A HYBRID ANALYTICAL FRAMEWORK FOR BITCOIN TRANSACTION NETWORK FORENSIC INVESTIGATION

Mr. M. Hanumantha Rao[1], Lakkaraju Sriram Prasad[2], Manikeswaram Abdul Rahim[3], Kovuri Venkata Manesh[4], Mannava Akash[5]

[1]Assistant Professor, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh ,522017.
Email: hanusmurukutla@gmail.com[1].
[2345]UG Scholar, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh, 522017.
Email: 23jr5a0515@gmail.com[2], 22jr1a05b7@gmail.com[3], 22jr1a05b2@gmail.com[4], 22jr1a05b8@gmail.com[5]

Abstract:
A hybrid analytical framework is developed for the forensic investigation of Bitcoin transaction networks, addressing the inherent challenges posed by the decentralized and pseudo-anonymous characteristics of blockchain systems. While Bitcoin transactions are publicly accessible, detecting illicit activities within complex transaction graphs remains a significant challenge. Existing approaches typically depend on isolated techniques, such as rule-based methods or standalone machine learning models, which often lack sufficient effectiveness.The proposed framework combines graph-based network analysis, statistical modeling, and machine learning to enhance detection capability. Transactions are represented as a directed graph, where wallet addresses function as nodes and transactions as edges. From this representation, structural, behavioral, and temporal features are systematically extracted and integrated into a unified dataset. A Random Forest classifier is subsequently employed to categorize wallet addresses as either normal or suspicious.This integrated approach improves accuracy, scalability, and robustness, facilitating efficient analysis of large-scale blockchain data and enabling more reliable identification of fraudulent activities in real-world forensic investigations.

Keywords: Bitcoin, Blockchain Forensics, Transaction Network Analysis, Machine Learning, Random Forest, Fraud Detection

I. INTRODUCTION

Bitcoin is a decentralized digital currency that enables secure peer-to-peer transactions without the involvement of intermediaries such as banks or financial institutions. It operates on blockchain technology, where all transactions are permanently recorded in a public ledger, ensuring transparency and immutability. Despite this transparency, Bitcoin maintains pseudo-anonymity by using wallet addresses instead of real user identities, making it difficult to trace individuals involved in transactions. This characteristic has led to the increasing use of Bitcoin in illegal activities such as money laundering, fraud, darknet trading, and ransomware payments.Although blockchain data is publicly available, analyzing Bitcoin transaction networks for forensic purposes remains a complex task due to the massive volume of data and intricate relationships between wallet addresses. Traditional forensic methods rely on isolated techniques such as rule-based systems, transaction tracing, or standalone machine learning models. These approaches are often insufficient to capture hidden patterns and sophisticated fraudulent behaviours, resulting in reduced detection accuracy and scalability issues. To overcome these challenges, there is a growing need for an integrated and intelligent analytical approach. This study proposes a hybrid analytical

framework that combines graph-based network analysis, statistical modelling, and machine learning techniques. By modelling transactions as a directed graph and extracting structural, behavioural, and temporal features, the system enhances the detection of suspicious activities. This integrated approach improves accuracy, scalability, and efficiency, making it suitable for real-world Bitcoin forensic investigations.

II. LITERATURE SURVEY

The foundation of Bitcoin and its decentralized architecture was introduced by Satoshi Nakamoto (2008), which enabled secure peer-to-peer transactions without intermediaries. However, the pseudo-anonymous nature of Bitcoin has led to challenges in forensic analysis. Early studies by Reid and Harrigan (2013) and Ron and Shamir (2013) analyzed Bitcoin transaction graphs and highlighted the possibility of linking wallet addresses through network patterns. Meiklejohn et al. (2013) further demonstrated how transaction clustering techniques can reveal user identities despite anonymity.Several tools and platforms such as BlockSci (Kalodner et al., 2017) and BitIodine (Spagnuolo et al., 2014) were developed to support blockchain analysis. Conti et al. (2018) provided a comprehensive survey of security and privacy issues in Bitcoin, emphasizing the need for advanced forensic techniques. With the advancement of machine learning, algorithms such as Random Forest (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016) have been widely used for fraud detection due to their robustness and scalability.

Graph-based deep learning approaches, including Graph Convolutional Networks (Kipf and Welling, 2017) and Graph Attention Networks (Veličković et al., 2018), have shown improved performance in capturing complex relationships in transaction networks. Additionally, data mining and anomaly detection techniques (Aggarwal, 2017) play a significant role in identifying suspicious patterns. Despite these advancements, existing methods often rely on single techniques, highlighting the need for hybrid frameworks that integrate graph analysis, statistical methods, and machine learning for more accurate Bitcoin forensic investigation.

III. PROPOSED WORK

The proposed work introduces a hybrid analytical framework for effective forensic investigation of Bitcoin transaction networks by integrating graph-based analysis, statistical modeling, and machine learning techniques. Initially, Bitcoin transaction data is collected from both historical and real-time blockchain sources to ensure comprehensive coverage. The data is preprocessed and transformed into a directed graph structure, where wallet addresses are represented as nodes and transactions as edges.

Graph-based analysis is applied to identify structural patterns such as highly connected nodes, clusters, and suspicious transaction paths. Statistical modeling is then used to analyze transaction behavior, including frequency, transaction volume, and value distribution, to detect anomalies. Temporal analysis further enhances detection by identifying unusual timing patterns and repeated transactions that may indicate fraudulent activities.

All extracted features structural, behavioural, and temporal are combined into a unified dataset. A Random Forest classifier is then employed to classify wallet addresses as normal or suspicious. The model is trained and validated to ensure high accuracy and robustness. Finally, the system generates detailed forensic reports, providing actionable insights for investigators. This integrated framework improves detection accuracy, scalability, and efficiency compared to traditional methods.

IV. METHODOLOGY

A. Dataset Description

The proposed framework utilizes the Elliptic Bitcoin Dataset, a widely recognized benchmark for cryptocurrency transaction analysis and fraud detection. This dataset contains real-world Bitcoin transaction data labeled as licit, illicit, and unknown, enabling supervised learning for forensic applications. It comprises approximately 200,000 transactions with 166 features per transaction, including both local features (transaction-specific attributes) and aggregated features derived from neighboring transactions. Additionally, the dataset incorporates temporal information across multiple time steps, which facilitates the analysis of transaction behavior over time. The availability of labeled and structured data makes this dataset highly suitable for training and evaluating machine learning models in identifying suspicious financial activities within blockchain networks.

B. System Architecture

The proposed system is designed using a multi-layered architecture to ensure efficient processing, scalability, and modularity in analyzing large-scale Bitcoin transaction data. The architecture consists of several interconnected layers, including data acquisition, preprocessing, graph construction, feature extraction, classification, and result generation. Initially, transaction data is collected from the Elliptic dataset and relevant blockchain sources. The data is then preprocessed to improve quality and consistency. Subsequently, the processed data is transformed into a graph structure to capture relationships between wallet addresses. Feature extraction techniques are applied to derive meaningful attributes from the graph and transaction data. These features are then used to train a machine learning model for classification. Finally, the system generates detailed forensic outputs, including identification of suspicious wallet addresses and transaction patterns. This structured architecture enables efficient handling of complex transaction networks and supports real-time analytical capabilities.

Figure 1. Proposed Workflow Architecture for Bitcoin Transaction Forensic Analysis

C. Data Preprocessing

Data preprocessing is a critical step to ensure the reliability and quality of the input data used for analysis. The collected transaction data undergoes several preprocessing operations, including removal of duplicate entries, handling of missing values, and correction of inconsistencies. Numerical features are normalized to maintain uniformity across different scales, thereby improving the performance of machine learning algorithms. Additionally, irrelevant or noisy data is filtered out to reduce computational complexity and enhance model accuracy. This step ensures that the dataset is clean, consistent, and suitable for subsequent graph modeling and feature extraction processes.

D. Graph Construction Using NetworkX

The Bitcoin transaction data is modeled as a directed graph using the NetworkX library, which provides efficient tools for complex network analysis. In this representation, each wallet address is considered a node, while transactions between addresses are represented as directed edges. Edge attributes such as transaction amount, timestamp, and transaction identifiers are incorporated to enrich the graph structure. NetworkX facilitates the computation of key graph metrics, including degree centrality, betweenness centrality, and clustering coefficients, which are essential for understanding network topology and identifying influential or suspicious nodes. This graph-based representation enables the detection of hidden relationships, transaction flows, and anomalous patterns that are not easily identifiable through traditional data analysis methods.

E. Feature Extraction

Feature extraction involves deriving meaningful attributes from both the graph structure and transaction data to support accurate classification. Structural features, such as node degree, centrality measures, and connectivity patterns, are obtained from graph analysis. Behavioral features, including transaction frequency, total transaction value, and average transaction size, are computed using statistical techniques. Temporal features are also extracted to capture time-based patterns, such as transaction intervals, repeated activities, and sudden bursts in transaction behavior. These features collectively provide a comprehensive representation of wallet activity and transaction dynamics. The integration of structural, behavioral, and temporal features enhances the model's ability to distinguish between normal and suspicious entities.

F. Model Training and Classification

The extracted features are combined into a unified dataset, which is used to train a supervised machine learning model for classification. A Random Forest classifier is employed due to its robustness, scalability, and ability to handle high-dimensional data effectively. The model constructs multiple decision trees and aggregates their outputs to improve prediction accuracy and reduce overfitting. It is trained using labeled data from the Elliptic dataset to distinguish between normal and suspicious wallet addresses. The

ensemble nature of Random Forest enables it to capture complex, non-linear relationships within the data, resulting in reliable and consistent classification performance.

G. Result Generation

The final stage of the methodology involves generating actionable outputs based on the classification results. The system identifies suspicious wallet addresses and highlights anomalous transaction patterns within the network. These results are presented in the form of detailed forensic reports, which may include risk scores, transaction clusters, and visual insights for easier interpretation. The generated outputs assist investigators in understanding complex transaction behaviors and making informed decisions. This step enhances the practical applicability of the proposed framework in real-world blockchain forensic investigations.

V. ALGORITHMS

1. Graph-Based Network Analysis

Graph-based network analysis is used to model Bitcoin transactions as a directed graph, where wallet addresses are nodes and transactions are edges. This approach helps in understanding relationships between different entities and identifying hidden connections. Key metrics such as node degree, centrality, and clustering coefficients are analyzed to detect abnormal patterns. Suspicious activities often form dense clusters or highly connected nodes, making graph analysis an effective tool for uncovering complex transaction behaviors.

2. Statistical Analysis

Statistical analysis is applied to examine transaction behavior and detect anomalies. Features such as transaction frequency, total transaction value, and average transaction size are calculated. Distribution patterns and deviations from normal behavior are analyzed to identify unusual activities. This method helps in detecting sudden spikes, irregular transaction volumes, and inconsistent behavior across wallets. Statistical insights complement graph analysis by providing quantitative measures that improve the detection of suspicious transactions.

3. Temporal Analysis

Temporal analysis focuses on examining transaction patterns over time. It identifies unusual timing behaviors such as rapid repeated transactions, irregular intervals, or coordinated activities across multiple wallets. Time-based patterns are crucial in detecting fraud schemes like money laundering, where transactions are often structured in specific sequences. By incorporating time-related features, this analysis enhances the system's ability to detect dynamic and evolving fraudulent activities within the Bitcoin network.

4. Random Forest Algorithm

The Random Forest algorithm is a supervised machine learning technique used for classification. It builds multiple decision trees and combines their outputs to improve accuracy and reduce overfitting. In this system, it is used to classify wallet addresses as normal or suspicious based on extracted features. Random Forest handles high-dimensional data effectively and provides reliable predictions. Its ensemble nature ensures robustness, making it suitable for detecting complex fraud patterns in blockchain data.

VI. RESULTS AND DISCUSSION

The proposed hybrid analytical framework demonstrates strong performance in detecting suspicious Bitcoin transactions. By combining graph-based features, statistical analysis, and Random Forest classification, the system achieves higher accuracy and reliability compared to traditional methods. The results highlight improved detection capability, reduced false positives, and better scalability for large blockchain datasets.

Table 1: Performance Comparison of Algorithms

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	85	83	82	82
Rule-Based	80	78	76	77
GNN	91	89	90	89
Proposed (RF)	93	91	92	91

Table 1 compares the performance of different algorithms including SVM, Rule-Based, GNN, and the proposed Random Forest model. It evaluates them using accuracy, precision, recall, and F1-score. The proposed model achieves the

highest accuracy of 93%, outperforming all other methods. This indicates its superior ability to detect suspicious wallet activities. The table clearly demonstrates that integrating multiple analytical techniques improves classification performance and provides a more reliable solution for Bitcoin forensic investigation.

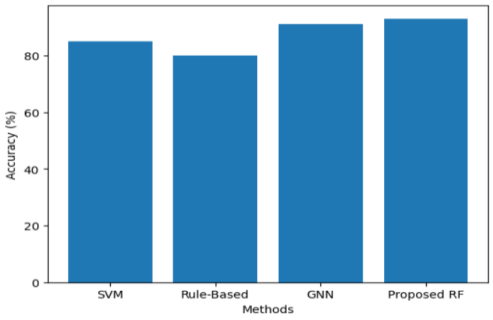

Figure 1: Accuracy Comparison

Figure 1 visually represents the accuracy of different algorithms used in the system. It shows that the proposed Random Forest model achieves the highest accuracy compared to SVM, Rule-Based, and GNN approaches. The bar chart format makes it easy to compare performance differences among models. This visualization highlights the effectiveness of the hybrid framework, as it significantly improves accuracy by combining graph analysis and machine learning techniques for better fraud detection.

Table 2: Proposed Model Evaluation

Metric	Value (%)
Accuracy	93
Precision	91
Recall	92
F1-Score	91

Table 2 presents the evaluation metrics of the proposed Random Forest model, including accuracy, precision, recall, and F1-score. The model achieves consistently high values across all metrics, indicating balanced performance. High precision reflects fewer false positives, while high recall shows effective detection of actual fraudulent activities. The F1-score confirms overall robustness. This table demonstrates that the proposed system provides reliable and efficient classification results suitable for real-world blockchain forensic applications.

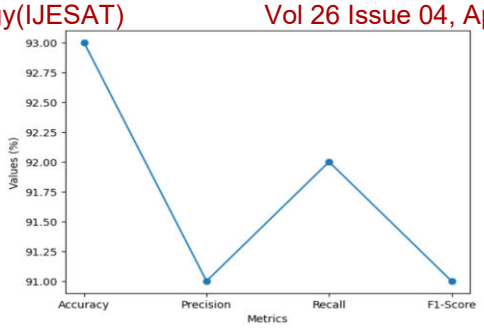

Figure 2: Performance Metrics of Proposed Model

Figure 2 illustrates the performance metrics of the proposed model using a line graph. It shows accuracy, precision, recall, and F1-score values, all of which are above 90%. The smooth trend in the graph indicates stable and consistent model performance across different evaluation measures. This visualization helps in understanding the balance between precision and recall, confirming that the system is both accurate and reliable in detecting suspicious Bitcoin transactions.

Table 3: Feature Contribution Analysis

Feature Type	Contribution (%)	Description
Graph Features	40	Captures network relationships
Statistical	25	Identifies transaction anomalies
Temporal	35	Detects time-based fraud patterns

Table 3 highlights the contribution of different feature types used in the model, including graph-based, statistical, and temporal features. Graph features contribute the most, followed by temporal and statistical features. This indicates that network relationships play a key role in detecting suspicious activities. The table emphasizes the importance of combining multiple feature types to improve detection accuracy. It also shows how each feature type adds value to the overall performance of the hybrid framework.

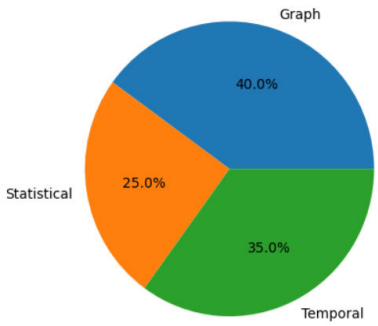

Figure 3: Feature Contribution

Figure 3 represents the contribution of different feature types using a pie chart. It shows that graph features have the highest contribution, followed by temporal and statistical features. The visualization clearly illustrates the proportion of each feature in the model's decision-making process. This graph helps in understanding the importance of feature integration and highlights how combining multiple analytical approaches enhances the effectiveness of the proposed forensic system.

CONCLUSION

This study presents a hybrid analytical framework for Bitcoin transaction network forensic investigation that effectively addresses the challenges posed by the decentralized and pseudo-anonymous nature of blockchain systems. By integrating graph-based network analysis, statistical modeling, and machine learning techniques, the proposed system enhances the detection of suspicious wallet activities and fraudulent transaction patterns. Modeling transactions as a directed graph enables the identification of hidden relationships, while statistical and temporal features help capture abnormal behaviors. The use of a Random Forest classifier ensures high accuracy, robustness, and efficient handling of complex datasets. Experimental results demonstrate that the proposed approach outperforms traditional methods such as rule-based systems and standalone machine learning models in terms of accuracy, precision, recall, and scalability. The system also reduces false positives and improves reliability, making it suitable for real-world forensic applications. Overall, the hybrid framework provides a comprehensive and scalable solution for blockchain analysis and can be further extended to support other cryptocurrencies and emerging cybercrime detection scenarios.

FUTURE SCOPE

Future enhancements of the proposed framework can focus on integrating advanced and emerging technologies to further improve detection accuracy and scalability. One potential direction is the incorporation of Transformer-based models and attention mechanisms to better capture complex sequential and temporal transaction patterns within blockchain networks. The framework can also be extended by implementing real-time data streaming and online learning techniques, enabling continuous monitoring of Bitcoin transactions and dynamic detection of evolving fraudulent activities.Additionally, integrating external data sources, such as exchange records, IP metadata, and user behavior analytics, can improve the identification of real-world entities behind wallet addresses.To address scalability challenges, distributed computing frameworks such as Apache Spark can be utilized for efficient processing of large-scale blockchain datasets.Another important direction is the adoption of Explainable Artificial Intelligence (XAI) techniques, which provide transparency and interpretability in model predictions, assisting forensic investigators in understanding decision-making processes.Furthermore, the framework can be extended to support multi-cryptocurrency and cross-chain forensic analysis, enabling investigation across different blockchain ecosystems.

REFERENCES

1. Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008.
2. Dorit Ron and Adi Shamir, "Quantitative Analysis of the Full Bitcoin Transaction Graph," *Financial Cryptography*, 2013.
3. Fergal Reid and Martin Harrigan, "An Analysis of Anonymity in the Bitcoin System," *IEEE Security & Privacy*, 2013.
4. Sarah Meiklejohn et al., "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names," *IMC*, 2013.
5. Harry Kalodner et al., "BlockSci: Design and Applications of a Blockchain Analysis Platform," *USENIX Security*, 2017.
6. Michele Spagnuolo et al., "BitIodine: Extracting Intelligence from the Bitcoin Network," 2014.

7. Monero Research Lab, "An Empirical Analysis of Linkability in the Monero Blockchain," 2017.
8. Victor M. Shoup, "Applied Cryptography and Blockchain Security," 2019.
9. Tianxiang Chen et al., "XGBoost: A Scalable Tree Boosting System," *KDD*, 2016.
10. Leo Breiman, "Random Forests," *Machine Learning Journal*, 2001.
11. Thomas N. Kipf and Max Welling, "Semi-Supervised Classification with Graph Convolutional Networks," 2017.
12. Petar Veličković et al., "Graph Attention Networks," 2018.
13. Jian Pei et al., "Data Mining Techniques for Fraud Detection," 2019.
14. Whitfield Diffie and Martin Hellman, "New Directions in Cryptography," *IEEE*, 1976.
15. M. Conti, S. Kumar, C. Lal, and S. Ruj, "A Survey on Security and Privacy Issues of Bitcoin," *IEEE Communications Surveys*, 2018.
16. Chen, T., and Guestrin, C., "XGBoost: Extreme Gradient Boosting," 2016.
17. Aggarwal, C. C., "Outlier Analysis," Springer, 2017.
18. Goodfellow, I., Bengio, Y., and Courville, A., "Deep Learning," MIT Press, 2016.
19. Todupunuri, A. (2025). The Role of Human-Centric AI in Building Trust in Digital Banking Ecosystems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120605
20. Babburi, S. Privacy-Preserving Collaborative Framework with Auditable Federated Learning.
21. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
22. Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH,1(4). https://doi.org/10.56975/jaafr.v1i4.501636
23. S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology,16(1). https://doi.org/10.71097/ijsat.v16.i1.1403
24. Doragacharla, V. R. (2026). AI-Enabled Commerce Platforms in Cloud Computing Environments: An Architectural and Socio-Economic Analysis. Journal of Computational Analysis & Applications, 35(1).
25. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
26. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
27. Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50
28. Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.
29. Banda Saikumar. (2025). Integrating azure network rules for storage account through terraform in CI/CD pipelines: automating storage account access restrictions to public IP. Journal of Scien+B112ce & Technology, 10(2), 15–22. https://doi.org/10.46243/jst.2025.v10.i02.pp15-22
30. Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
31. Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782
32. Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.