Anomalous Cryptocurrency Transaction Detection: A Systematic Review of Techniques, Datasets, and Future Directions

Md Saidul Islam[a], Syed Mohammed Shamsul Islam[a,b,c,*], Md Zakir Hossain[d,*], Mohiuddin Ahmed[e], Iqbal H. Sarker[a]

[a]*Discipline of Computing and Security, School of Science, Edith Cowan University, Perth, Western Australia, Australia*
[b]*Department of Computing and Information Systems, Daffodil International University, Dhaka, Bangladesh*
[c]*Department of Computer Science and Software Engineering, The University of Western Australia, Perth, Western Australia, Australia*
[d]*School of Medicine and Psychology, The Australian National University, Canberra, Australian Capital Territory, Australia*
[e]*School of Computer Science and Information Technology, Adelaide University, Adelaide, South Australia, Australia*

Abstract

The rapid adoption of blockchain-based financial systems has been accompanied by a surge in illicit activities, including money laundering, ransomware payments, phishing scams, and terrorist financing, necessitating robust anomalous transaction detection mechanisms. Detecting anomalies in cryptocurrency transactions is critical, as undetected illicit activity can result in significant economic losses and undermine trust in digital financial systems. This systematic review examines the state-of-the-art in cryptocurrency anomaly detection, with particular focus on methodological developments between 2008 and December 2025. A PRISMA-guided systematic literature search was conducted across IEEE Xplore, Scopus, Web of Science, ACM Digital Library, Google Scholar, and SpringerLink. From an initial set of 450 records, 32 empirical studies were selected after rigorous screening and eligibility assessment and included in the qualitative synthesis. Unlike prior surveys, this review provides a focused synthesis of empirical cryptocurrency transaction studies, a taxonomy of anomaly types, and a critical assessment of dataset bias and evaluation practices. The literature reveals a clear methodological shift from traditional feature-engineered machine learning approaches (e.g., Random Forest, XGBoost, and Support Vector Machines) toward graph-based deep learning architectures. Graph Neural Networks (GNNs), particularly Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), demonstrate competitive performance by capturing relational dependencies among blockchain addresses, while temporal graph models and hybrid GNN–transformer architectures enhance the detection of evolving, multi-hop laundering schemes. Unsupervised and semi-supervised approaches address the challenge of limited labeled data but introduce trade-offs in interpretability. Emerging research directions include privacy-preserving federated learning and cross-chain detection frameworks. Despite some studies reporting accuracies exceeding 90%, the field faces several limitations, including dataset bias, lack of standardized multi-chain benchmarks, inconsistency in evaluation metrics, limited adversarial robustness testing, scalability constraints, and insufficient explainability for regulatory compliance. This review aims to provide researchers and practitioners with a structured synthesis of current methodologies, a comprehensive taxonomy of anomalies, and a detailed roadmap for transitioning from experimental validation to real-world, scalable deployment.

Keywords: Blockchain, Cryptocurrency, Graph Neural Networks, Anomaly Detection, Deep learning

1. Introduction

1.1. Background and Motivation

Detecting anomalies in cryptocurrency transactions has emerged as a critical challenge in the rapidly evolving landscape of digital finance. Anomaly detection refers to "the problem of finding patterns in data that do not conform to expected behavior" [1, 2]. Although foundational research in anomaly detection predates cryptocurrency systems, this study focuses on empirical developments in cryptocurrency transaction analysis from 2008 onward, corresponding to the introduction of Bitcoin [3]. The importance of detecting anomalies in cryptocurrency transactions lies in the risk that undetected illicit activity may cause significant economic losses and undermine trust in digital financial systems. Bitcoin, Ethereum, and other cryptocurrencies operate on decentralized blockchain infrastructure, enabling peer-to-peer value transfer without centralized intermediaries [3, 4]. Blockchain technology employs distributed ledger systems where transactions are cryptographically validated, ensuring immutability and transparency. However, these benefits are accompanied by the pseudonymous nature of blockchain addresses, which has created significant opportunities for illicit financial activity.

The scale of cryptocurrency-related financial crime has reached substantial proportions. According to industry reports [5], approximately $40.9 billion in cryptocurrency flowed to illicit addresses in 2024, with money laundering, Ponzi schemes, and ransomware accounting for a large proportion of these activities. Criminal actors increasingly employ privacy-enhancing mechanisms, including mixing services,

*Corresponding authors
URL: syed.islam@ecu.edu.au (Syed Mohammed Shamsul Islam), zakir.hossain@anu.edu.au (Md Zakir Hossain)

tumblers, privacy coins, and cross-chain bridges, to obfuscate transaction traces [6, 7].

From a computational perspective, blockchain data is inherently graph-structured, where wallet addresses represent nodes and transactions form directed edges. This natural graph representation enables the application of graph theory, network analytics, and machine learning methodologies for anomaly detection [8, 9]. Early research predominantly employed conventional supervised learning models such as Random Forest, Support Vector Machines, and XGBoost, trained on manually engineered transaction features [10, 11]. Subsequent studies further demonstrated the effectiveness of supervised learning for cryptocurrency fraud detection [12]. Contemporary methodologies increasingly leverage deep learning and graph-based architectures, including Graph Neural Networks (GNNs), Graph Convolutional Networks (GCNs) [13], Graph Attention Networks (GATs) [14], and temporal graph models [15]. More recent work also explores graph-based representation learning approaches, including contrastive learning techniques, for improved modeling of transaction networks [16]. The emergence of decentralized finance (DeFi), smart contracts, and cross-chain ecosystems has further increased system complexity [17, 18]. Unlike traditional banking systems with centralized oversight, blockchain ecosystems require automated, real-time anti-money laundering (AML) compliance and cybersecurity monitoring frameworks. Consequently, anomalous cryptocurrency transaction detection has evolved into an interdisciplinary field integrating financial criminology, artificial intelligence, graph mining, and regulatory technology [7].

1.2. The Evolving Landscape of Financial Crime

Cryptocurrency-enabled financial crime manifests in diverse forms, each presenting unique detection challenges [19]. Money laundering remains one of the most significant threats, as criminals exploit cryptocurrency's pseudonymity to layer illicit funds through complex transaction chains, often involving mixing services and privacy-enhancing mechanisms [6, 7]. The absence of centralized oversight enables placement, layering, and integration stages to occur across multiple blockchain networks. Ransomware attacks increasingly demand cryptocurrency payments, with victims instructed to transfer funds to specific wallet addresses, after which these payments traverse laundering networks to obfuscate their origin [20].

Smart contract-enabled Ponzi schemes on Ethereum and other platforms have proliferated, exploiting automated execution to perpetrate fraud at scale [21, 22]. Attackers also employ social engineering techniques to obtain private keys or trick users into approving malicious transactions, leading to significant financial losses through phishing scams [23]. Decentralized finance (DeFi) ecosystems, characterized by complex smart contract interactions and cross-chain bridges, introduce additional attack surfaces, including flash loan attacks, oracle manipulation, and governance exploits [17, 24].

Cryptocurrency remains a primary payment mechanism for illicit goods and services on darknet platforms [25]. In addition, both state and non-state actors have increasingly leveraged cryptocurrency systems to bypass regulatory controls and facilitate cross-border financial activities outside traditional oversight mechanisms [5].

1.3. The Imperative for Anomaly Detection

Detecting anomalous cryptocurrency transactions is critically important for three fundamental reasons. First, global cryptocurrency adoption continues to accelerate, with both institutional and retail participants increasingly integrated into digital asset ecosystems [4, 5]. As transaction volumes grow, so do associated illicit finance flows. Without robust detection mechanisms, decentralized ecosystems face reputational damage and potential systemic risks.

Second, international regulatory frameworks, particularly the Financial Action Task Force (FATF) "Travel Rule," mandate anti-money laundering (AML) and counter-terrorist financing controls for Virtual Asset Service Providers (VASPs) [26]. Automated anomaly detection systems are essential for compliance and suspicious activity reporting obligations. Regulators increasingly require explainable and auditable detection decisions, which poses challenges for black-box machine learning models [27, 28].

Third, blockchain-based financial platforms are becoming embedded in global digital infrastructure. Threats including ransomware operations, decentralized finance (DeFi) exploits, and smart contract vulnerabilities carry significant economic and operational consequences [20, 17]. Advanced detection systems shift risk mitigation from forensic post-incident analysis toward proactive and real-time prevention.

1.4. Scope, Objectives, and Unique Contributions of This Review

Previous survey studies have examined anomaly detection in general graphs [8, 29] and fraud detection using machine learning [30, 31]. However, existing reviews do not comprehensively target the cryptocurrency transaction domain with its unique characteristics [32].

This review exclusively addresses cryptocurrency blockchain transactions, considering UTXO and account-based models, mixing services, and cross-chain bridges. It covers the evolution from traditional machine learning (pre-2015) to graph-based methods (2015–2020) and to temporal and hybrid architectures (2020–2025). The review includes deep learning techniques (autoencoders, graph neural networks, graph convolutional networks, graph attention networks, temporal networks, transformers, and federated learning) that are transforming the field.

Unlike previous reviews, this study addresses emerging threats in cross-chain bridges and decentralized finance protocols, includes a systematic evaluation of code availability and dataset transparency [33], follows rigorous PRISMA guidelines [34], and analyzes detection techniques from four perspectives: applications, machine learning algorithms, performance metrics, and anomaly classification types.

Distinct contributions of this review are as follows:

1. Focuses exclusively on empirical cryptocurrency transaction studies, rather than combining general anomaly detection with cryptocurrency-specific research.

2. Quantifies dataset usage and evaluation metric bias across the literature, revealing a strong concentration on benchmark datasets such as Elliptic and the widespread use of accuracy as a primary evaluation metric.
3. Synthesizes a comprehensive anomaly taxonomy specifically adapted for cryptocurrency transactions, distinguishing point, contextual, and collective anomalies.
4. Compares learning paradigms (supervised, unsupervised, and semi-supervised) and their prevalence in cryptocurrency contexts.
5. Highlights reproducibility and benchmark limitations through a systematic assessment of code availability and dataset transparency.

1.5. Paper Organization

The remainder of this paper is organized as follows. Section 2 presents a taxonomy of anomalies in cryptocurrency transactions. Section 3 describes the systematic review methodology based on PRISMA guidelines. Section 4 presents the results of the review structured around four research questions. Section 5 discusses the key findings, challenges, and future research directions. Section 6 concludes the paper with implications and recommendations.

2. Taxonomy of Anomalies in Cryptocurrency Transactions

Anomaly definition requires careful consideration in cryptocurrency contexts, where the relationship between statistical deviation and illicit activity is complex and context-dependent [8]. A statistically anomalous transaction may be entirely legitimate, while illicit transactions may be deliberately engineered to appear statistically normal. This section presents a comprehensive taxonomy of anomalies adapted from foundational literature. Figure 1 synthesizes the taxonomy of anomalies in cryptocurrency transactions into a conceptual framework, adapted from prior studies [8, 29].

2.1. Illicit vs. Anomalous Distinction

The distinction between illicit and anomalous transactions is fundamental. Illicit transactions violate laws or regulations but may be engineered to appear statistically normal through techniques such as structuring or mimicking legitimate behavioral patterns. Conversely, statistically anomalous transactions (e.g., unusually large legitimate transfers by institutional investors) may be entirely lawful. Detection systems must navigate this distinction by identifying anomalies as candidates for further investigation rather than as definitive indicators of illegality [29].

2.2. Point Anomalies

Point anomalies are individual transactions that deviate significantly from the global data distribution. Examples include transactions with extreme values, unusual fee structures, or unexpected interactions with previously dormant addresses [35]. Point anomalies are relatively straightforward to detect using statistical or feature-based approaches but may result in high false-positive rates in volatile cryptocurrency environments.

2.3. Contextual Anomalies

Contextual anomalies depend on context, where a transaction that appears normal in isolation may be anomalous given its temporal, behavioral, or network context [36]. For example, a transaction of typical size may be anomalous if it occurs at an unusual time, originates from an address with no prior activity, or coincides with abnormal network conditions. Detecting contextual anomalies requires modeling temporal dependencies and behavioral patterns across multiple dimensions.

2.4. Collective Anomalies

Collective anomalies involve groups of transactions that exhibit suspicious patterns collectively, while appearing normal individually. Examples include structured layering sequences used in money laundering [6], circular transaction patterns associated with wash trading [37], and temporal bursts of activity indicative of automated attacks [38]. Detecting such anomalies requires analysis of transaction sequences, graph structures, and temporal dynamics, with graph-based methods increasingly demonstrating strong effectiveness [13, 15].

2.5. Cryptocurrency-Specific Anomaly Types

Table 1 presents the cryptocurrency-specific anomaly types identified in this review. These categories represent commonly observed patterns in illicit cryptocurrency activities and provide a foundation for designing detection methodologies.

Money laundering chains and mixing service interactions are prominently represented in the literature, reflecting the maturity of research on traditional financial crime adapted to cryptocurrency ecosystems. In contrast, cross-chain laundering and decentralized finance (DeFi)-related exploits represent emerging areas of study, with increasing attention in recent years. This trend highlights the evolving nature of cryptocurrency-based financial crime and the need for detection methodologies capable of addressing complex, multi-platform transaction patterns.

3. Methodology

In this study, a Systematic Literature Review (SLR) was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [34]. PRISMA provides a structured framework for conducting and reporting systematic reviews, ensuring transparency, reproducibility, and methodological rigor.

3.1. Research Questions

This SLR systematically analyzes and synthesizes empirical computational techniques applied in cryptocurrency anomaly detection from January 2008 through December 2025. The following research questions (RQs) are defined:

- **RQ1:** What are the main applications of anomaly detection in cryptocurrency transactions?
- **RQ2:** What types of machine learning and deep learning algorithms are applied in cryptocurrency anomaly detection?

Table 1: Cryptocurrency-Specific Anomaly Types

Anomaly Type	Description	Detection Approach	Key References
Money Laundering Chains	Multi-hop transaction sequences obscuring fund origin	Graph-based, temporal analysis	[6, 7]
Mixing Service Interaction	Funds aggregated from multiple sources and redistributed	Behavioral analysis, entropy-based methods	[6]
Peel Chains	Structured value extraction with small amounts incrementally transferred	Sequential pattern mining	[10, 6]
Circular Transactions	Repeated transaction loops indicating potential wash trading	Graph cycle detection	[37]
Temporal Bursts	Sudden spikes in transaction activity suggesting automated behavior	Burst detection algorithms	[38, 36]
Smart Contract Exploits	Flash loan attacks and contract manipulation strategies	Smart contract analysis and anomaly scoring	[17, 23]
Cross-Chain Laundering	Funds transferred across multiple blockchains via bridges	Cross-chain graph analysis	[39, 24]

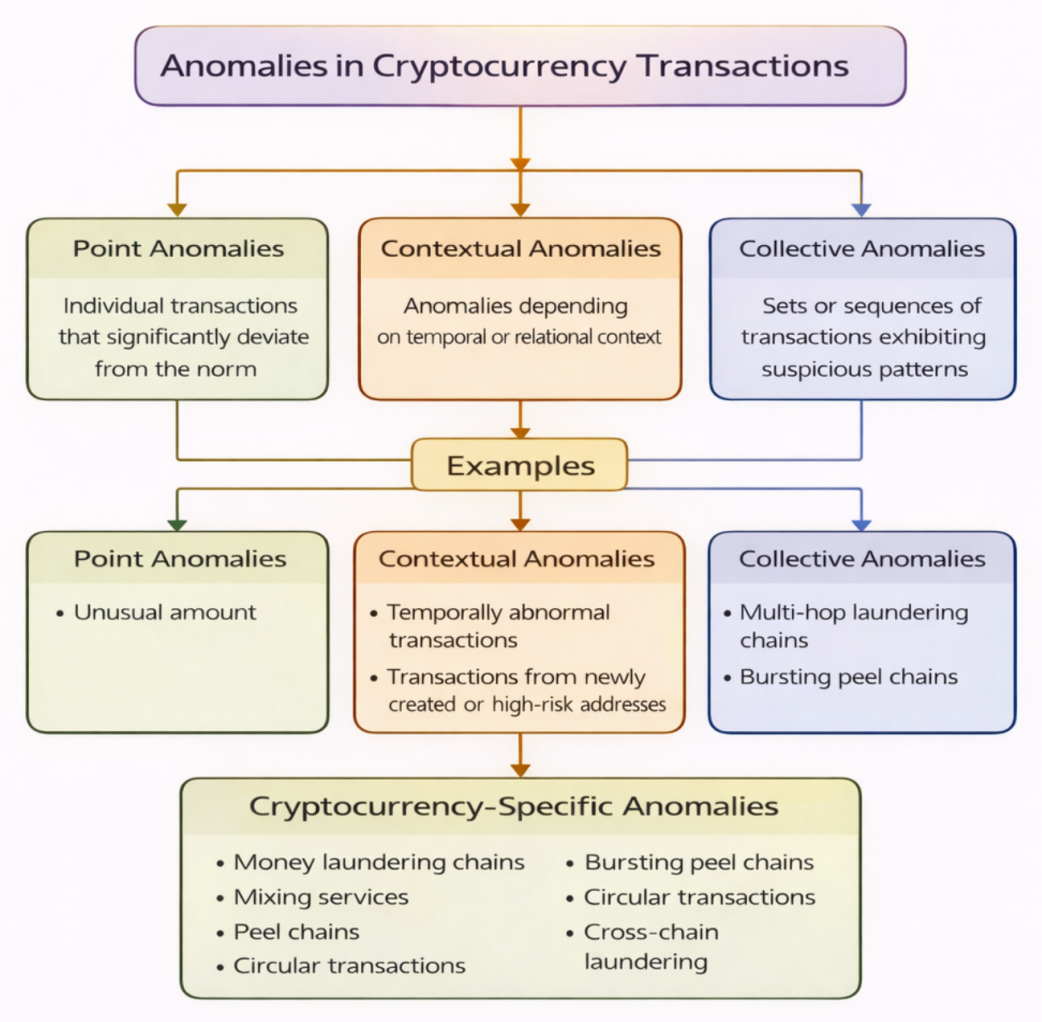


Figure 1: Conceptual taxonomy of anomaly categories in cryptocurrency transactions, illustrating point, contextual, and collective anomalies, along with cryptocurrency-specific extensions such as money laundering chains, mixing services, and cross-chain transaction patterns, adapted from foundational anomaly detection literature [8, 29].

- **RQ3:** What is the reported performance of machine learning models for cryptocurrency anomaly detection?
- **RQ4:** What is the distribution of supervised, unsupervised, and semi-supervised approaches in cryptocurrency anomaly detection?

3.2. Search Strategy

The search strategy followed a structured multi-step procedure. First, key search terms were derived directly from the research questions. Second, synonymous and related terms were incorporated, including "cryptocurrency," "blockchain," "fraud," "illicit," "anomaly," and "outlier." Third, Boolean operators (AND, OR) were applied to refine and combine search queries. Finally, the search strings were constructed to capture four main dimensions: anomaly detection, cryptocurrency systems, financial crime, and machine learning techniques.

The literature search was conducted on 31 December 2025 across the following digital libraries: IEEE Xplore, ACM Digital Library, Scopus, Web of Science, Google Scholar, and SpringerLink. The search was restricted to English-language publications. Eligible studies included peer-reviewed journal articles, conference proceedings, and high-quality preprints (e.g., arXiv) where they presented substantial empirical evaluation and were widely cited or methodologically relevant.

All search queries were adapted to the syntax and capabilities of each database to ensure comprehensive coverage.

The use of multiple complementary search strings ensures broad coverage of the research domain while reducing the risk of missing relevant studies across different subfields of cryptocurrency anomaly detection.

3.3. Study Selection Criteria

The inclusion and exclusion criteria were defined to ensure that only relevant, high-quality, and empirically validated studies were selected for analysis.

Inclusion Criteria (IC1–IC6):

These inclusion criteria ensure that only empirically validated and methodologically sound studies are considered, improving the reliability and consistency of the synthesized results.

Exclusion Criteria (EC1–EC6):

The exclusion criteria were designed to remove low-quality, non-empirical, or irrelevant studies, ensuring that the final dataset reflects rigorous and domain-specific research.

3.4. PRISMA Flow Diagram and Study Selection Process

The study selection process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [34]. The literature search was conducted on 31 December 2025 across the specified digital libraries. The screening process was performed in a structured and systematic manner, consisting of title and abstract screening followed by full-text eligibility assessment. Consistency was maintained throughout the screening process, and any disagreements were resolved through discussion and consensus. Each excluded study was assigned a single primary exclusion reason to ensure clarity and avoid overlap across categories.

Table 2: Search Strings Used in the Systematic Literature Review

Search String	**Purpose**
("cryptocurrency" OR "blockchain" OR "bitcoin" OR "ethereum") AND ("anomaly detection" OR "fraud detection")	General cryptocurrency anomaly detection
("money laundering" OR "illicit transaction") AND ("bitcoin" OR "cryptocurrency") AND ("machine learning" OR "deep learning")	Money laundering detection
("graph neural network" OR "GNN" OR "graph convolutional network") AND ("blockchain" OR "cryptocurrency") AND ("anomaly" OR "fraud")	Graph-based detection
("phishing" OR "ponzi" OR "ransomware") AND ("ethereum" OR "bitcoin" OR "cryptocurrency") AND ("detection")	Specific fraud type detection
("federated learning" OR "privacy preserving") AND ("blockchain" OR "cryptocurrency") AND ("anomaly detection")	Privacy-preserving detection

Table 3: Inclusion Criteria

Criteria	**Description**
IC1	Peer-reviewed journal articles, conference proceedings, or widely recognized preprints (e.g., arXiv) with clear empirical evaluation
IC2	Published between January 2008 and December 2025
IC3	Empirical evaluation using cryptocurrency or blockchain-related datasets
IC4	Reports standard performance metrics (e.g., accuracy, F1-score, AUC, precision, recall)
IC5	Full-text available in English
IC6	Focuses on anomaly or fraud detection in cryptocurrency or blockchain transactions

Table 4: Exclusion Criteria

Criteria	Description
EC1	Editorials, commentaries, blogs, or non-peer-reviewed opinion-based content
EC2	Industry reports or white papers without empirical validation
EC3	Purely theoretical or survey studies without experimental evaluation
EC4	Studies with insufficient methodological detail or unclear dataset usage
EC5	Duplicate publications across multiple databases
EC6	Studies not focused on cryptocurrency or blockchain transaction analysis

Figure 2 illustrates the study selection process. A total of 450 records were initially identified across multiple databases. After removing 68 duplicate records, 382 records remained for title and abstract screening. Of these, 238 records were excluded, resulting in 144 articles for full-text assessment. Following the application of the predefined inclusion and exclusion criteria, 112 articles were excluded due to reasons such as lack of empirical validation, use of non-cryptocurrency datasets, and absence of reported performance metrics. Ultimately, 32 empirical studies were included in the final qualitative synthesis.

Table 5 provides a structured summary of the study selection process, ensuring transparency in reporting and enabling reproducibility of the screening procedure.

Table 6 summarizes the primary reasons for exclusion during the full-text assessment stage, highlighting the strict filtering applied to ensure inclusion of only relevant and empirically validated studies.

The structured PRISMA-based selection process enhances transparency and reproducibility, reduces potential selection bias, and ensures a systematic and consistent filtering of relevant studies.

3.5. Quality Assessment

Following the study selection process, a structured quality assessment was conducted to ensure the methodological rigor and reliability of the included studies. This assessment followed established systematic review practices and was guided by PRISMA reporting principles [34]. Ten quality assessment criteria were defined to evaluate the methodological soundness, experimental design, and reporting quality of each study. The quality assessment was conducted independently and cross-verified to ensure consistency.

Each criterion was scored using a three-point scale: 1 (Yes), 0.5 (Partially), or 0 (No), resulting in a maximum possible score of 10 for each study. A threshold score of 5 (the midpoint of the scoring range) was selected to balance inclusion rigor with adequate coverage of relevant empirical studies. Studies with a total quality score below 5 were excluded to ensure that only methodologically sound and empirically validated studies were retained.

Dataset transparency was evaluated based on whether datasets were publicly available or sufficiently described to enable reproducibility. All 32 studies included in the final qualitative synthesis met or exceeded the defined quality threshold.

Table 7 outlines the criteria used to systematically evaluate the quality of the included studies, ensuring consistency, transparency, and methodological robustness in the assessment process.

3.6. Data Extraction Strategy

For each included study, the following information was systematically extracted: bibliographic details (authors, year, and publication venue), anomaly detection application type (RQ1), machine learning or deep learning technique employed (RQ2), dataset characteristics (dataset name and blockchain context), reported performance metrics and corresponding values (RQ3), anomaly detection paradigm (supervised, unsupervised, or semi-supervised) (RQ4), and reproducibility-related information (e.g., code availability, dataset accessibility, and experimental settings where reported).

This structured extraction process ensures consistency and enables systematic comparison across the selected studies.

3.7. Synthesis of Extracted Data

The extracted data were synthesized using a combination of narrative and quantitative approaches. Narrative synthesis was applied to address RQ1 and RQ2 by summarizing and comparing application domains and methodological trends across studies. Quantitative synthesis was conducted for RQ3 and RQ4 by analyzing the frequency and distribution of performance metrics and learning paradigms.

This combined approach enables both descriptive interpretation and structured comparison of findings, providing a comprehensive understanding of trends and patterns across the included studies.

3.8. Threats to Validity

Despite following a rigorous PRISMA-guided methodology, several limitations may affect the comprehensiveness of this review. First, the search was restricted to English-language publications, which may exclude relevant studies published in other languages. Second, although multiple digital libraries were queried, some relevant studies may not have been captured due to database coverage limitations or variations in indexing. Third, the inclusion of widely recognized preprints may introduce variability in peer-review quality. Fourth, heterogeneity in datasets, application domains, and evaluation metrics across studies may introduce overlap and limit direct comparability in quantitative summaries.

Table 5: Summary of PRISMA Study Selection Process

Stage	Description	Records
Stage 1	Initial records identified across all databases	450
Stage 2	Records after duplicate removal	382
Stage 3	Records screened by title and abstract	382
Stage 4	Records excluded (off-topic, non-cryptocurrency, survey, or theoretical studies)	238
Stage 5	Full-text articles assessed for eligibility	144
Stage 6	Full-text articles excluded based on predefined criteria	112
Stage 7	Final studies included in qualitative synthesis	32

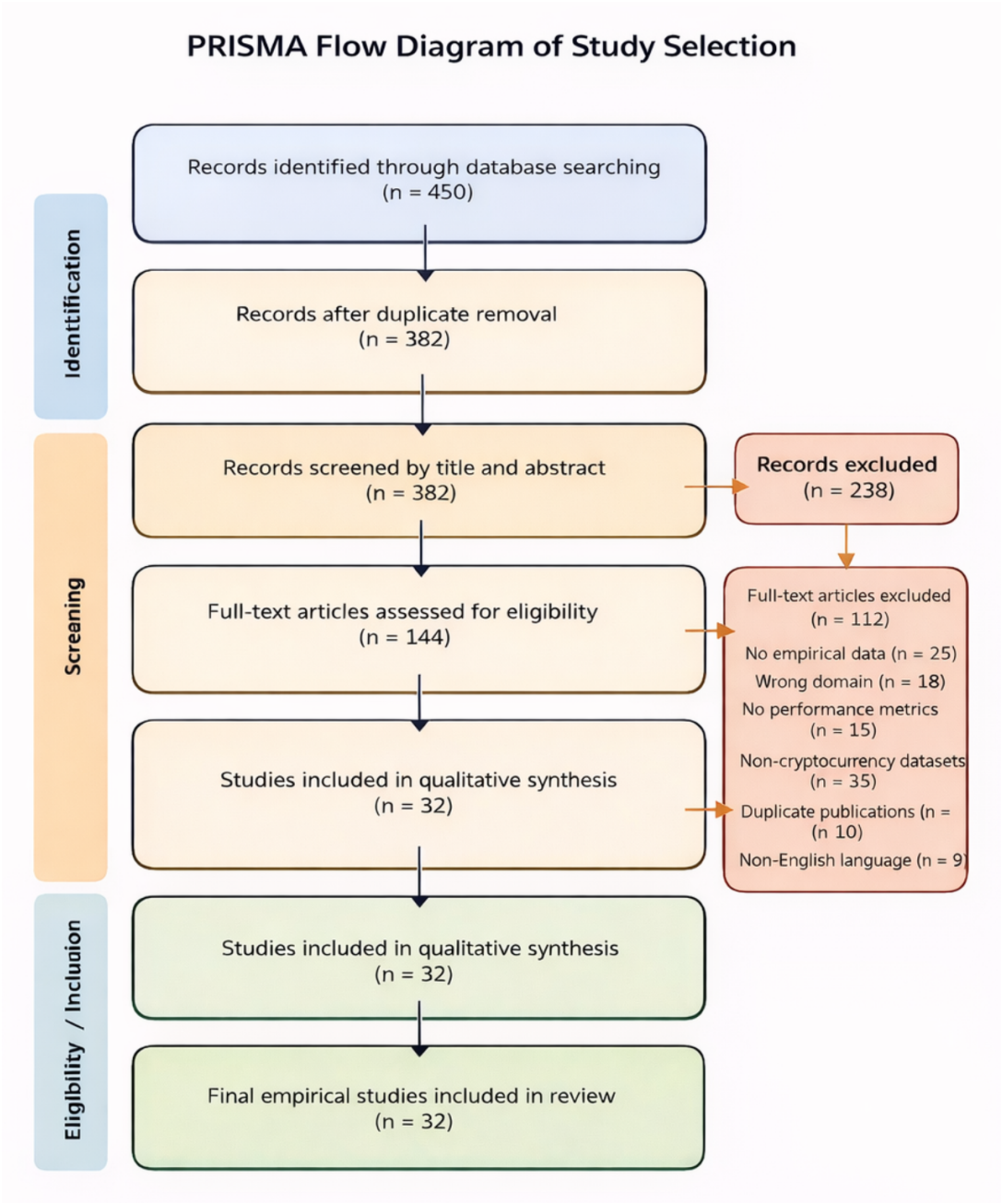


Figure 2: PRISMA flow diagram illustrating the study selection process for the systematic literature review, including identification, screening, eligibility assessment, and final inclusion of empirical studies (N = 32), adapted from PRISMA 2020 guidelines [34].

Table 6: Reasons for Exclusion at Full-Text Stage

Reason for Exclusion	Count
No empirical data (theoretical, survey, or review-only studies)	25
Wrong domain (not related to cryptocurrency or blockchain)	18
No performance metrics reported	15
Non-cryptocurrency datasets used (e.g., KDD, NSL-KDD, DARPA)	35
Duplicate publications across databases	10
Non-English language	9
Total	**112**

These limitations should be considered when interpreting the findings and highlight potential directions for improving future systematic reviews in this domain.

4. Results

In this section, the findings of the systematic review are presented. A total of 32 empirical studies applying machine learning techniques for cryptocurrency anomaly detection were selected. These studies span publications from 2016 to December 2025, although the review scope covered the period from 2008 to 2025. Earlier foundational works are referenced where relevant but are not included in the empirical study set. The results are organized primarily around the four research questions (RQ1–RQ4), along with additional synthesis subsections that support the overall interpretation of the reviewed studies.

4.1. Overview of Included Studies (RQ1)

Table 8 presents the distribution of anomaly detection applications identified in the cryptocurrency literature. The categories are not mutually exclusive, as many studies address multiple types of illicit activity. Therefore, individual studies may be counted in more than one application category, and total counts exceed the number of included studies (N = 32). This analysis addresses RQ1 by identifying the most frequently studied application domains.

Illicit transaction detection is the most frequently studied application, appearing in 59.4% of the reviewed studies. This dominance reflects the primary focus of the literature on identifying suspicious or fraudulent blockchain activity, particularly in Bitcoin transaction networks where labeled datasets such as Elliptic are available [10, 13]. However, this concentration also suggests a potential imbalance in research attention toward broader illicit activity categories, potentially limiting deeper exploration of more specific or emerging fraud types.

It is important to note that specific application areas such as money laundering, phishing, and Ponzi scheme detection represent subcategories of illicit transaction analysis and therefore overlap with the broader category of illicit detection.

Money laundering detection (34.4%) and phishing scam detection (28.1%) are also well represented, highlighting their importance in financial forensics and cybersecurity applications [23].

Ponzi scheme detection (18.8%) reflects early research on smart contract fraud in Ethereum ecosystems, where rule-based and machine learning approaches are commonly applied [21, 22].

Emerging areas such as decentralized finance (DeFi) fraud detection (9.4%) and cross-chain attack detection (6.3%) indicate a growing research focus on increasingly complex and interconnected blockchain environments. These trends suggest a shift toward addressing more sophisticated and multi-platform fraud scenarios.

In contrast, ransomware-related studies remain limited (6.3%), suggesting a potential research gap despite the increasing real-world significance of ransomware attacks.

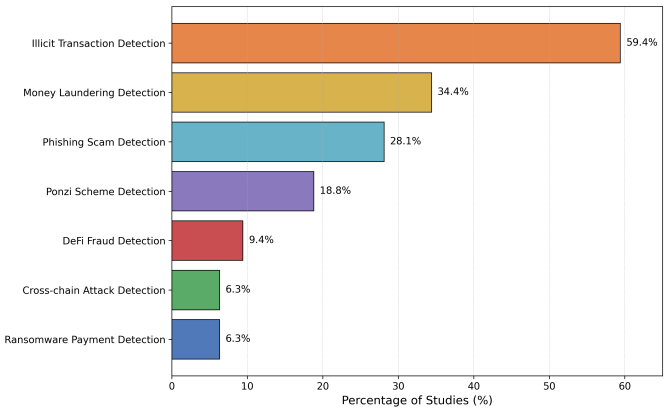


Figure 3: Distribution of anomaly detection applications in cryptocurrency research, based on the frequencies reported in Table 8.

4.2. Types of Machine Learning and Deep Learning Techniques (RQ2)

Table 9 presents the machine learning and deep learning techniques identified in the cryptocurrency anomaly detection literature. The table categorizes techniques into traditional machine learning, unsupervised or heuristic methods, graph neural network models, deep learning approaches, and advanced or emerging paradigms. This analysis addresses RQ2 by identifying the most commonly employed techniques across the reviewed studies.

As shown in Table 9, traditional machine learning techniques remain widely used, with Random Forest emerging as the most frequently applied method (25.0%). These approaches are often preferred due to their interpretability, robustness, and computational efficiency [40, 12]. However, their continued dominance may also indicate a reliance on feature-engineered approaches, potentially limiting the exploration of more complex representation learning methods.

Unsupervised approaches, including clustering and heuristic methods such as K-Means and rule-based analysis, are commonly applied in scenarios where labeled data are limited, particularly in early-stage anomaly detection studies [50, 49].

Table 7: Quality Assessment Criteria

QA#	Question
QA1	Are the study objectives clearly defined?
QA2	Are the anomaly detection techniques clearly defined and discussed?
QA3	Is the application domain of anomaly detection clearly specified?
QA4	Does the study include empirical evaluation of the proposed technique?
QA5	Are the experimental designs well-structured and justified?
QA6	Are the experiments conducted on sufficient and appropriate datasets?
QA7	Are evaluation metrics clearly reported?
QA8	Is the proposed method compared with existing approaches?
QA9	Are the methods used for result analysis appropriate?
QA10	Does the study provide meaningful contributions to academia or industry?

Table 8: Anomaly Detection Applications in Cryptocurrency Research (N = 32)

Application	No. of Studies	Percentage (%)	Representative References
Illicit Transaction Detection	19	59.4	[40, 10, 13, 15, 41, 7]
Money Laundering Detection	11	34.4	[10, 7]
Phishing Scam Detection	9	28.1	[23, 42, 16, 43]
Ponzi Scheme Detection	6	18.8	[21, 22]
DeFi Fraud Detection	3	9.4	[37, 39]
Cross-chain Attack Detection	2	6.3	[18, 24]
Ransomware Payment Detection	2	6.3	[6, 20]

Table 9: Machine Learning and Deep Learning Techniques in Cryptocurrency Anomaly Detection (N=32)

Technique Category	Specific Technique	No. of Studies	Percentage (%)	Key References
Traditional ML	Random Forest	8	25.0%	[40, 10, 12, 44, 45, 46, 47, 48]
	SVM	3	9.4%	[12, 48, 49]
	XGBoost / LightGBM	4	12.5%	[22, 45, 48, 46]
	KNN	2	6.3%	[12, 49]
	Logistic Regression	1	3.1%	[48]
	AdaBoost	1	3.1%	[12]
Unsupervised / Heuristic	K-Means / Trimmed K-Means	3	9.4%	[50, 49, 51]
	Heuristic / Analytical Methods	2	6.3%	[52, 21]
Graph Neural Networks	GCN	4	12.5%	[10, 13, 7, 53]
	GAT	3	9.4%	[41, 54, 7]
	Heterogeneous GNN	1	3.1%	[55]
	Subgraph Learning	1	3.1%	[56]
Deep Learning	Autoencoder-based	2	6.3%	[42, 57]
	MLP / Neural Networks	3	9.4%	[57, 46, 12]
Advanced / Emerging	Graph Contrastive Learning	3	9.4%	[16, 18, 43]
	Hybrid GNN-based Models	2	6.3%	[15, 39]
	Cross-chain Detection	1	3.1%	[24]

Graph-based deep learning approaches have gained significant attention, with Graph Neural Networks (GNNs), particularly GCN and GAT models, demonstrating strong capability in capturing relational structures within blockchain transaction networks [10, 13].

Recent studies increasingly explore advanced paradigms such as graph contrastive learning, hybrid GNN architectures, and cross-chain detection frameworks, reflecting a shift toward more sophisticated and scalable solutions [18, 39]. These approaches suggest a growing emphasis on representation learning and adaptability in complex blockchain environments.

Despite these advances, traditional machine learning methods remain relevant, particularly in applications where explainability and regulatory compliance are important considerations [27].

It should be noted that only the 32 selected empirical studies were considered for quantitative synthesis in this section, while additional supporting studies were used for contextual discussion.

Note: Frequencies are non-exclusive, as individual studies may employ multiple techniques across different categories. Therefore, total counts exceed the number of included studies (N=32).

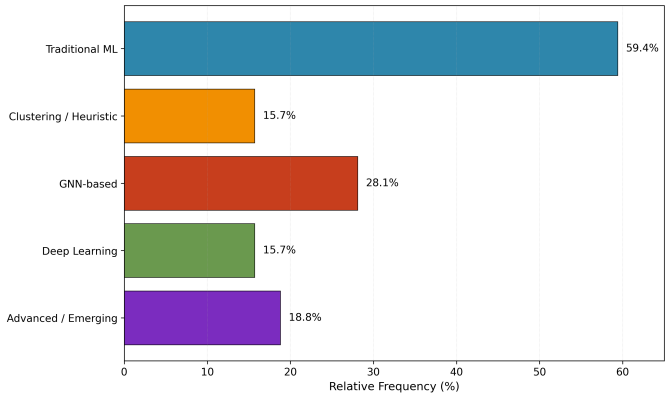


Figure 4: Distribution of machine learning and deep learning techniques in cryptocurrency anomaly detection, based on the frequencies reported in Table 9.

4.3. Datasets Used in Cryptocurrency Anomaly Detection

Table 10 summarizes the blockchain-specific datasets identified in the reviewed literature. Only datasets derived from cryptocurrency or blockchain environments are included, consistent with the defined inclusion criteria. Since some studies employ multiple datasets, the reported frequencies are non-exclusive.

As shown in Table 10, both Ethereum-based datasets and the Elliptic Bitcoin dataset represent the largest proportions (28.1%) of the reviewed studies, reflecting the strong research focus on smart contract vulnerabilities, phishing attacks, and anti-money laundering analysis.

The Elliptic Bitcoin dataset remains one of the most widely used benchmarks, supporting reproducible evaluation of anomaly detection models, particularly in financial crime detection. However, this concentration may introduce potential risks related to dataset bias and limited generalizability.

Bitcoin datasets outside Elliptic (21.9%) are also widely used, particularly in earlier studies and clustering-based anomaly detection approaches.

More recent studies increasingly adopt Elliptic++ and Elliptic2 datasets (9.4%), which provide enhanced temporal and structural representations of transaction networks.

Cross-chain and multi-platform datasets (6.3%) represent an emerging research direction, reflecting the growing importance of detecting illicit activities across interconnected blockchain ecosystems. However, their limited adoption suggests a gap between current research focus and the increasing complexity of real-world blockchain environments.

These observations motivate a more detailed analysis of dataset-related limitations, particularly in terms of bias and generalizability, which is discussed in the following subsection.

Note: Dataset usage frequencies are non-exclusive, as individual studies may utilize multiple datasets.

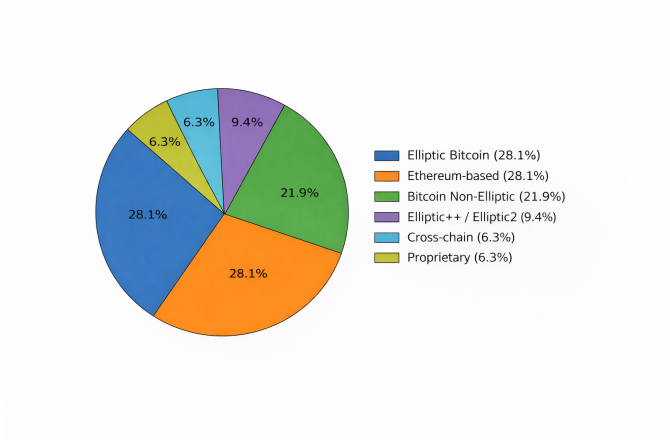


Figure 5: Dataset usage frequency in cryptocurrency anomaly detection research, based on the frequencies reported in Table 10.

4.4. Dataset Bias Analysis

Table 11 presents dataset usage statistics and associated bias risks based on the 32 empirical studies included in this review. The analysis highlights the concentration of research on specific benchmark datasets and the implications for generalizability and reproducibility.

The results indicate a strong concentration of studies on a limited number of datasets. This over-reliance on a small set of benchmark datasets may lead to overfitting to specific data distributions and reduces the external validity of proposed models. Ethereum-based datasets represent the largest proportion (28.1%), reflecting increased research focus on smart contracts, phishing attacks, and decentralized finance (DeFi) vulnerabilities.

The Elliptic Bitcoin dataset remains one of the most widely used benchmarks (25.0%), enabling reproducible evaluation across studies. However, this concentration introduces potential dataset bias, where models are repeatedly trained and tested on the same data distribution, limiting generalizability to unseen blockchain environments.

Table 10: Datasets Used in Cryptocurrency Anomaly Detection (N=32)

Dataset Category	Blockchain	No.	Percentage (%)	Key Refs.
Elliptic Bitcoin Dataset	Bitcoin	9	28.1%	[40, 10, 13, 15, 41, 7, 11, 47, 53]
Elliptic++ / Elliptic2	Bitcoin	3	9.4%	[54, 56, 46]
Ethereum-based Datasets	Ethereum	9	28.1%	[23, 42, 37, 22, 21, 16, 55, 43, 52]
Bitcoin (Non-Elliptic)	Bitcoin	7	21.9%	[48, 50, 12, 49, 51, 44, 45]
Cross-chain / Multi-platform	Various	2	6.3%	[18, 24]
Proprietary / Other Datasets	Various	2	6.3%	[57, 39]

Table 11: Dataset Usage Statistics and Bias Risk

Dataset Category	No. of Studies	% (of 32)	Blockchain	Typical Year	Bias Risk
Elliptic Bitcoin	8	25.0%	Bitcoin	2019	Medium
Elliptic++ / Elliptic2	3	9.4%	Bitcoin	2024–2025	Medium
Ethereum-based datasets	9	28.1%	Ethereum	2018–2025	Medium
Bitcoin (non-Elliptic)	7	21.9%	Bitcoin	2016–2024	Medium
Cross-chain / Multi-platform	2	6.3%	Various	2024	Medium
Proprietary / Private datasets	2	6.3%	Various	Various	High

Bitcoin datasets outside the Elliptic benchmark (21.9%) are also commonly used, particularly in earlier studies and unsupervised anomaly detection approaches. More recent datasets, such as Elliptic++ and Elliptic2 (9.4%), provide improved temporal and structural representations but remain less widely adopted.

Cross-chain and multi-platform datasets (6.3%) are underrepresented despite the growing importance of cross-chain financial crime, indicating a gap between real-world threats and current research focus. Additionally, the use of proprietary datasets (6.3%) introduces significant reproducibility challenges, as these datasets are not publicly accessible.

Overall, the findings highlight a clear need for more diverse, up-to-date, and publicly available datasets to improve the robustness, generalizability, and reproducibility of anomaly detection models in cryptocurrency ecosystems.

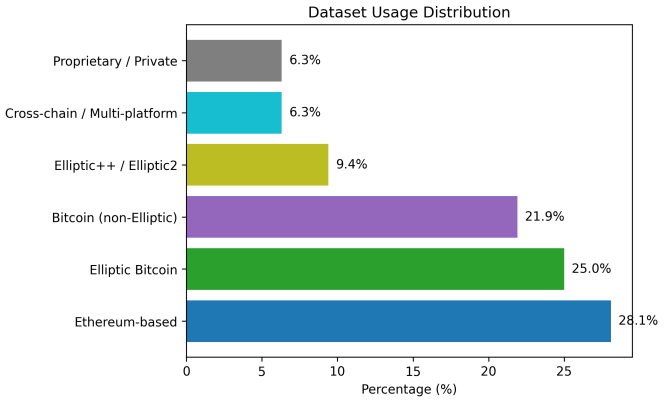


Figure 6: Dataset usage distribution in cryptocurrency anomaly detection studies, based on Table 11.

Note: Dataset usage frequencies are non-exclusive, as individual studies may utilize multiple datasets. Bias risk categories are defined as follows: Medium = publicly available datasets with potential generalization limitations; High = proprietary or non-public datasets with limited reproducibility.

4.5. Performance Metrics (RQ3)

Table 12 presents the frequency of performance metrics used across the reviewed studies. This analysis addresses RQ3 by identifying the most commonly adopted evaluation measures in cryptocurrency anomaly detection research.

Table 12: Performance Metrics Used in Cryptocurrency Anomaly Detection (N = 32)

Metric	No. of Studies	Percentage (%)
Accuracy	11	34.4
Precision	17	53.1
Recall (Sensitivity, TPR)	17	53.1
F1-score	21	65.6
AUC (ROC-AUC)	12	37.5
PR-AUC	2	6.3
FPR / Specificity	2	6.3
Computational Time	0	0.0

As shown in Table 12, F1-score is the most frequently reported evaluation metric, appearing in 65.6% of the reviewed studies. This reflects the importance of balancing precision and recall in highly imbalanced cryptocurrency transaction datasets.

Precision and recall are also widely reported (53.1% each), indicating that many studies explicitly consider the trade-off between detecting illicit transactions and minimizing false positives.

Accuracy is reported in 34.4% of studies, suggesting a reduced reliance on this metric due to its limitations in imbalanced classification problems [58]. In highly imbalanced datasets, accuracy can provide misleading performance estimates, as models

may achieve high accuracy by predominantly predicting the majority class.

AUC-based metrics are moderately used (37.5%), providing threshold-independent evaluation. However, more robust measures for imbalanced data, such as PR-AUC (6.3%), are rarely reported, indicating a gap in the adoption of more appropriate evaluation metrics.

Notably, no studies explicitly report computational efficiency metrics, such as execution time, which limits the assessment of real-world deployment feasibility. This lack of system-level evaluation highlights a gap between experimental performance and practical applicability.

It should be noted that not all studies report complete sets of standard evaluation metrics. Some studies provide only partial metrics (e.g., precision, recall, or F1-score without full numerical values), while others rely on qualitative or non-standard evaluation criteria. For consistency, only explicitly reported quantitative metrics were considered in the frequency analysis presented in Table 12. Studies without clearly reported numerical metrics were excluded from metric counting but remain included in the overall qualitative synthesis.

Note: Metric frequencies are non-exclusive, as individual studies may report multiple evaluation measures.

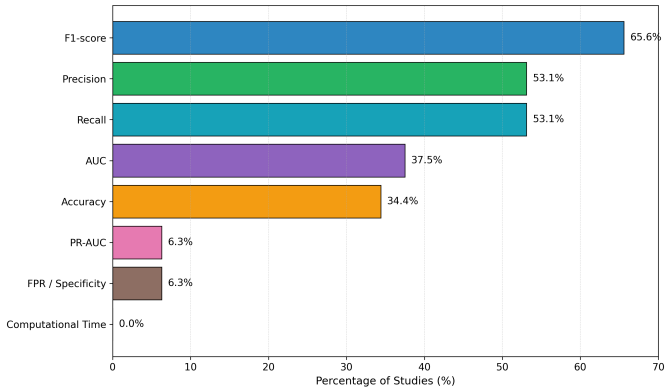


Figure 7: Frequency of performance metrics used in cryptocurrency anomaly detection studies, based on Table 12.

Table 13 summarizes the commonly used evaluation metrics and their mathematical definitions, providing a consistent reference framework for interpreting reported results across the reviewed studies.

4.6. Classification of Anomaly Detection Approaches (RQ4)

Table 14 presents the distribution of supervised, unsupervised, and semi-supervised learning approaches identified in the reviewed studies. This analysis addresses RQ4 by examining the prevalence of different learning paradigms in cryptocurrency anomaly detection.

As shown in Table 14, supervised learning is the dominant paradigm, accounting for 68.8% of the reviewed studies. This dominance is largely driven by the availability of labeled benchmark datasets, such as the Elliptic Bitcoin dataset, which enable the application of classification-based models [10, 13]. However, this reliance on labeled data may limit applicability in real-world scenarios where labeled datasets are scarce or incomplete.

Unsupervised approaches represent 21.9% of the studies and are primarily used in scenarios where labeled data is unavailable or limited, which remains a common challenge in cryptocurrency anomaly detection [50].

Semi-supervised methods account for 9.4% of the studies and aim to leverage both labeled and unlabeled data to improve detection performance [57]. Despite their potential advantages, these approaches remain comparatively less explored.

Overall, the distribution highlights a strong reliance on supervised learning, while indicating opportunities for further research in alternative paradigms that address label scarcity and scalability challenges. For consistency in quantitative analysis, each study was assigned to a single primary learning paradigm based on its dominant training strategy. Studies employing hybrid or multi-stage approaches were categorized according to their primary learning mechanism to avoid double counting and ensure comparability across studies. Emerging variants such as self-supervised learning were grouped under the closest broader category.

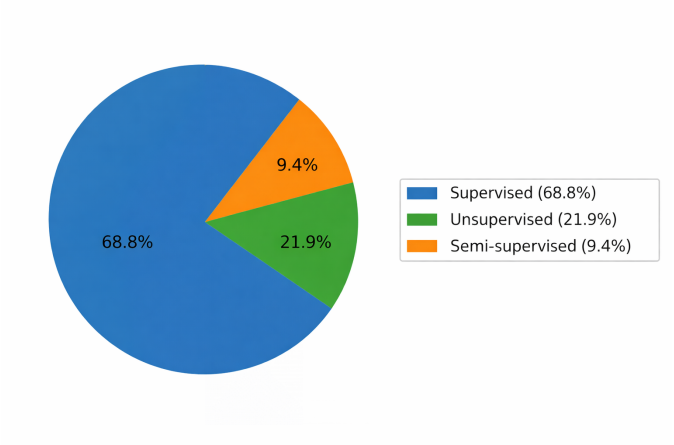


Figure 8: Distribution of anomaly detection learning paradigms in cryptocurrency research, based on Table 14.

4.7. Temporal Evolution of Detection Techniques

Table 15 presents a conceptual overview of the methodological evolution of anomaly detection techniques in cryptocurrency research. This synthesis is based on the temporal distribution of the included studies and broader literature context, and is not intended as a direct quantitative summary of the 32 empirical studies.

The evolution of detection techniques reflects a general progression from traditional machine learning approaches toward graph-based and deep learning models. Early studies primarily relied on feature engineering and classical classifiers, whereas more recent work increasingly incorporates graph structures and temporal dynamics to capture complex transaction patterns.

The most recent period (2023–2025) demonstrates growing methodological diversity, including hybrid architectures, self-supervised learning, and cross-chain analysis. This trend highlights the increasing complexity of cryptocurrency ecosystems

Table 13: Common Performance Metrics and Their Definitions

Metric	Formula	Description
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$	Overall classification correctness
Precision	$TP/(TP + FP)$	Proportion of predicted anomalies that are correct
Recall (TPR)	$TP/(TP + FN)$	Proportion of actual anomalies correctly detected
F1-score	$2 \times \frac{P \times R}{P+R}$	Harmonic mean of precision and recall
FPR	$FP/(FP + TN)$	Normal instances misclassified as anomalies
AUC	Area under ROC curve	Threshold-independent performance measure

Table 14: Anomaly Detection Classification Type (N=32)

Type	Number of Studies	Percentage (%)
Supervised	22	68.8%
Unsupervised	7	21.9%
Semi-supervised	3	9.4%

and the need for scalable and adaptive detection frameworks. It also suggests a shift toward models capable of handling dynamic, large-scale, and heterogeneous blockchain environments.

Note: This table provides a conceptual synthesis of methodological trends and is not derived from a direct quantitative breakdown of the 32 included studies.

4.8. Comparison of Deep Learning Techniques

Table 16 provides a comparative overview of deep learning techniques applied in cryptocurrency anomaly detection, highlighting their strengths, limitations, and typical application contexts. The comparison is based on the techniques identified in the 32 empirical studies included in this review and is supplemented by interpretive analysis.

Table 16 highlights the trade-offs between different deep learning techniques, indicating that model selection should be guided by application requirements such as interpretability, scalability, and temporal modeling capability.

The comparison indicates that no single deep learning technique consistently outperforms others across all scenarios within the reviewed studies. Graph-based models such as GCN and GAT are particularly effective for representing transaction networks, while temporal GNN models extend this capability by capturing time-dependent behavioral patterns. Autoencoder-based approaches are commonly applied in unsupervised settings, especially when labeled data are limited.

More recent studies increasingly explore advanced paradigms such as graph contrastive learning and hybrid GNN architectures, reflecting a shift toward more sophisticated and scalable detection techniques. These approaches improve representation learning and detection performance but often introduce additional computational complexity.

Overall, the findings suggest that model selection should be guided by application-specific requirements, including data availability, temporal dynamics, interpretability, and scalability considerations. The implications of these trade-offs are further discussed in Section 5.

Note: This comparison is derived from the deep learning techniques identified in the selected empirical studies and includes interpretive synthesis of their characteristics.

4.9. Resampling Techniques for Class Imbalance

Table 17 presents a comparison of different resampling techniques applied to the Elliptic Bitcoin dataset. This analysis provides insights into the effectiveness of imbalance handling strategies in improving anomaly detection performance.

As shown in Table 17, the ENN-all resampling technique demonstrates the best reported performance, achieving an F1-score of 93.93%. This improvement is primarily driven by a substantial increase in recall, indicating enhanced capability in identifying illicit transactions [40].

In contrast, commonly used oversampling techniques such as SMOTE and its variants show comparatively lower performance in this study. These findings suggest that noise reduction and instance cleaning approaches, such as ENN-based methods, can be more effective than synthetic data generation for handling class imbalance in blockchain transaction datasets [40].

It should be noted that this analysis is derived from a single study and is presented as an illustrative example rather than a generalized conclusion across all reviewed works. These findings have practical implications for real-world deployment, where maximizing recall is critical to ensure that illicit activities are not missed, even at the cost of minor reductions in precision.

4.10. Complete List of Included Empirical Studies

Table 18 presents a comprehensive overview of the 32 empirical studies included in this systematic review. The table summarizes key characteristics of each study, including the techniques applied, datasets used, reported performance metrics, and learning paradigms. It serves as the primary reference for the empirical evidence analyzed throughout the Results section.

Table 15: Temporal Evolution of Detection Techniques

Time Period	Representative Techniques	Key Characteristics
2016–2018	Traditional ML, Clustering	Early studies focused on classical machine learning methods (e.g., SVM, Random Forest) and unsupervised clustering approaches, with heavy reliance on handcrafted features
2018–2020	Graph-based Methods	Introduction of graph representations of blockchain transactions, enabling relational learning and the emergence of early Graph Convolutional Networks (GCNs)
2020–2023	Deep Learning and GNNs	Increased adoption of graph neural networks (GCN, GAT), hybrid deep learning models, and improved feature learning from transaction networks
2023–2025	Temporal, Hybrid, and Advanced Models	Emergence of temporal GNNs, graph contrastive learning, self-supervised learning, and cross-chain detection frameworks addressing dynamic and multi-platform environments

Table 16: Comparison of Deep Learning Techniques for Cryptocurrency Anomaly Detection

Technique	Strengths	Limitations	Best Use Case	Key References
GCN	Captures structural relationships in transaction graphs; efficient for node classification	Limited ability to model dynamic graphs; risk of oversmoothing	Static transaction graph analysis and illicit node classification	[10, 13, 7]
GAT / Multi-hop GAT	Attention mechanism improves neighbor importance modeling; better interpretability	Higher computational cost; scalability challenges	Identifying influential nodes and multi-hop transaction dependencies	[41, 54]
Temporal GNN (e.g., MDST-GNN, ATGAT)	Captures temporal evolution of transaction networks; models dynamic behavior	Complex architecture and training process	Time-evolving fraud detection and dynamic blockchain analysis	[15, 54]
Autoencoder-based Models	Effective for unsupervised anomaly detection; learns latent representations	Sensitive to threshold selection; may miss subtle anomalies	Detection of unknown or previously unseen anomalies	[42, 57]
MLP / Neural Networks	Flexible modeling capability; effective with engineered features	Limited interpretability; requires sufficient labeled data	Supervised classification tasks with structured features	[57, 46]
Graph Contrastive Learning	Learns robust representations without extensive labels	Requires careful design of contrastive objectives	Semi-supervised and representation learning tasks	[16, 18, 43]
Hybrid GNN-based Models	Combines structural and temporal features; improved performance	High computational complexity and model design overhead	Complex fraud detection scenarios and multi-feature environments	[39, 15]

Table 17: Performance Comparison of Resampling Techniques on the Elliptic Bitcoin Dataset (adapted from [40])

Resampling Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
No SMOTE (Baseline)	98.02	97.96	71.09	82.39
SMOTE	97.57	88.87	71.56	79.28
K-means SMOTE	98.02	97.96	71.09	82.39
Borderline-SMOTE1	97.25	84.00	71.28	77.12
ENN	98.01	99.34	69.89	82.05
ENN-all	99.42	99.31	89.10	93.93

Table 18: Complete List of Included Empirical Studies (N=32)

Ref	Authors	Year	Technique	Dataset	Key Performance	Type
[40]	Alarab & Prakoonwit	2022	Random Forest + Resampling (ENN-all)	Elliptic Bitcoin dataset	Accuracy: 99.42%, Precision: 99.31%, Recall: 89.10%, F1: 93.93%	Supervised
[10]	Weber et al.	2019	Random Forest, GCN	Elliptic Bitcoin dataset	Precision: 0.956, Recall: 0.670, F1: 0.788	Supervised
[13]	Asiri & Somasundaram	2025	GCN	Elliptic Bitcoin dataset	Accuracy: 98.56%, AUC: 0.9444, F1: 92.87%	Supervised
[15]	Chen et al.	2025	MDST-GNN	Elliptic Bitcoin dataset	AUC-ROC: 0.968, AUC-PR: 0.823, F1: 0.762	Supervised
[41]	Gu et al.	2025	Multi-hop GAT (OKD-GAT)	Elliptic Bitcoin dataset	F1: 0.909	Supervised
[48]	Lin et al.	2019	Supervised ML (RF, SVM, XGBoost, LightGBM)	Bitcoin transaction dataset	F1-score: 0.87	Supervised
[54]	Zhang et al.	2025	Temporal GAT (ATGAT)	Elliptic++ dataset	Accuracy: 0.9715, AUC: 0.9130, F1: 0.8427	Supervised
[56]	Bellei et al.	2024	Subgraph learning (GLASS benchmark)	Elliptic2 dataset	F1: 0.933, AUC: 0.889	Supervised
[53]	Lawal et al.	2025	Explainable GNN (GCN + GNNExplainer)	Elliptic Bitcoin dataset	AUC $\approx$ 0.5	Supervised
[57]	Son et al.	2024	Semi-supervised deep learning (DAE-MLP: Autoencoder + MLP)	Ethereum-based blockchain supply chain dataset	Accuracy: 96.5%, Precision, Recall, F1-score	Semi-supervised
[42]	Kang & Buu	2024	Graph autoencoder (DP-GCAE)	Ethereum	F1-score (+37.7% improvement over baselines)	Unsupervised
[52]	Victor	2020	Heuristic-based address clustering	Ethereum	Clustering-based evaluation (no standard quantitative metrics reported)	Unsupervised
[39]	Badjie et al.	2025	GAT+GPT-based hybrid (MGGPT)	Cryptocurrency transaction datasets	Precision, Recall, F1-score, AUC	Supervised
[18]	Che et al.	2024	Graph contrastive learning (ShadowEyes)	Cross-platform (BTC, ETH)	F1-score $\approx$ 90% (cross-platform)	Semi-supervised
[37]	Xia et al.	2021	Hybrid detection (guilt-by-association + ML)	Uniswap (Ethereum)	F1-score 96.62% (Precision 96.45%, Recall 96.79%)	Supervised
[23]	Chen et al.	2020	Graph-based deep learning (GCN + Autoencoder + LightGBM)	Ethereum transaction network	F1-score, Precision, Recall, AUC	Supervised
[21]	Bartoletti et al.	2019	Smart contract analysis (rule-based + empirical study)	Ethereum Ponzi schemes dataset	Transaction patterns, user gains/losses, economic impact	Unsupervised
[22]	Chen et al.	2018	Machine learning (XGBoost with account + opcode features)	Ethereum smart contracts	Precision, Recall, F1-score	Supervised
[50]	Monamo et al.	2016	Trimmed k-means clustering (robust anomaly detection)	Bitcoin transaction network	Detection of known fraud cases (unsupervised evaluation)	Unsupervised
[46]	Elmougy & Liu	2023	Graph-based machine learning (RF, MLP, LSTM, XGBoost)	Bitcoin (Elliptic++)	Precision, Recall, F1-score	Supervised
[24]	Augusto et al.	2024	Logic-based anomaly detection (XChainWatcher framework)	Cross-chain bridges (Ethereum, Ronin, Nomad)	No standard quantitative metrics reported	Unsupervised
[7]	Pocher et al.	2023	Graph-based machine learning (GCN, GAT) for AML forensics	Elliptic Bitcoin transaction dataset	Illicit transaction classification (F1-score, precision, recall)	Supervised
[11]	Lorenz et al.	2020	Active learning with supervised and unsupervised ML	Elliptic Bitcoin transaction dataset	F1-score reported	Semi-supervised

Continued on next page

Table 18: Complete List of Included Empirical Studies (N=32) (continued)

Ref	Authors	Year	Technique	Dataset	Key Performance	Type
[12]	Chen et al.	2021	KNN, SVM, RF, AdaBoost, MLP (with SMOTE)	Bitcoin transaction graph dataset	Precision, Recall, F1-score (RF ≈ 95.9%)	Supervised
[16]	Li et al.	2023	Transaction Graph Contrast Network (TGC)	Ethereum	Precision, Recall, F1-score	Supervised
[49]	Pham & Lee	2016	k-means, Mahalanobis distance, One-class SVM	Bitcoin	Anomaly detection performance (custom metrics)	Unsupervised
[51]	Shayegan et al.	2021	Collective anomaly detection + Trimmed k-means	Bitcoin	Anomaly detection performance (fraud cases detected)	Unsupervised
[44]	Hasan et al.	2024	Tree-based ensemble ML + SHAP explainability	Bitcoin	Accuracy, TPR, FPR, ROC-AUC	Supervised
[55]	Kanezashi et al.	2022	Heterogeneous GNN (RGCN, HAN, HGT)	Ethereum transaction network	Precision, Recall, F1-score, PR-AUC	Supervised
[47]	Gomes & Ferreira	2025	ML + dimensionality reduction (PCA, UMAP, ANOVA)	Elliptic Bitcoin dataset	Accuracy, Precision, Recall, F1-score, AUC	Supervised
[43]	Sui et al.	2025	Self-supervised GNN (EPAD)	Ethereum datasets (D1, D2)	Accuracy, Precision, Recall, F1-score, ROC-AUC	Semi-supervised
[45]	Ashfaq et al.	2022	Random Forest, XGBoost (ML + blockchain framework)	Bitcoin (Kaggle)	Precision, AUC, Accuracy (ROC, confusion matrix)	Supervised

Note: All studies listed in Table 18 satisfy the defined inclusion criteria (IC1–IC6) and represent empirical cryptocurrency anomaly detection research.

Table 18 highlights the diversity of approaches, datasets, and evaluation strategies employed in cryptocurrency anomaly detection research. Although the review scope covered the period from 2008 to 2025, no included empirical studies meeting the final selection criteria were identified prior to 2016, reflecting the relatively recent emergence and rapid growth of this research domain.

From a methodological perspective, the reviewed studies indicate a general shift from traditional machine learning techniques toward more advanced deep learning approaches, particularly graph-based and hybrid architectures. The growing adoption of Graph Neural Networks (GNNs) reflects the importance of capturing structural relationships within blockchain transaction networks.

The dataset distribution shows a noticeable reliance on benchmark datasets such as the Elliptic Bitcoin dataset, which facilitates comparative evaluation across studies but may limit generalizability to diverse and evolving blockchain environments.

In terms of evaluation practices, the table reveals notable heterogeneity in reported performance metrics. While some studies report comprehensive measures such as accuracy, precision, recall, and F1-score, others report only partial metrics (e.g., precision, recall, or F1-score without complete value disclosure) or rely on non-standard evaluation criteria. This lack of consistency complicates direct comparison across studies and highlights the need for standardized evaluation frameworks.

Regarding learning paradigms, supervised approaches dominate the literature, as reflected in Table 14. However, unsupervised, semi-supervised, and emerging approaches are also represented, indicating growing research interest in addressing challenges such as label scarcity, privacy constraints, and the dynamic nature of fraudulent activities.

5. Discussion

5.1. Synthesis of Key Findings

This systematic review synthesizes findings from 32 empirical studies on cryptocurrency anomaly detection and reveals several important trends shaping the field.

First, the literature indicates a shift from traditional feature-based machine learning classifiers (e.g., Random Forest, Support Vector Machines, and XGBoost) toward graph-based and temporally-aware deep learning architectures. This trend reflects the increasing complexity of blockchain transaction data and the need to capture relational and dynamic patterns in fraudulent behavior.

Second, Graph Neural Networks (GNNs) have become a prominent approach in the reviewed studies [10, 13] due to their ability to exploit the inherent graph structure of blockchain data. Models such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) enable the modeling of relationships between entities, making them well-suited for transaction network analysis. However, based on the reviewed studies, their effectiveness varies depending on dataset characteristics and experimental settings rather than consistently outperforming traditional methods.

Third, temporal modeling has become an increasingly important component in modern detection systems. Techniques incorporating temporal dependencies, including temporal graph models and hybrid architectures, demonstrate improved capability in capturing evolving transaction behaviors, such as multi-step laundering activities and coordinated fraud patterns.

Fourth, deep learning approaches, including autoencoders, GNNs, and hybrid models, provide advantages in automated feature learning and modeling of complex structures. However, these approaches often introduce higher computational costs and reduced interpretability, which may limit their applicability in real-world regulatory and financial environments.

Fifth, the analysis reveals a notable reliance on specific benchmark datasets, particularly the Elliptic Bitcoin dataset, which is used in approximately 28.1% of the reviewed studies. While this facilitates benchmarking and comparability, it also raises concerns regarding dataset bias and limited generalizability to newer and more diverse blockchain environments.

Sixth, evaluation practices across the literature remain inconsistent. The findings show that F1-score is the most commonly reported metric, reflecting the importance of balancing precision and recall in highly imbalanced datasets. However, other important metrics such as AUC and PR-AUC are less frequently used, and the variation in reported metrics complicates direct comparison across studies. This indicates the need for more standardized evaluation frameworks in cryptocurrency anomaly detection research.

Finally, emerging paradigms such as privacy-preserving learning, cross-chain detection frameworks, and explainable artificial intelligence (XAI) represent important future directions. These approaches aim to address practical challenges including data privacy, interoperability across blockchain platforms, and the need for transparent and interpretable decision-making systems.

5.2. Key Challenges

Despite significant progress, several critical challenges remain in the field of cryptocurrency anomaly detection.

Challenge 1: Dataset Availability and Bias. The literature exhibits a strong dependence on a limited number of benchmark datasets, particularly the Elliptic Bitcoin dataset, which is used in approximately 28.1% of the reviewed studies [10]. Only a small proportion of studies utilize multi-chain or cross-chain data, despite the growing importance of such environments [18]. Additionally, the use of proprietary datasets limits reproducibility and independent validation. This underscores the need for standardized, publicly available, and diverse datasets with temporal annotations and reliable ground truth to improve generalizability and benchmarking.

Challenge 2: Evaluation Metric Inconsistency. The findings reveal inconsistency in the selection of evaluation metrics across studies. While F1-score is the most frequently reported metric, reflecting the importance of balancing precision and

recall in imbalanced datasets, other robust measures such as PR-AUC and MCC are rarely used [59, 58]. Accuracy is reported in a relatively small proportion of studies and is often insufficient for evaluating highly imbalanced data. This lack of standardization complicates cross-study comparison and limits the reliability of reported performance. Future studies should adopt more consistent evaluation protocols incorporating F1-score, PR-AUC, MCC, and computational efficiency metrics.

Challenge 3: Adversarial Adaptation Gap. Cryptocurrency fraud detection operates in an adversarial environment where malicious actors continuously adapt their strategies [60, 61]. However, most studies are evaluated on static datasets and do not explicitly account for evolving adversarial behavior. This limits the robustness of proposed models in real-world settings. Incorporating adversarial simulation and robustness testing into evaluation pipelines is therefore essential.

Challenge 4: Explainability Requirements. Regulatory frameworks increasingly require explainable and auditable decision-making processes, particularly in financial applications [26, 62]. However, many advanced models, especially deep learning and graph-based approaches, lack interpretability [27]. This creates a trade-off between performance and transparency, highlighting the importance of integrating explainability techniques such as post-hoc explanation methods and interpretable model design [28].

Challenge 5: Scalability and Deployment Constraints. Although deep learning models demonstrate strong capability in modeling complex blockchain data, their computational complexity presents challenges for real-time deployment [39, 17]. Many approaches are evaluated in offline settings and may not scale effectively to high-throughput blockchain systems. Developing efficient architectures and leveraging optimization techniques such as model compression and hardware acceleration are essential for practical deployment.

Challenge 6: Temporal Generalization. Many studies rely on random train-test splits, which may overestimate model performance in real-world scenarios [63]. When evaluated on temporally separated data, performance can degrade due to evolving transaction patterns and adversarial adaptation [15]. Adopting temporal validation strategies and continual learning approaches is therefore important for improving model robustness and real-world applicability [64, 65].

5.3. Future Research Directions

Based on the limitations and research gaps identified from the 32 empirical studies, several directions for future work can be outlined across short-, medium-, and long-term horizons.

In the short term, a key priority is the development of standardized and diverse benchmark datasets. The results show a strong concentration on a limited number of datasets, particularly the Elliptic Bitcoin dataset, while cross-chain and multi-platform datasets remain underrepresented. Future datasets should therefore incorporate multiple blockchain platforms (e.g., Bitcoin, Ethereum, and cross-chain environments), include temporal information, and provide reliable ground truth annotations. In addition, there is a clear need for standardized evaluation protocols. The findings reveal inconsistency in metric selection, with limited use of robust measures such as PR-AUC and MCC. Future studies should adopt consistent reporting standards, including F1-score, PR-AUC, MCC, and system-level metrics such as computational efficiency. Furthermore, integrating explainability into detection models remains an important direction, particularly for graph-based approaches where interpretability is currently limited.

In the medium term, research should focus on improving model robustness and adaptability. The current literature largely evaluates models on static datasets, which does not fully reflect the adversarial and evolving nature of cryptocurrency fraud. Future work should incorporate adversarial evaluation frameworks to simulate adaptive attacker behavior and assess model resilience. Additionally, continual learning and temporal adaptation mechanisms are required to address concept drift and evolving transaction patterns. The scalability of advanced approaches, particularly graph neural networks and privacy-preserving methods such as federated learning, should also be investigated in large-scale and heterogeneous environments.

In the long term, research is expected to move toward unified and scalable detection frameworks capable of operating across multiple blockchain ecosystems. This includes cross-chain analysis, integration of transaction data from different platforms, and monitoring of complex financial interactions such as bridge transactions and decentralized finance (DeFi) activities. The development of regulatory-compliant AI systems that ensure explainability, transparency, and auditability will be essential for real-world adoption. Finally, designing real-time detection systems capable of processing high-throughput blockchain data streams with low latency remains a fundamental challenge and an important direction for future research.

6. Conclusion

6.1. Summary of Main Findings

This systematic review examined 32 empirical studies on cryptocurrency anomaly detection published between 2016 and December 2025, within a broader review scope covering 2008 to 2025. The findings indicate a shift from traditional machine learning approaches, such as Support Vector Machines and Random Forest, toward graph-based and deep learning architectures, including Graph Convolutional Networks, Graph Attention Networks, and hybrid models. These approaches leverage the inherent graph structure of blockchain data to capture relational dependencies and complex transaction patterns.

Deep learning techniques, including autoencoders, graph neural networks, and hybrid architectures, enhance detection capabilities by enabling automated feature learning and modeling of complex transaction networks. However, these approaches also introduce challenges related to computational complexity and limited interpretability, which may affect their practical deployment.

The analysis also highlights a concentration of studies on benchmark datasets, particularly the Elliptic Bitcoin dataset, which is used in approximately 28.1% of the reviewed studies. While this supports benchmarking and comparability, it raises

concerns regarding dataset bias and limited generalizability to evolving fraud patterns.

Furthermore, evaluation practices remain inconsistent across the literature. The findings show that F1-score is the most frequently reported metric, reflecting the importance of balancing precision and recall in highly imbalanced datasets. In contrast, other important metrics such as PR-AUC and MCC are relatively underutilized, and the lack of standardized evaluation practices complicates cross-study comparison.

Emerging paradigms, including privacy-preserving learning, cross-chain detection frameworks, and explainable artificial intelligence, represent promising directions for addressing real-world deployment challenges. Collectively, these findings provide a comprehensive overview of the current state of cryptocurrency anomaly detection, highlighting both methodological advances and remaining limitations.

6.2. Implications for Research and Practice

The findings of this review have several implications for researchers, industry practitioners, and policymakers.

From a research perspective, there is a clear need for more diverse and standardized benchmark datasets incorporating multi-chain environments and temporal information. The adoption of consistent evaluation protocols, including metrics suitable for imbalanced data and system-level performance measures, would improve the comparability and reliability of future studies. Additionally, integrating explainability into model design and ensuring reproducibility through data and code availability remain important priorities.

From an industry perspective, graph-based and hybrid detection models offer strong potential for capturing complex transaction patterns in real-world blockchain systems. However, practical deployment requires balancing performance with interpretability, scalability, and computational efficiency. Continuous model updating and monitoring are also necessary to address evolving fraud behaviors.

From a policy perspective, the development of regulatory frameworks supporting explainable and auditable artificial intelligence systems is critical. Encouraging privacy-preserving data sharing and collaboration across institutions can facilitate more robust and scalable detection systems. Public–private partnerships may further accelerate the development of trustworthy solutions in this domain.

6.3. Limitations of This Review

This systematic review has several limitations. First, the analysis is restricted to English-language studies, primarily peer-reviewed publications together with a limited number of widely recognized empirical preprints, which may exclude relevant studies published in other languages or within broader grey literature. Second, the reliance on studies using proprietary datasets limits reproducibility and cross-study validation. Third, inconsistencies in reporting evaluation metrics across studies complicate direct comparison of results. Fourth, foundational works prior to 2008 were not systematically included, as the focus of this review is on cryptocurrency-related research. Finally, given the rapid pace of development in this field, newly emerging techniques beyond December 2025 are not captured in this review.

Declaration of Generative AI and AI-assisted Technologies in the Manuscript Preparation Process

During the preparation of this manuscript, the authors used ChatGPT *solely* as a language-support tool to assist with grammar checking, sentence restructuring, and clarity improvement. After using this tool, the authors carefully reviewed, edited, and validated all content to ensure technical accuracy and originality. The authors take full responsibility for the content of the published article.

CRediT Authorship Contribution Statement

Md Saidul Islam: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft.

Syed Mohammed Shamsul Islam: Supervision, Formal analysis, Investigation, Validation, Writing – review & editing.

Md Zakir Hossain: Supervision, Formal analysis, Investigation, Validation, Writing – review & editing.

Mohiuddin Ahmed: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

Iqbal H. Sarker: Supervision, Writing – review & editing.

Data Availability

No new experimental dataset was generated for this study. All data analyzed in this review are drawn from previously published studies cited in the manuscript. Any extracted review tables or synthesis files are available from the corresponding author upon reasonable request. The search strategy, inclusion criteria, and extracted data summaries are also available upon request to support reproducibility of the review process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM Computing Surveys 41 (3) (2009) 71–97. doi:10.1145/1541880.1541882.

[2] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, K.-R. Muller, A unifying review of deep and shallow anomaly detection, Proceedings of the IEEE 109 (5) (2021) 756–795. doi:10.1109/JPROC.2021.3052449.

[3] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, white Paper (2008).
URL https://www.researchgate.net/publication/228640975_Bitcoin_A_Peer-to_Peer_Electronic_Cash_System

[4] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, An overview of blockchain technology: Architecture, consensus, and future trends, in: BigData Congress 2017, 2017, pp. 557–564. doi:10.1109/BigDataCongress.2017.85.

[5] Chainalysis, The chainalysis 2025 crypto crime report, Technical report, Chainalysis Inc. (2025).
URL https://www.chainalysis.com/blog/2025-crypto-crime-report-introduction/

[6] M. Möser, R. Böhme, D. Breuker, Towards risk scoring of bitcoin transactions, in: Financial Cryptography and Data Security (FC 2014), 2014, pp. 16–32. doi:10.1007/978-3-662-44774-1_2.

[7] N. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, S. Ferretti, Detecting anomalous cryptocurrency transactions: An aml/cft application of machine learning-based forensics, Electronic Markets 33 (1) (2023) 37. doi:10.1007/s12525-023-00654-3.

[8] L. Akoglu, H. Tong, D. Koutra, Graph based anomaly detection and description: A survey, Data Mining and Knowledge Discovery 29 (3) (2015) 626–688. doi:10.1007/s10618-014-0365-y.

[9] A. B. Nassif, M. A. Talib, Q. Nasir, F. M. Dakalbab, Machine learning for anomaly detection: A systematic review, IEEE Access 9 (2021) 78658–78700. doi:10.1109/ACCESS.2021.3083060.

[10] M. Weber, G. Domeniconi, J. Chen, D. Weidele, C. Bellei, T. Robinson, C. Leiserson, Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics (2019). arXiv:1908.02591.
URL https://arxiv.org/abs/1908.02591

[11] J. Lorenz, et al., Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity, Proceedings of the ACM Conference on AI in Finance (2020). doi:10.1145/3383455.3422549.

[12] Z. Chen, J. Li, Y. Wu, Bitcoin theft detection based on supervised machine learning algorithms, Security and Communication Networks 2021 (2021) 1–12. doi:10.1155/2021/6643763.

[13] A. Asiri, K. Somasundaram, Graph convolution network for fraud detection in bitcoin transactions, Scientific Reports 15 (1) (2025) 11076. doi:10.1038/s41598-025-95672-w.

[14] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: International Conference on Learning Representations (ICLR), 2018.
URL https://doi.org/10.48550/arXiv.1710.10903

[15] S. Chen, Y. Liu, Q. Zhang, Z. Shao, Z. Wang, Multi-distance spatial-temporal graph neural network for anomaly detection in blockchain transactions, Advanced Intelligent Systems (2025). doi:10.1002/aisy.202400898.

[16] Y. Li, X. Zhang, H. Wang, Tgc: Transaction graph contrastive learning for ethereum phishing scam detection, in: ACM International Conference Proceedings, 2023. doi:10.1145/3627106.3627109.

[17] B. Luo, Z. Zhang, Q. Wang, A. Ke, Ai-powered fraud detection in decentralized finance: A project life cycle perspective (2023). arXiv:2308.15992.
URL https://arxiv.org/abs/2308.15992

[18] Z. Che, M. Shen, Z. Tan, H. Du, L. Zhu, W. Wang, T. Chen, Q. Zhao, Y. Xie, Across-platform detection of malicious cryptocurrency transactions via account interaction learning, IEEE Transactions on Information Forensics and Security 20 (2024) 4783–4798. doi:10.1109/TIFS.2025.3565130.

[19] A. Abdallah, M. A. Maarof, A. Zainal, Fraud detection system: A survey, Journal of Network and Computer Applications 68 (2016) 90–113. doi:10.1016/j.jnca.2016.04.007.

[20] İ. Kara, M. Aydos, The rise of ransomware: Forensic analysis for windows-based ransomware attacks, Expert Systems with Applications 190 (2021) 116198. doi:10.1016/j.eswa.2021.116198.

[21] M. Bartoletti, S. Carta, T. Cimoli, R. Saia, Dissecting ponzi schemes on ethereum: Identification, analysis, and impact, Future Generation Computer Systems 102 (2019) 259–277. doi:10.1016/j.future.2019.08.014.

[22] W. Chen, Z. Zheng, J. Cui, E. Ngai, P. Zheng, Y. Zhou, Detecting ponzi schemes on ethereum: Towards healthier blockchain technology, in: WWW 2018, 2018, pp. 1409–1418. doi:10.1145/3178876.3186046.

[23] L. Chen, J. Peng, Y. Liu, J. Li, F. Xie, Z. Zheng, Phishing scams detection in ethereum transaction network, ACM Transactions on Internet Technology 21 (1) (2020) 1–16. doi:10.1145/3398071.

[24] A. Augusto, R. Belchior, J. Pfannschmidt, A. Vasconcelos, M. Correia, Xchainwatcher: Monitoring and identifying attacks in cross-chain bridges (2024). arXiv:2410.02029.
URL https://arxiv.org/abs/2410.02029

[25] N. Christin, Traveling the silk road: A measurement analysis of a large anonymous online marketplace, in: Proceedings of the World Wide Web Conference (WWW 2013), 2013, pp. 213–224. doi:10.1145/2488388.2488408.

[26] Financial Action Task Force (FATF), International standards on combating money laundering and the financing of terrorism and proliferation, fATF Recommendations (2023).

[27] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Computing Surveys 51 (5) (2018) 1–42. doi:10.1145/3236009.

[28] H. Yuan, H. Yu, S. Gui, S. Ji, Explainability in graph neural networks: A taxonomic survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (5) (2022) 1–20. doi:10.1109/TPAMI.2022.3204236.

[29] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, N. F. Samatova, Anomaly detection in dynamic networks: A survey, Wiley Interdisciplinary Reviews: Computational Statistics 7 (3) (2015) 223–247. doi:10.1002/wics.1347.

[30] S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, Data mining for credit card fraud: A comparative study, Decision Support Systems 50 (3) (2010) 602–613. doi:10.1016/j.dss.2010.08.008.

[31] T. Pourhabibi, K. L. Ong, B. H. Kam, Y. L. Boo, Fraud detection: A systematic literature review of graph-based anomaly detection approaches, Decision Support Systems 133 (2020) 113303. doi:10.1016/j.dss.2020.113303.

[32] O. Mounnan, O. Manad, L. Boubchir, A. El Mouatasim, B. Daachi, A review on deep anomaly detection in blockchain, Blockchain: Research and Applications 5 (4) (2024) 100227. doi:10.1016/j.bcra.2024.100227.

[33] V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, G. K. Sandve, Enhancing reproducibility for computational methods, Science 354 (6317) (2016) 1240–1241. doi:10.1126/science.aah6168.

[34] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, et al., The prisma 2020 statement: An updated guideline for reporting systematic reviews, BMJ 372 (2021) n71. doi:10.1136/bmj.n71.

[35] O. Patel, Anomaly detection in cryptocurrency transactions using machine learning, International Journal of Advanced Research in Engineering and Technology 12 (1) (2021) 1266–1282. doi:10.34218/IJARET.12.1.112.

[36] Y. Yan, J. Hou, Z. Song, E. E. Kuruoglu, Signal processing over time-varying graphs: A systematic review (2024). arXiv:2412.00462.
URL https://arxiv.org/abs/2412.00462

[37] P. Xia, H. Wang, B. Gao, W. Su, Z. Yu, X. Luo, C. Zhang, X. Xiao, G. Xu, Trade or trick? detecting and characterizing scam tokens on uniswap decentralized exchange, Proceedings of the ACM on Measurement and Analysis of Computing Systems 5 (3) (2021) 1–26. doi:10.1145/3491051.

[38] J. Kleinberg, Bursty and hierarchical structure in streams, Data Mining and Knowledge Discovery 7 (4) (2003) 373–397. doi:10.1145/775047.775061.

[39] A. Badjie, G. M. Ntuala, Q. Xia, J. Gao, H. Xia, Mg-gpt: A multi-graph gpt-enhanced framework for dynamic fraud detection in cryptocurrency networks, Computer Networks 270 (2025) 111508. doi:10.1016/j.comnet.2025.111508.

[40] I. Alarab, S. Prakoonwit, Effect of data resampling on feature importance in imbalanced blockchain data: Comparison studies of resampling techniques, Data Science and Management 5 (2) (2022) 66–76. doi:10.1016/j.dsm.2022.04.003.

[41] T. Gu, M. Han, S. He, X. Chen, H. Lu, Z. Wang, Online knowledge distillation enabled multi-hop graph attention networks for anonymous transaction regulation in blockchain, Applied Soft Computing 183 (2025) 113644. doi:10.1016/j.asoc.2025.113644.

[42] J. Kang, S. Buu, Graph anomaly detection with disentangled prototypical autoencoder for phishing scam detection in cryptocurrency transactions, IEEE Access 12 (2024) 91075–91088. doi:10.1109/ACCESS.2024.3419152.

[43] Y. Sui, H. Zhang, J. Liu, Epad: Ethereum phishing detection via graph contrastive learning, Expert Systems with Applications 288 (2025) 128227. doi:https://doi.org/10.1016/j.eswa.2025.128227.
URL https://www.sciencedirect.com/science/article/pii/S0957417425018470

[44] R. Hasan, S. Islam, M. Rahman, Detecting anomalies in blockchain transactions using machine learning classifiers and explainability analysis, arXiv preprint arXiv:2401.03530 (2024).
URL https://arxiv.org/abs/2401.03530

[45] M. Ashfaq, S. Khan, I. Ahmad, A machine learning and blockchain based efficient fraud detection mechanism, Sensors 22 (19) (2022) 7423. doi:10.3390/s22197162.

[46] Y. Elmougy, L. Liu, Demystifying fraudulent transactions and illicit nodes in the bitcoin network for financial forensics, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, ACM, 2023, p. 3979–3990. doi:10.1145/3580305.3599803.
URL http://dx.doi.org/10.1145/3580305.3599803

[47] A. Ferreira, J. Silva, Bitcoin fraud detection: A study with dimensionality reduction techniques, in: International Conference on Data Science and Applications, 2025.

[48] Y.-J. Lin, P.-W. Wu, C.-H. Hsu, I.-P. Tu, S. wei Liao, An evaluation of bitcoin address classification based on transaction history summarization (2019). arXiv:1903.07994.
URL https://arxiv.org/abs/1903.07994

[49] T. Pham, S. Lee, Anomaly detection in bitcoin network using unsupervised learning methods, arXiv preprint arXiv:1611.03941 (2016).
URL https://arxiv.org/abs/1611.03941

[50] P. Monamo, V. Marivate, B. Twala, Unsupervised learning for robust bitcoin fraud detection, in: ISSA 2016, 2016, pp. 129–134. doi:10.1109/ISSA.2016.7802939.

[51] M. Shayegan, K. Sabor, A collective anomaly detection method over bitcoin network, arXiv preprint arXiv:2107.00925 (2021).
URL https://arxiv.org/abs/2107.00925

[52] F. Victor, Address clustering heuristics for ethereum, in: Financial Cryptography and Data Security (FC 2020), 2020, pp. 617–633. doi:10.1007/978-3-030-51280-4_33.

[53] O. Lawal, A. Okolie, C. Obunadike, An explainable graph neural network framework for anti-money laundering in cryptocurrency transactions using the elliptic dataset, International Journal of Network Security & Applications 17 (6) (2025) 27–42. doi:10.5121/ijnsa.2025.17602.

[54] Y. Zhang, L. Wang, H. Liu, Temporal-aware graph attention network for cryptocurrency transaction fraud detection (2025). arXiv:2506.21382.
URL https://arxiv.org/abs/2506.21382

[55] H. Kanezashi, et al., Ethereum fraud detection with heterogeneous graph neural networks, arXiv preprint arXiv:2203.12363 (2022).
URL https://arxiv.org/abs/2203.12363

[56] C. Bellei, M. Xu, R. Phillips, T. Robinson, M. Weber, T. Kaler, C. E. Leiserson, Arvind, J. Chen, The shape of money laundering: Subgraph representation learning on the blockchain with the elliptic2 dataset (2024). arXiv:2404.19109.
URL https://arxiv.org/abs/2404.19109

[57] D. H. Son, B. D. Manh, T. V. Khoa, N. L. Trung, D. T. Hoang, T. Minh, Y. Alem, L. Q. Minh, Semi-supervised learning for anomaly detection in blockchain-based supply chains (2024). arXiv:2407.15603.
URL https://arxiv.org/abs/2407.15603

[58] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets, PLoS ONE 10 (3) (2015) e0118432. doi:10.1371/journal.pone.0118432.

[59] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation, BMC Genomics 21 (1) (2020) 1–13. doi:10.1186/s12864-019-6413-7.

[60] L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, J. D. Tygar, Adversarial machine learning, in: AISec 2011, 2011, pp. 43–58. doi:10.1145/2046684.2046692.

[61] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, A. Swami, The limitations of deep learning in adversarial settings, in: EuroS&P 2016, 2016, pp. 372–387. doi:10.1109/EuroSP.2016.36.

[62] European Commission, Proposal for a regulation laying down harmonised rules on artificial intelligence (artificial intelligence act) (2021).
URL https://tinyurl.com/27pd5e5c

[63] O. Shchur, M. Mumme, A. Bojchevski, S. Günnemann, Pitfalls of graph neural network evaluation (2018). arXiv:1811.05868.
URL https://arxiv.org/abs/1811.05868

[64] M. Delange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, T. Tuytelaars, A continual learning survey: Defying forgetting in classification tasks, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (7) (2021) 3366–3385. doi:10.1109/TPAMI.2021.3057446.

[65] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, S. Wermter, Continual lifelong learning with neural networks: A review, Neural Networks 113 (2019) 54–71. doi:10.1016/j.neunet.2019.01.012.