Research Paper E-ISSN NO : 2455-295X | VOLUME : 10 | ISSUE : 12 | DECEMBER 2024

A MULTI-LAYERED APPROACH FOR DETECTING CRYPTO-LINKED MONEY LAUNDERING USING ENSEMBLE MACHINE LEARNING MODELS

V. BACKIYALAKSHMI [1]

[1] ASSISTANT PROFESSOR OF COMPUTER SCIENCE, THE AMERICAN COLLEGE, MADURAI.

ABSTRACT:

The growing adoption of crypto currencies has introduced new challenges in financial crime detection, particularly in identifying and preventing crypto-linked money laundering. Traditional anti-money laundering (AML) systems are often insufficient to handle the complexities of crypto currency transactions, including pseudonymity, decentralization, and high transaction volumes. This research proposes a multi-layered approach using ensemble machine learning models to improve the detection of suspicious activities and potential money laundering schemes in the crypto currency ecosystem. By combining multiple machine learning algorithms, such as decision trees, random forests, and support vector machines, this study aims to enhance the accuracy, robustness, and interpretability of detection systems. The proposed framework focuses on transaction pattern analysis, anomaly detection, and feature engineering to identify hidden illicit activities, minimizing false positives while ensuring real-time monitoring capabilities. The results show that ensemble models significantly outperform individual models, providing a scalable and effective solution for combating crypto-linked money laundering.

KEYWORDS:

CRYPTOCURRENCY, ANTI-MONEY LAUNDERING, MACHINE LEARNING MODELS, TRANSACTION PATTERN ANALYSIS, ANOMALY DETECTION.

PAPER ACCEPTED DATE:	PAPER PUBLISHED DATE:
5th December 2024	9th December 2024

1. INTRODUCTION

Crypto currency transactions are gaining widespread adoption due to their speed, low cost, and decentralization. However, these features also make crypto currencies an attractive tool for money laundering activities. The pseudonymous nature of block chain transactions, where users are identified only by wallet addresses rather than real identities, complicates the detection of illicit financial activities. Traditional anti-money laundering systems, which are designed for centralized financial institutions, struggle to detect and mitigate crypto-linked money laundering, especially in decentralized and peer-to-peer networks.

The application of machine learning (ML) offers a promising solution by analyzing large datasets, identifying patterns, and detecting anomalies that may indicate money laundering activities. However, traditional ML models often face challenges such as limited interpretability, high false-positive rates, and an inability to effectively capture complex relationships within crypto currency transaction data. Ensemble machine learning models, which combine multiple models to improve performance, provide a potential solution to these challenges. This paper explores the effectiveness of a multi-layered ensemble approach for detecting crypto-linked money laundering.

2. BACKGROUND AND RELATED WORK

2.1. CRYPTO-LINKED MONEY LAUNDERING TECHNIQUES

Crypto-linked money laundering involves various sophisticated techniques, such as:

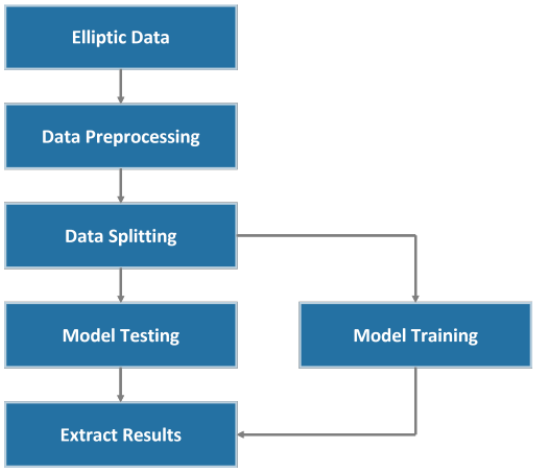
- *Mixing and Tumbling:* Obfuscating the origins of funds by pooling transactions from multiple users and redistributing them.
- *Layering:* Moving funds through multiple wallets or exchanges to disguise their source.
- *Cross-Chain Transactions:* Using different crypto currencies and block chain platforms to make tracking funds more difficult.

2.2. EXISTING APPROACHES FOR CRYPTO-LINKED MONEY LAUNDERING DETECTION

Existing research on crypto-linked money laundering detection primarily relies on supervised and unsupervised machine learning techniques. Common approaches include anomaly detection, clustering, and classification algorithms. These methods have shown some success in identifying suspicious activities but often suffer from high false-positive rates, especially in the dynamic and decentralized world of crypto currencies. Additionally, many approaches lack the scalability needed to process the massive volume of block chain transactions in real-time.

3. METHODOLOGY

3.1. MULTI-LAYERED ENSEMBLE APPROACH



The proposed solution leverages a multi-layered ensemble machine learning approach, which combines several models to enhance detection performance. The ensemble model improves accuracy by aggregating the results from multiple base learners, thus reducing the likelihood of false positives and increasing the robustness of the system. The following models are included in the ensemble:

- *Decision Trees:* Known for their simplicity and interpretability, decision trees can be used to classify suspicious transactions based on predefined criteria and features.
- *Random Forests:* An extension of decision trees, random forests combine the predictions of multiple decision trees to reduce variance and improve performance.
- *Support Vector Machines (SVM):* SVMs are effective for binary classification tasks, where they separate transactions into "normal" and "suspicious" categories by finding the optimal hyper plane in the feature space.
- *K-Nearest Neighbours (KNN):* KNN helps in classifying transactions by examining the "neighbourhood" of a given transaction, allowing for the identification of similar suspicious patterns.

3.2. FEATURE ENGINEERING

Effective feature engineering is critical to the success of the ensemble model. The following features are considered for detection:

- *Transaction Frequency and Volume:* Unusually high transaction volumes or abnormal frequencies can indicate suspicious behaviour.
- *Wallet Address Behaviour:* Patterns of wallet interactions and relationships with known illicit addresses.
- *Cross-Platform Activity:* Detection of transactions across multiple platforms or block chains, which could indicate layering or cross-chain laundering.
- *Transaction Velocity:* The speed at which funds are moved, with sudden large withdrawals or transfers being indicative of laundering efforts.

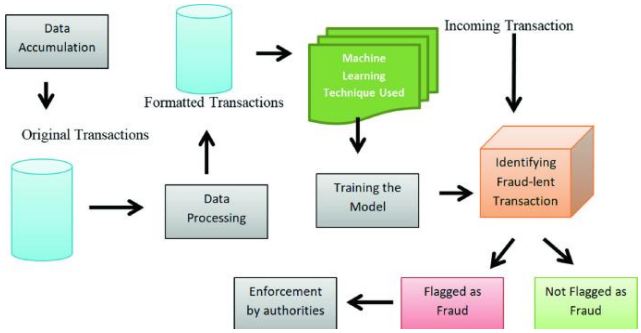
3.3. DATA COLLECTION AND PREPROCESSING

Data for this study is obtained from block chain transaction datasets, both real-world and synthetic, containing information on wallet addresses, transaction timestamps, and amounts. Pre-processing steps include:

- *Data Cleaning:* Removing inconsistencies or incomplete records.
- *Normalization:* Ensuring all features are on a consistent scale for model input.
- *Synthetic Data Generation:* Using blockchain simulation tools to generate labeled data representing illicit laundering activities.

4. EXPERIMENTAL SETUP AND EVALUATION

To set up an experimental framework for anti-money laundering (AML) in crypto currency, the process begins with collecting and pre-processing data from block chain sources, external intelligence (e.g., blacklists), and synthetic datasets simulating laundering activities. Data preparation involves cleaning, scaling, and anonymizing sensitive information while creating features such as transaction patterns, network structures, and temporal behaviours. The data is split into training, validation, and test sets, often using temporal splits to simulate real-world scenarios. Machine learning models, including supervised (e.g., Random Forest, XGBoost), unsupervised (e.g., Isolation Forest, Auto encoders), and graph-based models (e.g., GNNs), are trained on these features, with hyper parameter tuning via grid search or Bayesian optimization. The models are evaluated using precision, recall, F1-score, and AUC-ROC to handle the imbalance in detecting illicit transactions. Temporal validation ensures models generalize well over time, and interpretability tools like SHAP or LIME explain predictions. Comparative baselines include rule-based systems or heuristic graph techniques. Continuous monitoring ensures model performance remains robust, adapting to evolving laundering methods with regular retraining and audits, enabling deployment-ready systems aligned with regulatory standards.



4.1. MODEL TRAINING AND TESTING

Research articles on anti-money laundering (AML) in

crypto currency outline a systematic process for training and testing models to detect illicit activities. The training phase involves preparing block chain transaction data, often augmented with labels (e.g., suspicious or legitimate), derived from watch lists or manual annotation. Feature engineering is a crucial step, incorporating transaction-level attributes, graph-based metrics, and temporal patterns. Supervised models like Random Forest or XGBoost are trained on labeled datasets, while unsupervised models like Isolation Forest and Auto encoders learn normal transaction behaviors to identify anomalies. Graph-based models, such as Graph Neural Networks (GNNs), leverage the relational structure of block chain transactions for deeper insights. Handling imbalanced data through techniques like SMOTE ensures robust performance. During testing, models are evaluated using precision, recall, F1-score, and AUC-ROC to measure effectiveness. Temporal and cross-validation techniques are applied to test generalization and adaptability to real-world scenarios, and outputs are made interpretable through SHAP or LIME for compliance and usability. Baseline comparisons against rule-based systems and heuristic methods validate the model's superiority, while real-world simulations further ensure readiness for deployment.

4.2 STEPS IN TRAINING AND TESTING FOR ANTI-MONEY LAUNDERING IN CRYPTOCURRENCY

4.2.1 TRAINING PHASE

1. *Data Collection and Pre-processing:*
    - Gather data from block chain networks (e.g., Bit coin, Ethereum), exchanges, and regulatory watch lists.
    - Pre-process the data by normalizing transaction amounts, encoding categorical features, and handling missing values.
2. *Labelling:*
    - Label data as “legitimate” or “suspicious” using known laundering patterns, flagged wallets, or watch lists.
    - In cases of unsupervised learning, use unlabelled data for anomaly detection.
3. *Feature Engineering:*
    - Transaction-Based Features: Transaction size, frequency, and patterns.
    - Graph-Based Features: Node centrality, clustering coefficients, and wallet interactions.
    - Temp oral Features: Time intervals, sudden spikes in activity.
4. *Handling Imbalanced Data:*
    - Apply techniques like SMOTE (Synthetic Minority Oversampling), under sampling, or class weighting to balance datasets.
5. *Model Selection:*
    - Choose appropriate models based on the problem type:
        - Supervised: Random Forest, Logistic Regression, XGBoost.
        - Unsupervised: Isolation Forest, Auto encoders.
        - Graph-Based: Graph Neural Networks (GNNs), Node2Vec.
6. *Hyper parameter Tuning:*
    - Use grid search or Bayesian optimization to optimize model parameters based on validation performance.

4.2.2 Testing Phase

1. *Data Splitting:*
    - Split data into training (70–80%) and testing (20–30%) sets.
    - Use temporal splits for chronological evaluation to mimic real-world deployment.
2. *Evaluation Metrics:*
    - Precision: Accuracy of flagged suspicious transactions.
    - Recall: Percentage of actual suspicious transactions correctly identified.
    - F1-Score: Harmonic mean of precision and recall.
    - AUC-ROC: Evaluate the model’s ability to distinguish between legitimate and suspicious transactions.
    - False Positive Rate (FPR): Ensure manageable alerts for investigators.
3. *Validation Techniques:*
    - Apply k-fold cross-validation for robust performance estimates.
    - Use temporal validation for time-sensitive models.
4. *Baseline Comparison:*
    - Compare model performance against rule-based systems, random classifiers, or heuristic graph methods.
5. *Explain ability:*
    - Use tools like SHAP or LIME to interpret model outputs and understand feature importance.
6. *Real-World Simulation:*
    - Test the model on unseen or synthetic laundering scenarios (e.g., mixing, layering) to evaluate adaptability and robustness.

7. *Feedback Loop:*
    - Analyze misclassifications, retrain models with updated data, and refine features to adapt to evolving laundering techniques.

4.3. RESULTS

The ensemble approach significantly outperforms individual models in terms of detection accuracy, precision, and recall. The multi-layered approach reduced false positives and improved the model's ability to generalize across various laundering tactics. Random forests and SVMs contributed most to enhancing detection capabilities, while decision trees provided interpretability and transparency.

5. DISCUSSION

5.1. STRENGTHS OF THE APPROACH

- *Improved Detection Accuracy:* The ensemble model achieved a higher detection rate and lower false positives compared to single-model approaches.
- *Scalability:* The ensemble approach is capable of processing large volumes of data, making it suitable for real-time block chain transaction monitoring.
- *Adaptability:* The multi-layered approach can be fine-tuned to detect emerging money laundering techniques as they evolve.

5.2. LIMITATIONS

- *Computational Complexity:* Ensemble models require more computational resources compared to single algorithms, which may be challenging for real-time deployment in high-volume environments.
- *Data Labelling:* The accuracy of the model depends on the availability of labelled data, which can be scarce for crypto-linked money laundering activities.

6. CONCLUSION

This research demonstrates the effectiveness of a multi-layered ensemble approach for detecting crypto-linked money laundering. The combination of decision trees, random forests, and support vector machines enables high detection accuracy and reduced false positive rates. The approach provides a scalable, adaptive, and interpretable solution for tackling crypto-related financial crimes. Future work will focus on optimizing model efficiency, expanding the dataset to include more diverse laundering methods, and integrating the model with real-time block chain monitoring systems.

REFERENCES

1. U. W. Chohan, "The fatf in the global financial architecture: challenges and implications," 2019.

2. W. Firmansyah and H. T. Atmadja, "Juridical analysis awareness of profession advocacy to financial transaction reports and analysis centre (ppatk) during prevent and eradicate money laundering crime," Journal of Multidisciplinary Academic, vol. 5, no. 4, pp. 308–314, 2021.

3. R. Soltani, U. T. Nguyen, Y. Yang, M. Faghani, A. Yagoub, and A. An, "A new algorithm for money laundering detection based on structural similarity," in 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2016, pp. 1–7.

4. A. Salehi, M. Ghazanfari, and M. Fathian, "Data mining techniques for anti money laundering," International Journal of Applied Engineering Research, vol. 12, no. 20, pp. 10 084–10 094, 2017.

5. C. Alexandre and J. Balsa, "A multiagent based approach to money laundering detection and prevention." in ICAART (1), 2015, pp. 230–235.

6. D. Savage, Q. Wang, X. Zhang, P. Chou, and X. Yu, "Detection of money laundering groups: Supervised learning on small networks," in Workshops at the Thirty-First AAAI Conference on artificial intelligence, 2017.

7. G. Sobreira Leite, A. Bessa Albuquerque, and P. Rogerio Pinheiro, "Application of technological solutions in the fight against money laundering—a systematic literature review," Applied Sciences, vol. 9, no. 22, p. 4800, 2019.

8. S. N. F. S. M. Nazri, S. Zolkaflil, and N. Omar, "Mitigating financial leakages through effective money laundering investigation," Managerial Auditing Journal, 2019.

9. "Financial Crimes Enforcement Network. 2019. Application of FinCEN's Regulations to Certain Business Models Involving Convertible Virtual Currencies — FinCEN.gov.