INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Machine Learning in Money Laundering Detection over Blockchain Technology

Dr.K.Upendra Babu,P.Vamsi,B.Durga Shyam,C.Rajasekhara Reddy,B.Praveen
ASST. Professor (CSE), UG Scholar, UG Scholar, UG Scholar, UG Scholar
Department of Computer Science & Engineering
Bharath Institute of Science and Technology, BIHER
173, Agaram Road, Selaiyur, Tambaram, Chennai, Tamil Nadu, India

Abstract - Money laundering through cryptocurrency networks presents a growing challenge due to the inherent anonymity of blockchain transactions. Layering techniques using decentralized platforms obscure sender and receiver identities, complicating compliance and enforcement. To address this issue, a dual-layered approach combining Machine Learning (ML) and Value-driven Transactional Tracking Analytics for Crypto-compliance (VTAC) is proposed. This framework enables effective de-anonymization and classification of illicit transactions. The ML component includes data preprocessing, standard scaling normalization, model training with supervised classifiers, and hash address identification from transaction IDs. VTAC enhances detection by analyzing transaction frequency and behavioral anomalies. Three baseline classifiers—Random Forest, XGBoost, and AdaBoost—are evaluated. Further improvements are introduced using a hybrid ensemble of XGBoost with Random Forest, and advanced learners such as LightGBM and CatBoost. Performance is benchmarked using accuracy, precision, recall, F1-score, and confusion matrix on the Elliptic Bitcoin dataset. Results demonstrate superior detection rates with LightGBM. A web-based interface facilitates real-time illicit transaction prediction, hash traceability, and compliance reporting. This method contributes to enhancing blockchain forensics and regulatory oversight by integrating analytics with blockchain infrastructure

Keywords: Borderline SMOTE, class imbalance, credit card, fraud detection, sampling techniques, Tomek links.

I. INTRODUCTION

Cryptocurrencies have rapidly evolved into a mainstream financial medium, widely adopted for both lawful and unlawful purposes. Their decentralized architecture, cryptographic security, and user anonymity make them attractive not only to investors and businesses but also to malicious actors seeking to exploit financial systems without regulatory oversight [1]. Among the various cybercrimes enabled by digital currencies, money laundering has emerged as one of the most critical and challenging threats. The absence of centralized control and Know Your Customer (KYC) regulations in many crypto exchanges creates an ideal environment for laundering illicit funds [2].

Criminals utilize the anonymity and pseudonymity offered by blockchain networks to layer, disguise, and move funds obtained through illegal means. These funds are often transferred across multiple digital wallets, decentralized exchanges (DEXs), and privacy coins in an effort to erase transaction trails and avoid detection by financial watchdogs [3]. Cryptocurrencies like Bitcoin, Monero, and Zerocash are frequently used in dark web transactions involving drug trafficking, identity theft, arms deals, and various other illegal activities [4]. Privacy-centric cryptocurrencies, in particular, offer enhanced obfuscation techniques such as ring signatures, stealth addresses, and zero-knowledge proofs, making forensic blockchain analysis even more complex [5].

While blockchain itself is inherently secure due to its distributed ledger and cryptographic mechanisms, vulnerabilities still exist within the ecosystem. Attackers often exploit flaws in smart contracts, conduct phishing campaigns, and utilize deceptive malware or trojans to gain access to private keys and digital wallets [6]. Once private keys are compromised, cybercriminals can move digital assets undetected across multiple addresses, leveraging mixers, tumblers, and privacy coins to further obscure their tracks [7].

Regulatory agencies and financial institutions face growing difficulties in tracing such activities due to the pseudonymous nature of most blockchain transactions. Although every transaction is recorded publicly, the absence of real-world identity association poses a significant challenge. Traditional anti-money laundering (AML) tools fall short when applied to blockchain systems, especially in tracking complex, multi-layered crypto transactions [8].

RELATED WORK

Scharfman [9] emphasizes the pressing need for integrating anti-money laundering (AML) frameworks into cryptocurrency platforms to mitigate the exploitation of digital assets for illegal activities. He discusses the compliance challenges facing virtual asset service providers (VASPs) and presents practical methodologies for ensuring AML adherence within crypto exchanges. Scharfman identifies the structural gaps between traditional finance and blockchain-based assets, proposing a compliance-centric operational framework for cryptocurrency ecosystems. This framework includes customer due diligence (CDD), transaction monitoring, and suspicious activity reporting, aiming to harmonize crypto operations with global regulatory expectations.

Chang et al. [10] explored how blockchain technology is impacting financial services by conducting a comprehensive study involving interviews with subject-matter experts. Their findings highlight that while blockchain introduces transparency and security into financial systems, its decentralized and immutable nature poses major challenges for fraud detection and compliance enforcement. The study underlines the urgent requirement for industry-wide standardization and suggests a multi-stakeholder approach involving regulators, developers, and institutions to formulate AML guidelines tailored to decentralized environments. The authors argue that successful integration of blockchain in finance depends on balancing innovation with risk management.

Jullum et al. [11] proposed a machine learning-based solution for detecting money laundering transactions, focusing on identifying suspicious behavior in large-scale banking transaction data. Their work employed supervised learning algorithms trained on historical data labeled as either suspicious or non-suspicious. Through experiments, they demonstrated that machine learning models significantly outperform traditional rule-based systems in terms of accuracy and adaptability. The authors further emphasized the need for explainable models, especially in regulatory settings where auditability and transparency are paramount. Their approach serves as a foundation for applying similar methodologies in cryptocurrency transaction monitoring.

Gerbrands et al. [12] studied the effectiveness of AML policies through empirical network analysis. Their research aimed to quantify the impact of regulatory interventions by examining transaction networks across various jurisdictions. The study concluded that while AML policies have some measurable effects, they are often limited by inconsistent implementation and lack of global coordination. Gerbrands and colleagues suggested that enhanced information-sharing frameworks and cross-border regulatory partnerships are essential for robust AML enforcement. This research supports the idea that technical solutions must be accompanied by policy harmonization to be fully effective.

Serena, Ferretti, and D'Angelo [13] introduced a graph-based approach for analyzing cryptocurrency activities by modeling transaction histories as complex networks. Their work provided valuable insights into the structure and evolution of digital financial behavior. By applying network science principles, the authors could detect abnormal transaction patterns, identify central actors, and assess systemic risk. Their study supports the argument that blockchain's transparency can be leveraged for forensic investigation when coupled with advanced graph analytics. The research forms a key basis for designing automated systems capable of real-time risk scoring of wallets and transactions.

Pareja et al. [14] developed EvolveGCN, a novel method for processing dynamic graph data using evolving graph convolutional networks. Although originally applied in dynamic graph scenarios such as social networks, EvolveGCN has significant implications for transaction monitoring in blockchain, where the network structure evolves rapidly. Their work introduced a temporal component into GCNs, enabling the model to learn from both structural and temporal changes in graphs. This dynamic adaptation is particularly useful for money laundering detection, where transactions often follow time-sensitive patterns designed to obscure tracking.

Lo et al. [15] introduced Inspection-L, a self-supervised graph neural network (GNN) framework that generates node embeddings for detecting money laundering in Bitcoin. Their research utilized blockchain data to model financial activity as a graph and applied unsupervised learning to identify anomalous behavior without relying on labeled datasets. This method addresses the common issue of label scarcity in financial fraud detection by allowing the system to learn meaningful patterns from raw transaction data. The experimental results showed that Inspection-L outperformed other models in terms of precision and recall, demonstrating the potential of GNN-based approaches in cryptocurrency compliance.

Adewumi and Akinyelu [16] presented a comprehensive survey of machine learning and nature-inspired techniques for credit card fraud detection. Although focused on traditional financial systems, the methodologies discussed—such as artificial neural networks, genetic algorithms, and swarm intelligence—offer transferable insights for digital asset monitoring. The authors highlighted the importance of combining multiple models (ensemble learning) and adapting algorithms to evolving fraud techniques. Their analysis reinforces the need for hybrid models and continuous learning frameworks, which are increasingly relevant in the volatile and rapidly evolving landscape of cryptocurrency transactions.

MATERIALS AND METHODS

The proposed system aims to detect money laundering activities over blockchain by integrating advanced Machine Learning [17] techniques with Value-driven Transactional Tracking Analytics for Crypto-compliance (VTAC). The system begins with loading and preprocessing the Elliptic Bitcoin dataset, including handling missing values and converting non-numeric attributes. Data normalization is performed using Standard Scaling to prepare it for model training. Multiple machine learning [18] algorithms are trained to classify transactions as legal or illegal based on transaction behavior. A de-anonymization module maps transaction IDs to hash addresses, enabling identification of potential sources involved in illicit activities. The VTAC component strengthens detection by flagging transactions with unusually high frequency or suspicious value patterns. Furthermore, the system incorporates enhancements through hybrid ML models and state-of-the-art algorithms such as LightGBM and CatBoost. A user-friendly interface is provided to upload and process new transaction data, predict its legality, and retrieve related hash addresses for further blockchain-level investigation.

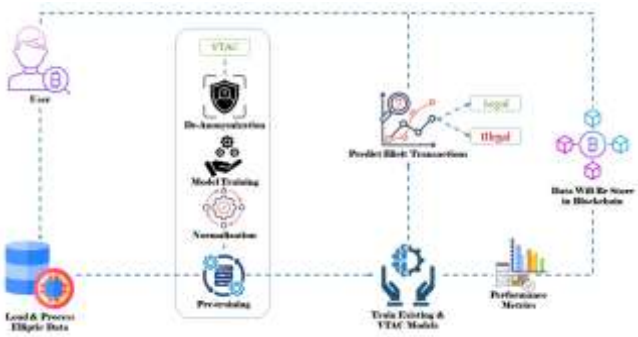


Fig. 1. System Architecture

The diagram illustrates a decision-tree-based model for detecting illegal cryptocurrency transactions. It processes crypto transactional data using multiple tree models to identify patterns. Features like wallet hashes, value, and transaction frequency guide prediction. The system determines potential illegal activity and outputs results for identification, enhancing accuracy in fraud detection and prevention.

A) Modules:

1. Signup: This module allows new users to register by entering necessary details such as username, email, and password. Upon successful registration, user information is stored securely in the database, enabling access to the system's features and ensuring personalized interaction for monitoring blockchain transactions and model predictions.

2. User Login: This module enables registered users to log in using their credentials. It verifies user authenticity by checking details against the database. Upon successful login, users gain access to various functionalities like data loading, model training, and transaction prediction, ensuring secure and controlled access to the system.

3. Load & Process Elliptic Dataset: This module facilitates uploading and preprocessing the Elliptic Bitcoin Dataset. It handles data cleaning, missing value replacement, and normalization. After processing, the data is split into training and testing sets, preparing it for machine learning [19] tasks and ensuring efficient data handling for accurate analysis.

4. Train Existing & VTAC Models: This module trains various machine learning [20] models using the processed dataset. It includes traditional, proposed, and extended models, both with and without VTAC integration. The module evaluates each model's performance using metrics like accuracy, precision, recall, and F-score to identify the most effective approach.

5. Predict Illicit Transactions: This module allows users to upload new transaction data for analysis. The trained model predicts whether each transaction is legal or illegal, identifies hash addresses, and displays transaction details. It helps in de-anonymizing suspicious activities and supports proactive detection of money laundering.

Logout: This module ends the user session securely, clearing access tokens or session data. It ensures that user credentials and activity data remain protected, preventing unauthorized access and maintaining system integrity after a user completes their interaction with the system.

B) Methods/ Algorithms:

Existing Random Forest: Random Forest is used for classifying transactions as legal or illegal by constructing multiple decision trees and combining their output. It handles imbalanced data well and reduces overfitting. Its ensemble nature improves accuracy by averaging predictions from different decision paths, making it suitable for detecting patterns in blockchain transaction behavior.

Existing XGBoost: XGBoost is applied for high-performance classification due to its ability to handle large datasets with speed and accuracy. It uses gradient boosting with regularization, reducing both bias and variance. It efficiently identifies suspicious transaction patterns by learning from data iteratively and optimizing for performance with fewer errors.

Existing AdaBoost: AdaBoost works by combining multiple weak learners to form a strong classifier. It adjusts weights after each iteration, focusing more on difficult transactions to classify. This makes it effective for identifying subtle differences between legal and illegal blockchain activity, though it may be sensitive to noise in data.

Propose VTAC Random Forest: This method enhances Random Forest with VTAC, which detects frequent high-value transfers. VTAC flags transactions based on volume and frequency, adding a rule-based filter before classification. It strengthens the accuracy of the model in recognizing repetitive laundering patterns and increases reliability in identifying high-risk transactions.

Propose VTAC XGBoost: Combining XGBoost with VTAC allows early filtering of potentially illegal transfers before classification. VTAC identifies suspicious frequency and value patterns, and XGBoost classifies them accurately. This integration significantly boosts detection power by aligning transaction characteristics with advanced model learning.

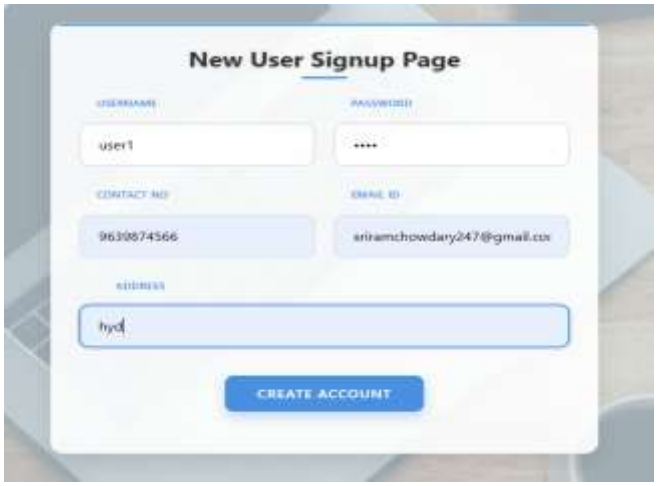
Propose VTAC AdaBoost: AdaBoost is enhanced with VTAC to prioritize suspicious transactions based on volume. VTAC pre-filters data, and AdaBoost focuses learning on high-risk records. This increases the model's sensitivity to laundering behavior and improves overall classification quality despite AdaBoost's typical sensitivity to noisy data.

Extension1 Hybrid Model: The hybrid model combines XGBoost and Random Forest to utilize the strengths of both—boosting accuracy and reducing variance. XGBoost's optimization and Random Forest's ensemble stability create a powerful classifier that outperforms individual models, especially for complex transaction patterns involving hidden laundering activities.

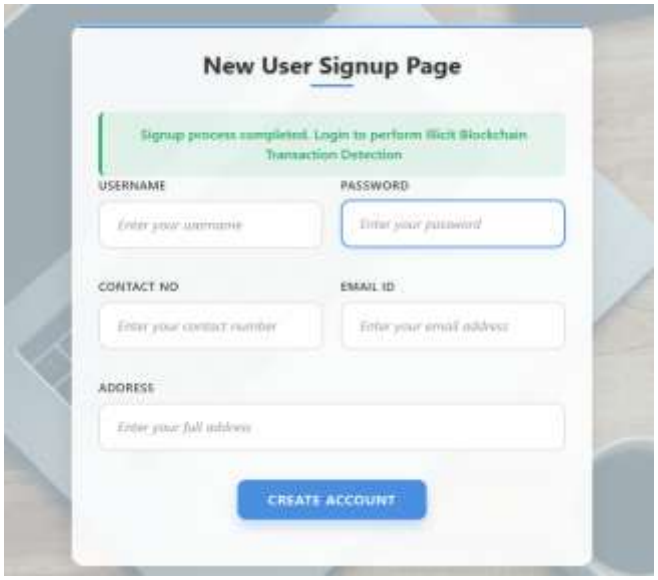
Extension2 LightGBM: LightGBM is a fast, high-performance gradient boosting algorithm that handles large-scale datasets efficiently. It uses leaf-wise tree growth, which improves accuracy and reduces training time. It achieved the best performance in detecting illicit transfers due to its ability to learn complex patterns and scale effectively.

IV. EXPERIMENTAL RESULTS

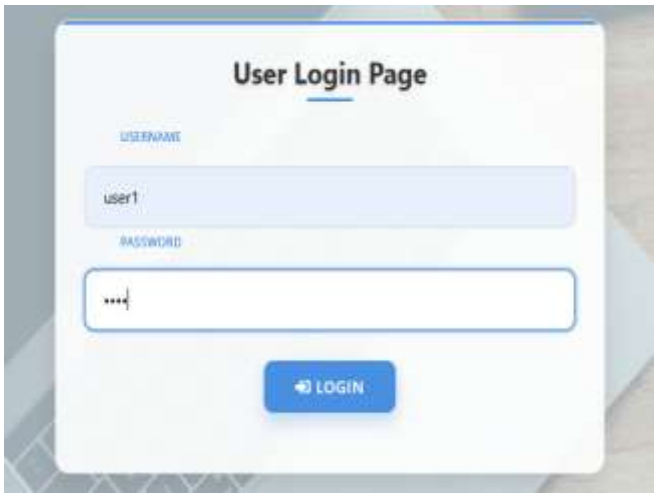
Subsequently, the project interface appears as illustrated in the above screen.



Enter the required new user information for the signup process displayed on the screen below.

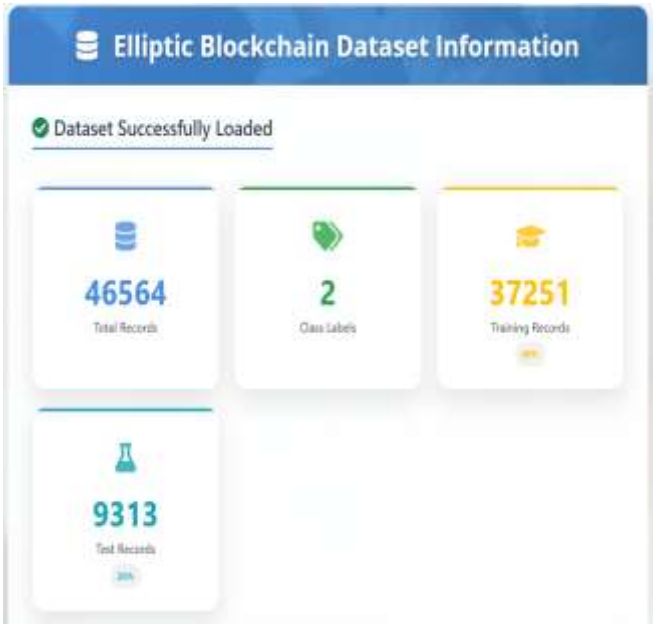


User registration has been successfully completed, and the details are securely recorded on the blockchain.

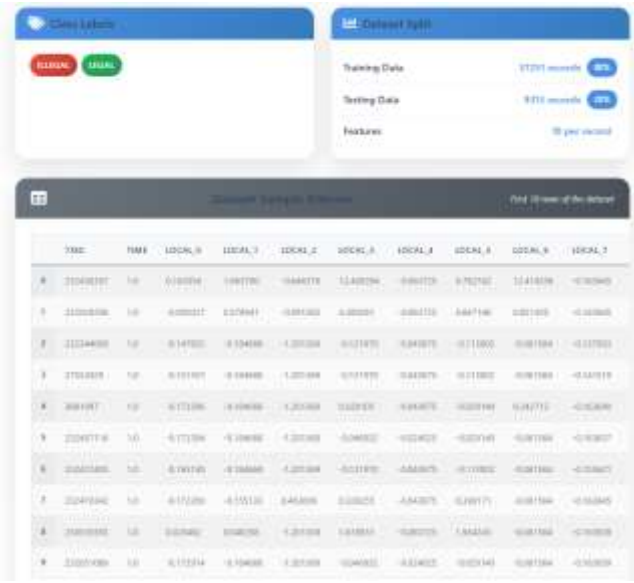


Users authenticate by entering their username and password into the designated input fields.

Upon successful login, the user is redirected to the home page shown above to continue the process.

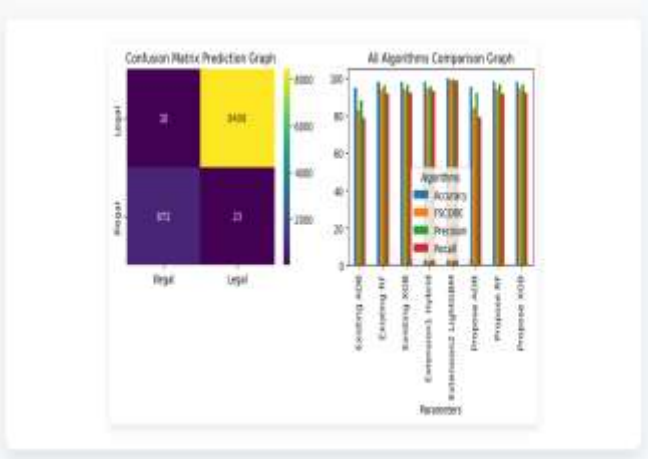


The Elliptic dataset is loaded, comprising two classes with a total of 46,564 records.

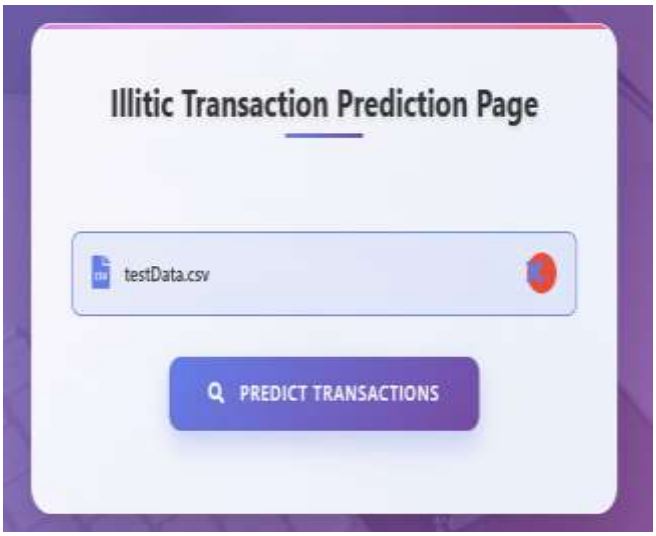


The dataset contains class labels along with train–test data splits and corresponding feature values for each column.

The performance metrics of all models are evaluated, and the best-performing model is identified below.



Presented below are the confusion matrix of the extension model and the performance metrics comparison graph for all implemented models.



Choose the test dataset from the test cases directory to test the model's performance.

The model's predictions for the test dataset, including the corresponding output classes, are presented below.

V. CONCLUSION

In conclusion, this work demonstrates an effective and scalable approach for detecting money laundering within blockchain ecosystems by combining advanced Machine Learning (ML) techniques with the Value-driven Transactional Tracking Analytics for Crypto-compliance (VTAC) framework. Utilizing the Elliptic Bitcoin dataset—a comprehensive, labeled graph dataset representing real-world cryptocurrency transactions—the system performs data normalization and preprocessing to enhance model performance. It implements and compares multiple classification algorithms, including standalone models like Random Forest,

XGBoost, AdaBoost, LightGBM, and CatBoost, along with a proposed hybrid ensemble model combining XGBoost and Random Forest. The integration of VTAC enables deeper transactional behavior analysis by incorporating frequency, directionality, and flow patterns of transactions, leading to better distinction between legal and suspicious activity. The de-anonymization of transaction IDs reveals high-risk hash addresses, aiding in traceability and compliance enforcement. Experimental results indicate that the hybrid model significantly improves detection accuracy and robustness, outperforming traditional classifiers in identifying illicit transactions. A web-based user interface further facilitates real-time prediction and performance visualization. Overall, the system provides a technically sound and practically viable solution to enhance blockchain transparency and support anti-money laundering (AML) efforts.

Future advancements can enhance this system by integrating real-time blockchain monitoring, expanding support to multiple cryptocurrencies beyond Bitcoin, and incorporating advanced deep learning techniques like Graph Neural Networks (GNNs) for improved pattern recognition. Enhancing the VTAC framework with dynamic behavioral profiling and anomaly detection can further increase detection precision. Additionally, collaboration with regulatory bodies and financial institutions could lead to the development of standardized tools for compliance, while improved de-anonymization techniques may aid law enforcement in tracing illicit financial networks more effectively.

REFERENCES

Ferretti, S., D'Angelo, G., & Ghini, V. (2025). Enhancing anti-money laundering frameworks: An application of graph neural networks in cryptocurrency transaction classification. IEEE Access.

Radhan, A., Sumith, N., & Srividya, S. (2025, February). Exploring the Effectiveness of Machine Learning Models in Detecting Anomalous Transactions. In 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE) (pp. 85-89). IEEE.

Li, G., Mi, Y., Zhou, J., Zheng, X., & Wu, W. (2025). Group Based Detection of Cryptocurrency Laundering Using Multi-Persona Analysis. IEEE Transactions on Information Forensics and Security.

Yu, Q., Xu, Z., & Ke, Z. (2024, November). Deep learning for cross-border transaction anomaly detection in anti-money laundering systems. In 2024 6th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) (pp. 244-248). IEEE.

Irshad, F., Alkhalifah, T., Alturise, F., & Khan, Y. D. (2024). GCF-MLD: integrated approach for money laundering detection using machine learning and graph network analysis. IEEE Access.

Wu, J., Liu, J., Zhao, Y., & Zheng, Z. (2021). Analysis of cryptocurrency transactions from a network perspective: An overview. Journal of Network and Computer Applications, 190, Article 103139. https://www.sciencedirect.com/science/article/pii/S1084804521001557

Marasi, S., & Ferretti, S. (2024, January). Anti-money laundering in cryptocurrencies through graph neural networks: A comparative study. In Proceedings of the IEEE 21st Consumer Communications & Networking Conference (CCNC) (pp. 272–277).

Rathore, M. M., Chaurasia, S., & Shukla, D. (2022, December). Mixers detection in Bitcoin network: A step towards detecting money laundering in crypto-currencies. In Proceedings of the IEEE International Conference on Big Data (Big Data) (pp. 5775–5782).

Scharfman, J. (2022). Anti-money laundering compliance for cryptocurrencies. In Cryptocurrency Compliance and Operations (pp. 91–114). Springer, Berlin, Germany.

Chang, V., Baudier, P., Zhang, H., Xu, Q., Zhang, J., & Arami, M. (2020). How blockchain can impact financial services—The overview, challenges and recommendations from expert interviewees. Technological Forecasting and Social Change, 158, Article 120166.

Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., & Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. Journal of Money Laundering Control, 23(1), 173–186.

Gerbrands, P., Unger, B., Getzner, M., & Ferwerda, J. (2022). The effect of anti-money laundering policies: An empirical network analysis. EPJ Data Science, 11(1), Article 15.

Serena, L., Ferretti, S., & D'Angelo, G. (2022). Cryptocurrencies activity as a complex network: Analysis of transactions graphs. Peer-to-Peer Networking and Applications, 15(2), 839–853.

Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T. B., & Leiserson, C. E. (2020, April). EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In Proceedings of the AAAI Conference on Artificial Intelligence, 34(4), 5363–5370.

Lo, W. W., Kulatilleke, G. K., Sarhan, M., Layeghy, S., & Portmann, M. (2023). Inspection-L: Self-supervised GNN node embeddings for money laundering detection in Bitcoin. Applied Intelligence, 53(16), 19406–19417.

Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), 937–953.

Popat, R. R., & Chaudhary, J. (2018, May). A survey on credit card fraud detection using machine learning. In Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1120–1125).

Sinayobye, J. O., Kiwanuka, F., & Kyanda, S. K. (2018, May). A state-of-the-art review of machine learning techniques for fraud detection research. In Proceedings of the IEEE/ACM Symposium on Software Engineering in Africa (SEiA) (pp. 11–19).

Mekterović, I., Brkić, L., & Baranović, M. (2018). A systematic review of data mining approaches to credit card fraud detection. WSEAS Transactions on Business and Economics, 15, 437–444.

Sadgali, I., Sael, N., & Benabbou, F. (2018). Detection of credit card fraud: State of art. International Journal of Computer Science and Network Security, 18(11), 76–83.