rifà-Pous, and R. Clarisó, "Graph neural networks for anomaly detection: a systematic review of dynamic temporal approaches," *Artif. Intell. Rev.*, vol. 59, no. 5, May 2026, doi: 10.1007/s10462-026-11532-7.
[105] S. V. K. Gummadi, "Temporal Graph Neural Networks for Real-Time Fraud Detection in Cross-Border Transactions," *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, vol. 6, 2025, doi: 10.63282/3050-9262.ijaidsml-v6i4p111.
[106] C. Ju, X. Ma, and B. Dong, "Real-time Cross-border Payment Fraud Detection Using Temporal Graph Neural Networks: A Deep Learning Approach," *Academic Journal of Sociology and Management*, vol. 3, no. 2, 2025, doi: 10.70393/616a736d.323639.
[107] F. Wang, Y. Chen, H. Gao, Q. Li, and C. Wang, "Self-supervised contrastive representation learning for classifying Internet of Things malware," *Eng. Appl. Artif. Intell.*, vol. 150, 2025, doi: 10.1016/j.engappai.2025.110299.
[108] A. E. J. Villegas-Espinoza and J. I. Necochea-Chamorro, "Using Deep Learning in Student Performance Prediction: A Systematic Review," 2025. doi: 10.18421/TEM143-51.
[109] M. Krishnan, Y. Lim, S. Perumal, and G. Palanisamy, "Detection and defending the XSS attack using novel hybrid stacking ensemble learning-based DNN approach," *Digital Communications and Networks*, vol. 10, no. 3, 2024, doi: 10.1016/j.dcan.2022.09.024.
[110] G. Zhu, D. He, H. An, M. Luo, and C. Peng, "The governance technology for blockchain systems: a survey," 2024. doi: 10.1007/s11704-023-3113-x.
[111] A. Karamanou, P. Brimos, E. Kalampokis, and K. Tarabanis, "Explainable Graph Neural Networks: An Application to Open Statistics Knowledge Graphs for Estimating House Prices," *Technologies (Basel).*, vol. 12, no. 8, 2024, doi: 10.3390/technologies12080128.
[112] J. Kim, S. Lee, Y. Kim, S. Ahn, and S. Cho, "Graph Learning-Based Blockchain Phishing Account Detection with a Heterogeneous Transaction Graph," *Sensors*, vol. 23, no. 1, 2023, doi: 10.3390/s23010463.
[113] N. M. A. Chisty *et al.*, "Blockchain, Big Data, and Deep Learning-based Fraud Detection System for Credit Card Fraud," in *Coresource 4*, 2026. doi: 10.2174/9789815324211126010011.
[114] M. Zhao, G. Wang, Q. Fu, X. Guo, and T. Li, "Deep Reinforcement Learning-Based Air Defense Decision-Making Using Potential Games," *Advanced Intelligent Systems*, vol. 5, no. 10, 2023, doi: 10.1002/aisy.202300151.

BIOGRAPHIES OF AUTHORS

Oluwaseun Adeniyi Ojerinde is an Associate Professor of Computer Science at the Federal University of Technology Minna (FUTMinna), Nigeria. He holds a BSc in Computer Technology from Babcock University, Nigeria (2006), and an MSc and PhD in Mobile Communication Systems from Loughborough University, UK (2008 and 2014). His research interests span blockchain technology, digital transformation and innovation, process optimisation, antenna systems, radio propagation, MIMO systems, 5G technology, on-body systems, telecommunications, Specific Absorption Rate (SAR), radiation, and computer science education. He has authored over 70 peer-reviewed publications in reputable journals and international

conferences and has supervised master's and doctoral students. He is a member of the Computer Professionals Registration Council of Nigeria (CPN) and the Institute of Electrical and Electronics Engineers (IEEE), and serves as Tech Lead at TTF Limited, where he oversees product delivery and technical partnerships. He can be contacted at email: o.ojerinde@futminna.edu.ng

Ramatu Abubakar is a postgraduate researcher in Computer Science at the Federal University of Technology, Minna (FUTMinna), Nigeria. She obtained her undergraduate degree from the Federal University of Technology, Minna, and is currently pursuing a master's degree in computer science at the same institution. She has been actively involved in academic research and scholarly writing, with research activities spanning blockchain fraud detection, phishing detection, cybersecurity, digital phenotyping, ontology engineering, and other applications of artificial intelligence. Her work reflects a growing interest in developing trustworthy, adaptive, and explainable AI solutions for real-world problems. She is particularly interested in advancing her research career through postgraduate studies, academic collaborations, publications, and the practical application of artificial intelligence and emerging technologies. She can be contacted at email: ramatu.abubakar@st.futminna.edu.ng.

Abimbola Susan Ajagun received her B. Eng. and M. Eng. degrees from the Federal University of Technology, Minna, Nigeria, in 2010 and 2016, respectively. She is currently a Ph.D. scholar in the Department of Energy and Electrical Engineering at Hohai University, Nanjing, China with a background in Electrical & Computer Engineering. She is also a lecturer at the Federal University of Technology, Minna, Nigeria. Ajagun is a passionate advocate for clean energy and gender equality in STEM, and she founded the Female in Clean Energy (FiCE) Foundation. Her research interests include power system planning and operations. She can be contacted at email: bimbo.ajagun@futminna.edu.ng.

Enesi Femi Aminu is a Senior Lecturer in the Department of Computer Science at the Federal University of Technology, Minna, Niger State, Nigeria, with over fifteen years' experience in teaching and research. He obtained his PhD degree in Computer Science from Federal University of Technology, Minna, Niger State in 2023. He specialized in Ontology and reasoning algorithm design, AI and machine learning. He has published over 40 peer reviewed publications in reputable journals, international conferences, and book chapters. He has supervised master's and doctorate students. He can be contacted at email: enesifa@futminna.edu.ng.