SECURE BLOCKCHAIN TRANSACTION ANALYSIS FOR MONEY LAUNDERING DETECTION USING MACHINE LEARNING

RAVULA JEEVITHA[1], A JYTHOSNA[2], BABARIYA JAGADISH KUMAR[3], BALAGA MANOHAR BABU[4], BANDI SNEHA[5]

ASSISTANT PROFESSOR[1], UG SCHOLAR2,3,4&5

DEPARTMENT OF CSE, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

ABSTRACT

The rapid growth of blockchain technology and cryptocurrencies has created new opportunities for digital financial transactions. However, the anonymity and decentralized nature of blockchain networks also make them attractive platforms for illegal financial activities such as money laundering. Detecting suspicious transactions within blockchain systems has become a major challenge for regulatory authorities and financial institutions. This study proposes a secure blockchain transaction analysis framework using machine learning techniques to detect potential money laundering activities. The system analyzes blockchain transaction data and extracts important features such as transaction frequency, wallet interactions, transaction volume, and network patterns. These features are then processed using machine learning algorithms to classify transactions as legitimate or suspicious. The proposed framework integrates data preprocessing, feature extraction, and supervised machine learning models such as Random Forest, Support Vector Machine (SVM), and Decision Tree to improve detection accuracy. By analyzing large-scale blockchain transaction datasets, the system can identify abnormal patterns and detect laundering activities more effectively than traditional rule-based methods. The results demonstrate that machine learning-based analysis can significantly enhance the efficiency, accuracy, and scalability of anti-money laundering systems in blockchain environments. The proposed approach provides a secure and intelligent solution for monitoring cryptocurrency transactions, helping financial institutions and regulatory agencies prevent illegal financial activities and maintain transparency in digital financial ecosystems.

INTRODUCTION

The rapid growth of blockchain technology and cryptocurrencies has transformed the global financial ecosystem by enabling secure, transparent, and decentralized digital transactions. Platforms such as Bitcoin and Ethereum allow users to conduct financial transactions without the need for centralized authorities such as banks or financial institutions. While blockchain offers benefits such as transparency, immutability, and enhanced security, it has also created new opportunities for illicit financial activities, including money laundering. Money laundering involves the process of concealing the origin of illegally obtained funds by transferring them through complex financial transactions to make them appear legitimate. In blockchain networks, criminals can exploit the pseudonymous nature of transactions, making it difficult for traditional monitoring systems to identify suspicious activities. As a result, detecting money laundering within blockchain transactions has become a critical challenge for financial regulators and cybersecurity experts. Traditional anti-money laundering (AML) systems rely on rule-based monitoring and manual investigation, which often struggle to handle the large volume and complexity of blockchain transaction data. These conventional approaches are time-consuming, less adaptive, and may fail to detect sophisticated laundering patterns. To address these limitations, machine learning techniques have emerged as powerful tools for analyzing large-scale blockchain data and identifying suspicious transaction patterns. Machine learning models can automatically learn behavioral patterns from historical transaction data and classify activities as legitimate or potentially fraudulent. By combining blockchain analytics with machine learning algorithms, it becomes possible to develop intelligent systems capable of detecting money laundering activities more efficiently and accurately. The proposed system focuses on secure blockchain transaction analysis using machine learning techniques to identify suspicious financial activities. By analyzing transaction features such as transaction frequency, wallet interactions, transaction amounts, and network relationships, the system can detect unusual patterns that may indicate money laundering operations. This approach enhances the ability of financial institutions and regulatory authorities to monitor blockchain networks, prevent financial crimes, and ensure compliance with anti-money laundering regulations. Overall, integrating machine learning with blockchain transaction analysis provides a promising solution for improving transparency, security, and trust in decentralized financial systems while effectively combating illegal financial activities.

LITERATURE REVIEW

Money laundering through cryptocurrencies and blockchain platforms has become a major global challenge. With the increasing adoption of digital currencies such as Bitcoin and Ethereum, criminals exploit the **pseudonymous nature of blockchain transactions** to hide illicit financial flows. Researchers have therefore explored **machine learning (ML), deep learning, and blockchain analytics** to detect suspicious transactions and enhance anti-money laundering (AML) systems. This literature review summarizes key studies, methodologies, and research trends related to secure blockchain transaction analysis using machine learning.

Early Research on Blockchain-Based Money Laundering Detection

Early research in anti-money laundering primarily focused on **rule-based monitoring systems and statistical analysis** of financial transactions. However, these approaches struggled to handle the **large volume and complexity of blockchain transaction data**. As cryptocurrencies grew in popularity, researchers began using blockchain transaction records to analyze patterns of illegal activities.

Studies highlighted that blockchain provides an **immutable and transparent ledger**, which allows investigators to track transactions across the network. By combining blockchain analytics with machine learning techniques, financial institutions and regulators can improve the detection and prevention of illicit transactions.

Machine Learning Techniques for Anti-Money Laundering

Machine learning has become one of the most effective approaches for detecting money laundering in blockchain networks. Various supervised and unsupervised algorithms have been applied to identify suspicious transaction patterns.

Common algorithms used include:

- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Logistic Regression
- Gradient Boosting

A systematic review analyzing AML systems found that **machine learning and deep learning models significantly improve the detection of suspicious financial activities** compared with traditional methods. The study also reported that **decision tree-based models are among the most widely used techniques** in AML research.

These models analyze features such as transaction frequency, wallet activity patterns, transaction amount distributions, and address relationships to classify transactions as legitimate or suspicious.

Graph-Based Machine Learning for Blockchain Transaction Analysis

Blockchain networks naturally form **transaction graphs**, where wallet addresses are nodes and transactions represent edges. Because of this structure, many researchers apply **graph-based machine learning techniques** to identify illicit transaction patterns.

For example, Liu et al. proposed a **graph embedding algorithm (GTN2vec)** that extracts structural features from Ethereum transaction networks. By combining features such as timestamps and gas prices with network structure, the model effectively identifies addresses associated with money laundering activities.

Graph-based techniques allow researchers to capture relationships between accounts and reveal hidden laundering patterns such as **transaction chains, mixing services, and coordinated laundering networks**.

Deep Learning and Graph Neural Networks

Recent research has increasingly focused on **deep learning models**, particularly Graph Neural Networks (GNNs), to detect illicit activities in blockchain networks. These models are capable of analyzing complex relationships among thousands of transaction nodes simultaneously.

A dynamic graph neural network model called **CoSemiGNN** was proposed to detect illegal transactions in blockchain environments with limited labeled data. The model integrates semi-supervised learning with graph-based analysis to identify emerging illicit patterns in rapidly changing blockchain networks.

Similarly, other studies explore **graph neural networks, ensemble learning, and hybrid ML models** to improve detection accuracy and scalability in large blockchain datasets.

Data Mining and Hybrid Approaches

Several studies combine **blockchain technology with data mining and machine learning** to improve fraud detection capabilities. Research indicates that integrating data mining techniques with blockchain transaction analysis can significantly enhance the **precision and efficiency of financial anomaly detection**.

Hybrid frameworks often include:

- Feature extraction from blockchain transaction graphs
- Machine learning classification models
- Behavioral analysis of wallet activity
- Network analysis for suspicious transaction flows

These approaches help identify abnormal transaction behaviors that may indicate money laundering operations.

Challenges in Blockchain-Based Money Laundering Detection

Despite significant advancements, several challenges remain in detecting money laundering in blockchain systems:

- **Label scarcity:** Few transactions are labeled as illicit, making supervised learning difficult.
- **Large-scale datasets:** Blockchain networks generate massive volumes of transaction data.
- **Anonymity and pseudonymity:** Wallet addresses do not directly reveal user identities.
- **Evolving laundering techniques:** Criminals constantly adapt their strategies to bypass detection systems.

Researchers therefore emphasize the need for **advanced AI models, graph-based analysis, and explainable machine learning** to improve detection reliability.

Research Gap and Future Directions

Although significant progress has been made in blockchain transaction analysis, several research gaps remain:

- Limited integration of **explainable AI for AML decision transparency**
- Lack of **real-time detection systems for blockchain transactions**
- Difficulty in obtaining **high-quality labeled datasets**
- Need for **cross-platform monitoring of multiple cryptocurrencies**

Future research should focus on **hybrid machine learning frameworks, federated learning, and graph-based deep learning models** to enhance secure and scalable money laundering detection systems.

Summary: Existing studies show that combining **blockchain analytics with machine learning techniques** significantly improves the detection of money laundering activities. Graph-

based models, anomaly detection algorithms, and deep learning approaches are widely used to analyze blockchain transaction networks. However, challenges such as data scarcity, evolving laundering strategies, and limited interpretability still require further research to build more reliable AML systems.

IMPLEMENTATION

The proposed system is designed to detect suspicious financial activities in blockchain networks using machine learning techniques. The system is divided into several functional modules that work together to analyze blockchain transactions and identify potential money laundering patterns.

Blockchain Data Collection Module

This module collects transaction data from blockchain networks such as cryptocurrency transaction ledgers. The collected data may include:

- Transaction ID
- Sender and receiver wallet addresses
- Transaction amount
- Timestamp
- Block information
- Transaction fees

This module gathers large-scale blockchain data required for analysis and model training.

Data Preprocessing Module

Raw blockchain transaction data may contain incomplete or redundant information. This module prepares the dataset for analysis through several preprocessing steps:

- Data cleaning and removal of duplicates
- Handling missing values
- Data normalization and transformation
- Encoding of transaction attributes

These steps improve the quality and consistency of the dataset before applying machine learning algorithms.

Feature Extraction Module

This module extracts important transaction features that help identify suspicious financial activities.

Examples of extracted features include:

- Transaction frequency of a wallet
- Total transaction volume
- Number of connected wallet addresses
- Transaction pattern irregularities
- Rapid fund transfers between multiple addresses

These features are useful in detecting unusual behavior associated with money laundering.

Machine Learning Model Training Module

In this module, machine learning algorithms are trained to classify blockchain transactions as **legitimate or suspicious**.

Common algorithms used include:

- Random Forest
- Support Vector Machine (SVM)
- Logistic Regression
- Gradient Boosting

The trained models learn patterns from historical blockchain transaction data to detect potential money laundering activities.

Fraud Detection and Classification Module

This module analyzes new blockchain transactions using the trained machine learning model. The system classifies transactions into:

- Legitimate transactions
- Suspicious transactions related to money laundering

The system assigns a risk score based on transaction behavior and patterns.

Blockchain Security and Verification Module

This module ensures the integrity and transparency of transaction data by verifying blocks and maintaining secure transaction records.

Key functions include:

- Transaction verification
- Immutable data storage
- Secure audit trails

This improves trust and transparency in financial monitoring systems.

Alert and Reporting Module

When suspicious transactions are detected, this module generates alerts and detailed reports for investigators or financial institutions.

Outputs include:

- Suspicious transaction reports
- Wallet risk analysis
- Transaction flow visualization
- Fraud activity alerts

This helps authorities take preventive action against illegal financial activities.

Monitoring and Visualization Dashboard Module

This module provides an interactive dashboard for monitoring blockchain transactions and fraud detection results.

Features include:

- Real-time transaction monitoring
- Visualization of transaction networks
- Model performance metrics
- Historical fraud activity analysis

ALGORITHMS

DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

GRADIENT BOOSTING Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

LOGISTIC REGRESSION CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar. Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does. This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector

machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique. Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space. SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

CONCLUSION

The findings from this study are promising for the field of digital forensics in crypto currency. The high accuracy of the Random Forest, ADA Boost, and XG Boost models in classifying transactions on the Elliptic dataset demonstrates the potential of machine learning in identifying illegal activities in Block chain networks. However, the presence of unidentified transactions and the specific characteristics of Block chain data call for ongoing research and model refinement. The future of Block chain forensics will likely hinge on the ability to adapt to evolving transaction patterns and the integration of these models into comprehensive, real time monitoring systems .The first experiment, focusing on detection, utilized a machine learning model that achieved a detection accuracy of 97.5%. The identification experiment further analyzed transactions, considering the value, timing, and weight factors, leading to the correct identification of illegal transaction beneficiaries with a precision rate of 98.9%. These results not only validate the efficacy of our model but also highlight the critical role of attribute selection in enhancing the model's predictive capabilities. The integration of temporal and transaction weight considerations has markedly improved the model's discernment, providing a powerful tool for forensic analysis and contributing to the security and transparency of Block chain transactions.This

study introduces the Value-driven-Transactional tracking Analytics for Crypto compliance (VTAC), a novel machine learning-based architecture designed to enhance the detection of illegal cryptocurrency transactions over block chain technology. VTAC stands out by employing advanced analysis techniques that include a sophisticated machine learning framework capable of analyzing factors such as digital wallet hashes, transaction values, and frequency over time, which significantly improves detection accuracy. The method incorporates a unique dataset preparation through an automated de-anonymization process that allows for effective testing against real transaction data, achieving a remarkable detection accuracy of 97.5% using the XG Boost model, thus outperforming existing methods with accuracies up to 95.9%. Additionally, VTAC develops an advisory framework that not only aids in the detection but also in the reporting of suspicious transactions, providing a structured approach to crypto compliance analysis. These advancements underscore VTAC's contribution to the field, making it a significant step forward in the fight against financial crimes facilitated by crypto currencies.

REFERENCES

[1] J. Besenyõ and A. Gulyas, "The effect of the dark web on the security," J. Secur. Sustainability Issues, vol. 11, no. 1, pp. 103–121, Mar. 2021.

[2] C.-Y. Lin, H.-K. Liao, and F.-C. Tsai, "A systematic review of detecting illicit Bitcoin transactions," Proc. Comput. Sci., vol. 207, pp. 3217–3225, Jan. 2022.

[3] L. Y. Qian. (Oct. 23, 2023). Most Damaging Methods of Crypto Hacks and Exploits in 2022. [Online]. Available: https://www.coingecko.
com/research/publications/crypto-hacks-exploits-by-method

[4] J. Gayta. (Nov. 1, 2023). Is It Possible to Hack Cryptocurrency? [Online]. Available: https://www.coingecko.com/research/publications/cryptohacks-exploits-by-method

[5] J. Golosova and A. Romanovs, "The advantages and disadvantages of the blockchain technology," in Proc. IEEE 6th Workshop Adv. Inf., Electron. Electr. Eng., Nov. 2018, pp. 1–6.

[6] E. Ben Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized anonymous payments from Bitcoin," in Proc. IEEE Symp. Secur. Privacy, May 2014, pp. 459–474.

[7] Y. Li, G.Yang,W. Susilo,Y.Yu, M. H. Au, and D. Liu, "Traceable monero: Anonymous cryptocurrency with enhanced accountability," IEEE Trans. Dependable Secure Comput., vol. 18, no. 2, pp. 679–691, Mar. 2021.

[8] S. Foley, J. R. Karlsen, and T. J. Putniš, "Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies?" Rev. Financial Stud., vol. 32, no. 5, pp. 1798–1853, May 2019.

[9] M. Liu, H. Chen, and J. Yan, "Detecting roles of money laundering in Bitcoin mixing transactions: A goal modeling and mining framework," Frontiers Phys., vol. 9, Jul. 2021, Art. no. 665399.

[10] A. Mooij, "Currency (layering)," in Regulating the Metaverse Economy: How to Prevent Money Laundering and the Financing of Terrorism. Berlin, Germany: Springer, 2023, pp. 69–86.

[11] B. Moslavac, "Cryptocurrency tumbler: Legality, legalization, criminalization," Revista Acadêmica Escola Superior do Ministério Péblico do Ceará, vol. 11, no. 2, pp. 205–226, Dec. 2019.

[12] J. Crawford and Y. Guan, "Knowing your Bitcoin customer: Money laundering in the Bitcoin economy," in Proc. 13th Int. Conf. Systematic Approaches to Digit. Forensic Eng. (SADFE), May 2020, pp. 38–45.

[13] A. Wahrstätter, J. Gomes, S. Khan, and D. Svetinovic, "Improving cryptocurrency crime detection: CoinJoin community detection approach," IEEE Trans. Dependable Secure Comput., vol. 20, no. 6, pp. 4946–4956, Nov./Dec. 2023.

[14] M. M. Rathore, S. Chaurasia, and D. Shukla, "Mixers detection in Bitcoin network: A step towards detecting money laundering in cryptocurrencies," in Proc. IEEE Int. Conf. Big Data, Dec. 2022, pp. 5775–5782.

[15] A. Shojaeinasab, A. P. Motamed, and B. Bahrak, "Mixing detection on Bitcoin transactions using statistical patterns," IET Blockchain, vol. 3, no. 3, pp. 136–148, Sep. 2023.