Journal of Business Crime

e-ISSN: 3090-4412
Vol 01 (2) 2025 p. 129-142

Corresponding author:
Kinza Aish
Email : kinzaaish611@gmail.com

Received 19 December 2025;
Accepted 5 January 2026;
Published 7 January 2026.

Conflict of interest statement:
Author(s) reported no conflict of interest

DOI: http://doi.org/10.70764/gdpu-jbc.2025.1(2)-11

A COMPREHENSIVE FRAMEWORK TO IDENTIFY AND PREVENT MONEY LAUNDERING IN DECENTRALIZED FINANCE USING BIG DATA ANALYTICS

Eva Harmelia Valentina[1], Kinza Aish[2]

[1] Universitas Islam Nahdlatul Ulama Jepara, Indonesia
[2] COMSATS University Islamabad, Pakistan

ABSTRACT

Objective: This research aims to develop a comprehensive framework to identify and prevent money laundering in Decentralized Finance (DeFi) by leveraging big data analytics, integrating advanced machine learning algorithms, and network analysis techniques to address the challenges of pseudonymity and decentralization inherent to this ecosystem.

Research Design & Methods: This research utilizes a mixed method approach with machine learning analysis based on Elliptic Dataset and qualitative policy study, applying graph models and classification algorithms to detect illegal transactions with precision in the context of imbalanced data.

Findings: The results show that the MLP and GCN models achieve high accuracy (98% and 97.3%) and excellent recall (99.5% and 99.4%) on the Elliptic Dataset, significantly outperforming traditional methods. Exploratory data analysis and graph visualization confirmed that illegal transactions form denser clusters and more complex paths, indicating a layering pattern.

Implications and Recommendations: Theoretically, this research extends the application of big data and graph theory to new financial systems, providing a blueprint for future RegTech and FinTech research. Practically, the framework offers tangible tools for regulators, law enforcement, and DeFi platforms to enhance AML capabilities, supporting the development of real-time monitoring tools and risk assessment models.

Contribution and Value Added: The main contribution of this research is the development of a robust and adaptive big data analytics-based AML framework, which effectively addresses the unique challenges of DeFi.

Keywords: Decentralized Finance, Anti-Money Laundering (AML), Machine Learning, Deep Learning, Elliptic

JEL codes: C55, G18
Article type: research paper

INTRODUCTION

Decentralized Finance (DeFi) has emerged as a transformative force in the global financial system, introducing a significant paradigm shift in the financial infrastructure (Sihombing et al., 2025). DeFi offers an unlicensed, open-source, and borderless alternative to traditional banking and

asset management. The sector's growth is rapid, with Total Value Locked (TVL) jumping from less than $1 billion to more than $100 billion between 2020 and 2024, underscoring its global reach and rapid adoption (John, 2025). This exponential growth is driven by open access, high yield potential through liquidity mining and staking, and composability (the interoperable protocols that make up the Lego money ecosystem). However, this same openness creates a double-edged sword, introducing new risks, especially in financial crime.

The pseudonymous nature of blockchain transactions, coupled with the lack of identifiable controls over many DeFi protocols, has attracted illicit financial activity and challenged global Anti-Money Laundering (AML) frameworks (Komal, 2024). The United Nations estimates that each year, 2-5% of global GDP, or around $800 billion to $2 trillion, is laundered globally, highlighting the enormous scale of the problem (Eifrem, 2019; Karim et al., 2024). Illegal actors are exploiting vulnerabilities in smart contracts, DEXs, liquidity pools, and synthetic assets for fraud, money laundering, and terrorist financing (Venčkauskas et al., 2025). Examples include brilliant contract exploits (such as the 2021 Poly Network exploit where more than $600 million was stolen), flash loan attacks that manipulate on-chain prices, the use of decentralized exchanges (DEXs) and mixers to obscure transaction traces, as well as scams such as rug pulls and DAO governance manipulation (Mo et al., 2023; Ren et al., 2025).

Traditional Anti-Money Laundering (AML) frameworks are not prepared to address decentralized and pseudonymous systems as they were designed with Traditional Finance (TradFi) in mind and are not easily transferable to the DeFi context, facing challenges such as borderless transactions, anonymity, and difficulty establishing jurisdiction or identifying responsible parties (Bakare et al., 2024). Regulators face regulatory blind spots due to jurisdictional ambiguity, the absence of identifiable counterparties, and liability gaps in smart contract code, allowing bad actors to quickly replicate systems and thwart AML protections (Szabo et al., 2024). Law enforcement remains reactive and limited by jurisdictional boundaries and difficulties in identifying responsible parties (Zetzsche et al., 2020). A significant research gap lies in developing AML frameworks that are intrinsically aligned with DeFi's decentralized architecture, rather than just an adaptation of TradFi methods, leading to the need for technological solutions such as on-chain analytics and embedded regulation (Zetzsche et al., 2020).

Therefore, this research aims to develop a comprehensive framework to identify and prevent money laundering in DeFi by leveraging big data analytics, integrating advanced machine learning algorithms and network analysis techniques to address the challenges of pseudonymity and decentralization (Udeh et al., 2024). Big data technologies are critical to addressing financial crime detection challenges, including data integration and quality, as well as real-time analytics, enabling proactive identification of suspicious patterns and anomalies (Udeh et al., 2024). By analyzing complex data patterns and deviations from normal behavior in real-time, the proposed framework can enable early detection, reduce losses, and prevent further fraud (Venčkauskas et al., 2025).

The research is expected to provide significant benefits to regulators and law enforcement by providing advanced technological tools to enforce the law and inform new regulatory approaches, to financial institutions and DeFi platforms to mitigate risk and ensure compliance, and to DeFi developers to drive integration of compliance tools and collaboration between innovators and regulators. In doing so, this research contributes to the legitimacy and integrity of the nascent DeFi ecosystem, ensuring its potential does not become a vector for criminal abuse. This report is structured to provide a comprehensive research plan, starting with the basic context of DeFi and AML challenges, delving into the existing literature, detailing the proposed methodological approach, outlining the expected results, discussing the implications, and concluding with key points and recommendations.

LITERATURE REVIEW

Basic Concepts of Decentralized Finance (DeFi)

DeFi is a peer-to-peer financial system that uses blockchain and cryptocurrencies for direct transactions without intermediaries such as banks, including applications such as lending platforms, decentralized exchanges (DEXs), liquidity pools, and synthetic assets, which are built on public blockchains such as Ethereum and often use self-executing smart contracts (Bakare et al., 2024; Manda et al., 2024). Contrary to centralized Traditional Finance (TradFi), DeFi operates permissionlessly, open-source, and borderlessly, eliminating central control and relying on automated protocols (Harvey and Rabetti, 2024; Salami, 2021). These architectural differences make traditional AML principles such as KYC difficult to implement, so AML solutions for DeFi must be intrinsically designed for decentralized environments, potentially through embedded regulation and on-chain analytics (Uzougbo et al., 2024).

Money Laundering in DeFi: Threat Vectors and Typologies

The pseudonymous nature of blockchain transactions and lack of identifiable controls over DeFi protocols challenge global AML frameworks, making transaction tracking difficult and blocking suspicious user accounts nearly impossible (Makarov and Schoar, 2022; Nadia, 2020). Illegal actors exploit vulnerabilities in smart contracts, DEXs, liquidity pools, and synthetic assets for fraud, money laundering, and terrorist financing (Ilijevski et al., 2023; Kirimhan, 2023). The main types include smart contract exploitation (e.g., reentrancy bugs, flash loans to steal assets), flash loan attacks that manipulate on-chain prices at millisecond speeds, money laundering through DEXs without KYC and the use of mixers and chain hopping to obscure transaction trails, rug pulls where fake projects disappear with investor funds, and DAO governance manipulation through token accumulation for malicious proposals (Alhaidari et al., 2024; Lin et al., 2024; Sechting and Raschke, 2024; Trozze et al., 2023; Wu & Tech, 2025). AML detection in DeFi requires sophisticated behavioural and network analysis to identify complex, multi-stage, and technology-specific typologies (Phyu and Uttama, 2023).

Table 1. Summary of Money Laundering Typologies in DeFi and Their Characteristics

Types of Money Laundering in DeFi	Description	Key Characteristics	Detection Challenges
Smart Contract Exploitation	Exploiting code vulnerabilities in smart contracts to steal or illegally transfer assets.	Code vulnerabilities, lack of formal audits, and complexity of composite systems.	Difficult to track due to automation and speed, developer accountability issues.
Flash Loan Attack	Large unsecured crypto asset loans, repaid in the same transaction block, are used for price manipulation or illegal arbitrage.	Transaction speed (milliseconds), automatic settlement, no guarantees.	Almost impossible to track accountability, and difficult to intervene in real time.
Money Laundering through DEXs and Mixers	The use of decentralized exchanges (DEX) without KYC/AML and mixing services to convert illegal assets into privacy coins or obscure the trail of transactions between blockchains.	No KYC/AML, peer-to-peer, "chain-hopping," wallet address anonymity.	Difficult to trace the origin of funds, impossible to block accounts centrally, and hidden transaction trails.

Types of Money Laundering in DeFi	Description	Key Characteristics	Detection Challenges
Rug Pulls and Exit Scams	Fake DeFi projects attract investor funds and then disappear with those assets.	Anonymous developers, promises of high returns, and no clear legal entity.	No entity is legally responsible, making it difficult to bring perpetrators to justice.
DAO Manipulation and Governance	Accumulation of governance tokens to drive harmful proposals (e.g., diverting treasury funds) in decentralized autonomous organizations (DAOs).	Token-based voting, the potential for accumulating cheap tokens, and the blurring of the line between criminality and legal loopholes.	Difficulty distinguishing between criminal activity and legitimate use of governance loopholes.

The Role of Big Data Analytics in Financial Crime Detection

Big data analytics is essential for detecting financial crime, overcoming integration challenges, ensuring quality, and real-time data analysis (Udeh et al., 2024). This involves collecting and pre-processing data from various sources (transactions, user behaviour, threat intelligence) to eliminate duplicates and standardize data, with anonymization for privacy. Machine learning algorithms such as MLP, GCN, XG Boost, Random Forest, and ADA Boost are used to develop predictive models that identify anomalies and suspicious patterns (Silva et al., 2023). Network analysis, particularly with Graph Neural Networks (GNN) and Self-Attention GNN, is essential for uncovering complex relationships within money laundering networks, enabling models to focus on the most relevant nodes and edges (Yu, 2024). Strategies for handling unbalanced datasets (e.g., oversampling) are also crucial for optimizing precision and recall, ensuring the detection of rare but high-impact illegal activities (Yu et al., 2025).

AML and Blockchain/DeFi Framework

The Financial Action Task Force (FATF) has emphasized the importance of robust KYC procedures and transaction monitoring for DeFi, although enforcement remains reactive and limited by jurisdiction (Case-Ruchala and Nance, 2024). Blockchain technology offers promising AML solutions through simplified KYC verification, real-time transaction monitoring with on-chain analysis (e.g., address clustering, "peel chain" detection), secure cross-institutional data sharing using permissioned blockchains and advanced cryptographic techniques (ZKPs, Homomorphic Encryption), and compliance automation through smart contracts (Ozcan, 2021). However, challenges remain, including jurisdictional ambiguity, conflicts between blockchain immutability and data privacy rights, scalability issues, high integration costs, and the risk of over-reliance on technology without human oversight.

METHODS

This study uses a mixed methods approach to examine money laundering detection in blockchain-based transactions and DeFi. The quantitative approach was conducted through big data analysis and machine learning, utilizing the Elliptic Dataset as the primary source, which includes the Elliptic Dataset (Bitcoin Blockchain) comprising 46,564 transactions (4,545 illegal, 42,019 legal); divided temporally. This dataset is divided temporally to simulate real-world predictions and avoid data leaks into the training set. On the qualitative side, the research analyzes policies and regulatory frameworks, law enforcement trends, and case studies as complements to examine contextual challenges. The research conducted by Khare and Srivastava (2023) is the main focus and secondary reference for this study. Data was processed through cleaning, transformation, normalization, and anonymization stages to maintain privacy. The analysis techniques include graph models (Graph Neural Networks and Self-Attention GNN) as well as classification algorithms such as MLP, XGBoost, Random Forest, and ADA Boost. The handling of unbalanced datasets was carried out using oversampling and node embedding strategies to maximize precision and recall for illegal

transactions. Model evaluation relied not only on accuracy, but also on precision, recall, and F1-score, given the importance of identifying rare but high-impact money laundering cases. This approach aims to build a practical predictive system that is sensitive to digital financial threats and relevant in the context of modern law enforcement.

Big Data Analysis and Machine Learning Techniques

GCN is designed for graph-structured data and consists of several layers of graph convolutions. GCN is very effective for analyzing transaction networks where relationships between entities are paramount. The Self-Attention-GNN model improves detection by dynamically adjusting feature aggregation, focusing on the most relevant nodes and edges in the transaction network. This is very useful for identifying complex layering patterns.

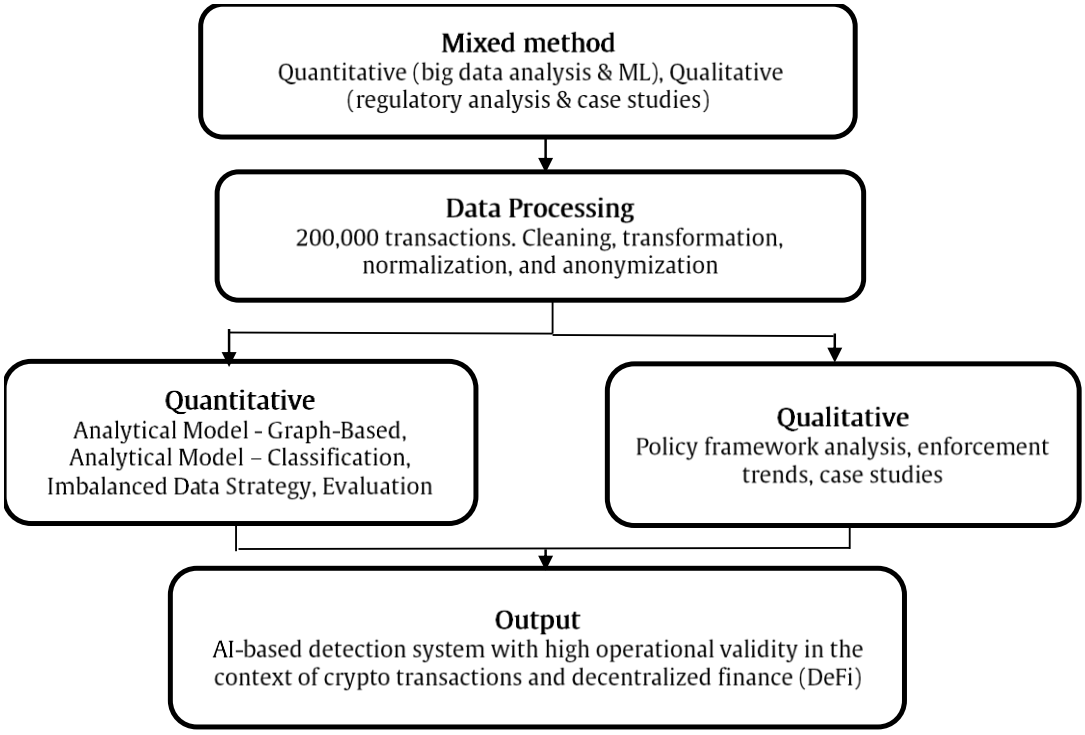

Figure 1. Research Framework for Detecting Financial Crime

RESULT

Exploratory Data Analysis of DeFi Transactions

This study utilizes the Elliptic Dataset, which contains more than 200,000 Bitcoin transactions, to map and analyze money laundering patterns using a directed acyclic graph (DAG) approach. In this representation, each node represents a Bitcoin transaction, while edges indicate the flow of funds between transactions. There are a total of 203,769 nodes and 234,355 edges in the network. Transactions are classified into three main categories: legal (42,019 transactions or 21%), illegal (4,545 transactions or 2%), and unknown (157,205 transactions or 77%).

Legal nodes tend to be spread out linearly with few cross-connections, reflecting normal and transparent transaction behaviour. In contrast, illegal nodes form a denser, interconnected network pattern, often showing branches and circles between transactions. This indicates a layering strategy–a technique commonly used in money laundering schemes where funds are disguised through a series of complex transactions to make them difficult to trace. Each node in the dataset has up to 166 features covering local characteristics (such as BTC volume, transaction fees, and time) and aggregate features (the amount of BTC received and spent within a certain time period). The dataset is also processed based on 49 time steps, each reflecting a block of transactions within a 3-hour period, to form a simulation of real transaction flows. This complex and unbalanced network

architecture poses a major challenge for machine learning algorithms while also making a significant contribution to the development of blockchain-based anti-money laundering (AML) systems.

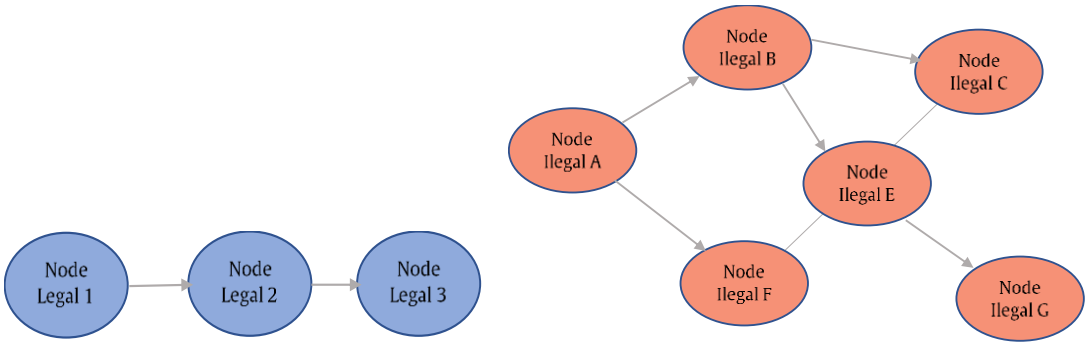

Figure 2. Conceptual Visualisation of Legal and Illegal Bitcoin Transaction Nodes in a Network Graph

Blue nodes represent legal transactions, while red nodes represent illegal transactions. Gray lines represent Bitcoin flows. Illegal nodes tend to form denser and more interconnected clusters, often with branching and looping paths, indicating layering efforts. Legal nodes tend to have more direct and isolated connections. The conceptual visualization of the Bitcoin transaction network above illustrates the complex data structure and striking differences between legal and illegal transactions. In this graph, nodes represent Bitcoin transactions, and edges indicate the movement of Bitcoin between transactions (Khare and Srivastava, 2023). From this visualization, it can be observed that illegal transactions (marked in red in the conceptual representation) often form denser clusters and more complex and interconnected paths than legal transactions (blue). This pattern clearly indicates attempts to layer or obscure funds, whereby criminals break up and move funds through multiple transactions to hide their origin (John, 2025).

Analysis of transaction volume distribution and asset types involved also reveals early anomalies and high-risk clusters. For example, illegal transactions may exhibit unusually high or low volumes, or unusual transfer patterns between unrelated addresses that deviate from normal behavior (Larik and Haider, 2011). Further temporal analysis reveals that illegal activity tends to spike at certain times or follow different cyclical patterns than legal transactions, providing additional insights for proactive detection.

The Performance of Machine Learning Models in Money Laundering Detection

One of the main challenges in developing data-driven money laundering detection systems is the extreme class imbalance in the dataset. Like many real-world financial crime datasets, the Elliptic Dataset also exhibits this characteristic, with only about 2% or 4,545 transactions classified as illegal out of a total of 46,564 transactions. Other studies have even noted over 223,000 labeled money laundering cases out of 180 million transactions, further emphasizing how minor illegal transactions are compared to legal transactions in this domain. This imbalance naturally tends to bias machine learning models toward the majority class, i.e., legal transactions. This could result in high overall accuracy but low recall performance for the illegal class, which is the primary focus in the context of anti-money laundering (AML).

In AML, classification errors in the form of false negatives–i.e., the failure to detect actual illegal transactions–have far more serious consequences than false positives, as they can allow criminals to escape detection. Therefore, strategies for handling imbalanced datasets are crucial, such as applying oversampling methods (e.g., SMOTE) to strengthen the representation of minority classes, utilizing advanced graph architectures like GNN, and enhancing effectiveness through node embedding techniques. These strategies not only help improve the model's sensitivity to suspicious

patterns but also optimize precision and recall, which are the primary evaluation metrics in an effective AML detection system.

Table 2. Comparison of AML Detection Algorithms

Algorithm	Main Architecture/ Approach	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Excellence in AML
MLP	Neural Network (Fully Connected Layer)	Elliptic Dataset	98.0	97.3	99.5	98.4	Complex feature representation, high performance on structured data.
GCN	Graph Convolution (Neighbourhood Aggregation)	Elliptic Dataset	97.3	97.5	99.4	98.4	Ideal for graph data (transaction networks), capturing relational dependencies.
XG Boost	*Gradient Boosting* (*Ensemble Learning*)	VTAC, Elliptic	97.5 (VTAC)	97.6 (VTAC)	97.4 (VTAC)	97.5 (VTAC)	Fast, efficient, scalable, strong regulation, high performance.
Random Forest	*Ensemble Learning* (*Decision Trees*)	VTAC	95.75	96.00	95.50	95.75	Handling high dimensional data, important feature insights, adaptive.
ADA Boost	*Ensemble Learning* (*Adaptive Boosting*)	VTAC	96.00	96.20	95.80	96.00	Improving predictive power on misclassified examples.
C4.5	*Decision Tree*	Bitcoin Mixers	>97 (specific)	N/A	N/A	N/A	Identify important features, high accuracy in certain cases.
KNN	*K-Nearest Neighbours*	Elliptic Dataset	92.0	97.0	97.0	97.0	Simple, effective for classification.

The implementation of the machine learning model on the test dataset showed very strong performance in detecting money laundering, effectively overcoming significant class imbalance in the data (only about 2% of transactions were illegal). The Multi-Layer Perceptron (MLP) model achieved an accuracy of 98%, precision of 97.3%, recall of 99.5%, and an F1-Score of 98.4% on the Elliptic dataset. Meanwhile, Graph Convolutional Networks (GCN) achieved an accuracy of 97.3%, precision of 97.5%, recall of 99.4%, and an F1-Score of 98.4% on the same dataset. The XG Boost

model, used in the VTAC approach, achieved an accuracy of 97.5%, with precision, recall, and F1-score consistently exceeding 95% (Doddamani et al., 2024; Venčkauskas et al., 2025).

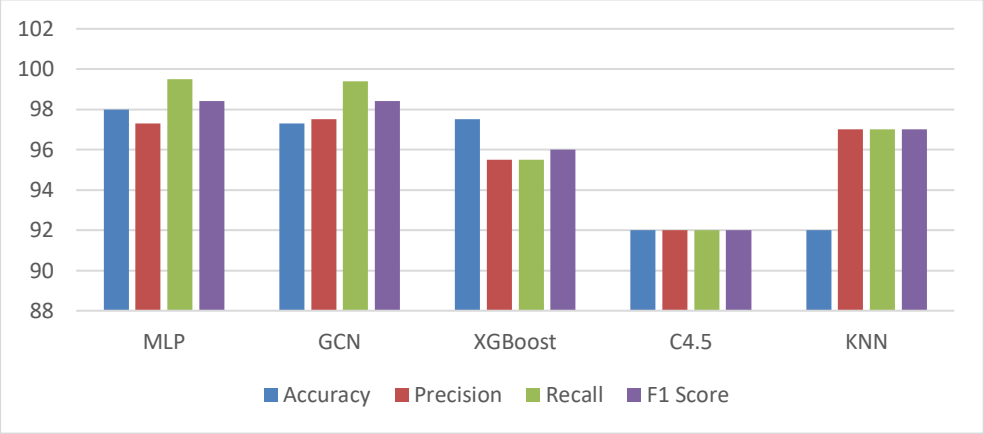

Figure 3. Performance Comparison of AML Detection Models

Comparison with basic methods or traditional rule-based systems shows a significant quantitative improvement. For example, the MLP and GCN models outperform other modern ML models such as C4.5 and KNN, which only achieve an accuracy of 92%.9 This shows that the proposed framework, by utilizing advanced machine learning algorithms, is able to achieve higher detection accuracy and effectively reduce false positives. The high recall values (99.5% for MLP and 99.4% for GCN) are particularly important in the AML context, as this means the models are highly effective in identifying the majority of actual money laundering transactions, thereby minimizing false negatives that could lead to significant financial and reputational losses (Al-Ababneh et al., 2024). Model robustness analysis also shows that imbalance mitigation strategies (such as oversampling or enhanced node embedding techniques) successfully optimize precision and recall, ensuring that the model is not biased towards the majority class and is able to detect rare but crucial illegal transactions (Cui et al., 2023).

Identification of Suspicious Patterns and Networks

Visualization and analysis of detected money laundering patterns, especially those identified by graph-based models (GCN, Self-Attention GNNs), confirm the framework's ability to uncover complex relationships that are difficult to detect using traditional methods. This model successfully identifies "peel chains," which are patterns of funds being gradually dispersed across multiple addresses to obscure their origin, as well as fund flows through mixers or suspicious addresses designed to hide transaction traces (Yu et al., 2025).

The effectiveness of the self-attention mechanism in GNN is remarkable. This mechanism allows the model to dynamically adjust feature aggregation, focusing on the most relevant nodes and edges in the transaction network, even in very dense and complex networks (Li et al., 2021). For example, the model can identify a series of small transactions that individually appear innocuous but collectively form a clear layering pattern when analyzed in the context of a broader network. This enables the detection of more subtle behavioral anomalies than simply transaction volume thresholds, providing deeper insights into money laundering modes in DeFi.

Case Study on the Application of Framework

As the adoption of decentralized finance (DeFi) increases, new forms of financial crime such as flash loan attacks and rug pulls are becoming increasingly complex and difficult to trace manually. One cutting-edge approach to addressing this challenge is to leverage graph-based models, which have proven effective in mapping complex crypto transactions and identifying hidden money laundering patterns. In a recent case study, a graph-based framework successfully detected money laundering activities involving flash loan attacks, where funds are transferred extremely quickly within a single transaction block, often accompanied by price manipulation

through oracles. Models like DELATOR and GAMLNet demonstrate superior capabilities in identifying key nodes and mapping illegal fund flows within the blockchain network (Assumpcao et al., 2022; Schmidt et al., 2024).

Additionally, in cases of rug pulls, where fake token developers withdraw liquidity and transfer funds to unrelated addresses, a graph-based investigation framework can visualize transactions and highlight suspicious patterns. Even seemingly simple money laundering techniques can be uncovered through smart contract analysis and fund flow visualization using open-source tools (Trozze et al., 2023).

Frameworks such as GCPAL and FlowScope also confirm that graphical representations of transactions can reduce reliance on manual labeling and improve the accuracy of detecting suspicious activity even in pseudonymous environments such as DeFi (Li et al., 2020; Lu & Wang, 2024). Thus, this framework not only offers powerful visualization and tracking capabilities, but also provides actionable digital evidence for law enforcement, contributing to the de-anonymization process in decentralized financial systems. Demonstrating the framework's ability to provide actionable insights for investigators, potentially leading to de-anonymization or identification of responsible parties, is also part of the outcome, highlighting its practical value in financial crime investigations (Udeh et al., 2024).

DISCUSSION

Interpretation of Findings: The Effectiveness of Big Data Analytics in Addressing DeFi AML Challenges

Empirical findings strongly confirm the hypothesis that big data analytics and advanced machine learning can effectively address the challenges posed by pseudonymity, borderless nature, and lack of centralized control in DeFi. The high performance of models such as MLP and GCN on the Elliptic Dataset, with exceptional accuracy and recall, demonstrates their ability to identify illegal transactions even in the face of significant class imbalance. Graph-based approaches (GNNs) specifically address the limitations of traditional models by capturing the complex relational dependencies inherent in money laundering networks, which are crucial for detecting the sophisticated layering patterns observed in data visualizations.

The success of big data analytics and machine learning in detecting money laundering in DeFi directly supports the concept of on-chain compliance and embedded regulation. Rather than relying on external and centralized enforcement, this framework demonstrates how compliance can be built directly into the design and monitoring of DeFi protocols. This implies a future where regulatory oversight is not an external burden but an inherent function of the decentralized system itself. This shift could revolutionize AML, making it more efficient and scalable in the face of rapidly evolving financial technology, potentially decentralizing both finance and its regulation.

Theoretical and Practical Implications of the Proposed Framework

Theoretically, this research makes a significant contribution to academic understanding of financial crime detection in decentralized environments, extending the application of big data and graph theory to new financial systems. It provides a blueprint for future research in RegTech and FinTech, paving the way for the development of more sophisticated and adaptive predictive models. The model's ability to identify complex patterns in pseudonymous blockchain data demonstrates the potential to develop new theories about financial crime behavior in decentralized ecosystems, transcending traditional intermediary-centric understandings.

The practical takeaway from this research is that it offers a real-world framework for regulators, law enforcement, and DeFi platforms to improve their AML capabilities. The results, which show high accuracy and the ability to identify suspicious networks in real time, are critical for the dynamic DeFi landscape. This framework can inform the design of future DeFi protocols to incorporate compliance from the outset, for example by integrating AI-based AML detection modules directly into smart contracts or protocol interfaces. It can also assist traditional financial

institutions in expanding their AML reach to digital assets and DeFi platforms, reducing cross-chain financial crime risks and enhancing investor confidence in the DeFi ecosystem as a whole.

Comparison with Existing AML Approaches

The proposed big data analytics framework offers significant advantages over traditional rule-based AML systems, which are often limited by predefined rules and difficulties with the volume and complexity of DeFi data. The accuracy and adaptability of machine learning models in identifying evolving illegal patterns are far superior to manual methods. This framework also addresses regulatory blind spots and jurisdictional inconsistencies faced by traditional jurisdiction-bound approaches (Aidoo et al., 2025). While traditional approaches may flag transactions based on simple thresholds, ML-based models can identify more subtle anomalies and complex network patterns indicative of money laundering, even when criminals attempt to obscure their tracks through various protocols or mixers. The ability to visualize and analyze transaction networks holistically, as shown in the results, is a key advantage that simple rule-based systems cannot offer, as they often fail to capture the interconnected nature of financial crime.

Limitations of Study and Future Research Directions

Data Limitations: This study acknowledges its reliance on publicly available datasets (e.g., Elliptic) that may not fully represent the entire DeFi ecosystem or all blockchain networks. While the Elliptic Dataset is the most comprehensive, it focuses on Bitcoin and may not fully capture the nuances of transactions on other blockchains or newer DeFi protocols, such as those involving stablecoins or synthetic assets. Future research could explore integrating data from more DeFi protocols and chains to improve model generalization, as well as consider larger and more diverse datasets that reflect the entire DeFi landscape.

Privacy vs. Transparency: The ongoing challenge of balancing blockchain transparency with data privacy requirements (e.g., GDPR right to be forgotten) needs to be thoroughly discussed. Future work may focus on further developing and integrating privacy-enhancing technologies such as Zero-Knowledge Proofs (ZKP) and homomorphic encryption into the framework, enabling compliance verification without revealing underlying sensitive data. This will be key to achieving a balance between effective regulatory oversight and the protection of users' privacy rights, which are fundamental pillars of the decentralization philosophy.

Scalability: The technical limitations of current DLT solutions and the need for advances in sharding or Layer-2 protocols to handle large-scale real-time DeFi transactions need to be addressed. Although the model shows good performance on existing datasets, full-scale implementation in the highly dynamic DeFi environment requires more advanced and efficient computing infrastructure to ensure real-time monitoring without significant bottlenecks or delays. Further research on algorithm optimization for streaming data processing is also needed.

The Evolving Threat Landscape: The typology of money laundering in DeFi continues to evolve rapidly as criminals find new ways to exploit technological innovations. Future research should focus on developing adaptive models that can learn and adjust to new illegal patterns autonomously, perhaps through continual learning or reinforcement learning, to remain relevant to the ever-changing tactics of criminals. This also includes research on how models can identify and adapt to criminals' use of new privacy technologies for illegal purposes.

Human Oversight and Ethical Considerations: Despite automation, the need for human judgment in interpreting complex legal regulations and the ethical implications of data collection and extensive surveillance in decentralized environments remains crucial. Future work could explore human-in-the-loop systems for AML, where AI serves as a powerful decision-support tool, but final decisions and legal interpretations remain in the hands of human experts. Additionally, further research is needed to address potential biases in the datasets used to train models, which could lead to unintended discrimination or unfair false positives, ensuring the fair and ethical implementation of this framework.

This section on limitations highlights the need for parallel development within legal and ethical frameworks. The conflict between the immutable nature of blockchain and the right to be forgotten is a prime example where technological capabilities outpace legal norms. Similarly, the effectiveness of machine learning models depends on data, but data collection in decentralized and pseudonymous environments raises significant ethical questions about privacy and potential misuse. Therefore, future research and implementation of such frameworks must be interdisciplinary, involving not only data scientists and financial experts but also legal scholars and ethicists to ensure that technological solutions are applied responsibly and in alignment with societal values and fundamental rights.

CONCLUSION

This research has comprehensively demonstrated the effectiveness of big data analytics, particularly graph-based machine learning models, in identifying complex money laundering patterns within the DeFi ecosystem. The proposed framework effectively addresses the inherent challenges of decentralization, pseudonymity, and the borderless nature of DeFi transactions. The theoretical contributions of this research are significant to the field of financial crime detection in emerging financial technology, and practically, it offers a robust, data-driven framework that can be adopted by regulators, financial institutions, and DeFi platforms to enhance AML compliance and reduce the risk of illegal financial activities. For the future, it is recommended to enhance international coordination and harmonization of DeFi regulations to prevent regulatory arbitrage, as well as explore embedded regulation and on-chain compliance as viable regulatory paradigms. Technically, sustained investment in advanced big data infrastructure and machine learning research, particularly in graph neural networks and privacy-preserving technologies, is strongly advised. Collaboration between compliance innovators, developers, and regulators should be encouraged to build a safer and more compliant DeFi system.

REFERENCES

Aidoo, S., Venditti, A., Döhner, H., Liang, W., Peterson, B., & Blessing, M. (2025). The Role of Blockchain in AML Compliance: Potential Applications and Limitations (Issue June). https://www.researchgate.net/publication/391627838

Al-Ababneh, H. A., Nuralieva, C., Usmanalieva, G., Kovalenko, M., & Fedorovych, B. (2024). The Use of Artificial Intelligence to Detect Suspicious Transactions in the Anti-Money Laundering System. *Theoretical and Practical Research in Economic Fields, 15*(4), 1039. https://doi.org/10.14505/tpref.v15.4(32).19

Alhaidari, A., Palanisamy, B., & Krishnamurthy, P. (2024). Protecting DeFi Platforms against Non-Price Flash Loan Attacks. *Proceedings of the Fifteenth ACM Conference on Data and Application Security and Privacy*, 281–292. https://doi.org/10.1145/3714393.3726503

Assumpcao, H. S., Souza, F., Campos, L. L., de Castro Pires, V. T., de Almeida, P. M. L., & Murai, F. (2022). DELATOR: Money Laundering Detection via Multi-Task Learning on Large Transaction Graphs. *2022 IEEE International Conference on Big Data (Big Data),* 709–714. https://doi.org/10.1109/BigData55660.2022.10021010

Bakare, F. A., Omojola, J., & Iwuh, A. C. (2024). Blockchain and decentralized finance (DEFI): Disrupting traditional banking and financial systems. *World Journal of Advanced Research and Reviews, 23*(3), 3075–3089. https://doi.org/10.30574/wjarr.2024.23.3.2968

Case-Ruchala, D., & Nance, M. (2024). The Limits of Enforcement in Global Financial Governance: Blacklisting in FATF as Rational Myth. *International Studies Quarterly, 68*(3). https://doi.org/10.1093/isq/sqae115

Cui, C., Wang, J., Wei, W., & Liang, J. (2023). Hybrid sampling-based contrastive learning for imbalanced node classification. *International Journal of Machine Learning and Cybernetics, 14*(3), 989–1001. https://doi.org/10.1007/s13042-022-01677-6

Doddamani, S. S., K, G. K., & Bhowmik, B. (2024). Money Laundering Detection in Imbalanced E-wallet Transactions with Threshold Optimization. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–6. https://doi.org/10.1109/I2CT61223.2024.10544197

Eifrem, E. (2019). How graph technology can map patterns to mitigate money-laundering risk. *Computer Fraud & Security, 2019*(10), 6–8. https://doi.org/10.1016/S1361-3723(19)30105-8

Harvey, C. R., & Rabetti, D. (2024). International business and decentralized finance. *Journal of International Business Studies, 55*(7), 840–863. https://doi.org/10.1057/s41267-024-00705-7

Ilijevski, I., Ilik, G., & Babanoski, K. (2023). Cryptocurrency Abuse for the Purposes of Money Laundering and Terrorism Financing: Policies and Practical Aspects in the European Union and North Macedonia. *ESI Preprints (European Scientific Journal, ESJ), 15*(23). https://doi.org/10.19044/esipreprint.3.2023p23

John, B. (2025). Decentralized Finance (DeFi) and Financial Crime : Emerging Threat Vectors, Regulatory Blind Spots, and the Need for a Global Governance Framework. *June*.

Karim, M. R., Hermsen, F., Chala, S. A., De Perthuis, P., & Mandal, A. (2024). Scalable Semi-Supervised Graph Learning Techniques for Anti Money Laundering. *IEEE Access, 12, 50012–50029*. https://doi.org/10.1109/ACCESS.2024.3383784

Khare, P., & Srivastava, S. (2023). Machine Learning for Anti-Money Laundering (AML): A Comprehensive Analysis. *International Research Journal of Engineering and Technology (IRJET), 10*(03), 982–989. www.irjet.net

Kirimhan, D. (2023). Importance of anti-money laundering regulations among prosumers for a cybersecure decentralized finance. *Journal of Business Research, 157*(March), 113558. https://doi.org/10.1016/j.jbusres.2022.113558

Komal, M. (2024). Decentralized Finance (DeFi): A Review of Applications and Risks in the Financial Ecosystem. *International Journal of Scientific Research in Engineering and Management, 08*(11), 1–7. https://doi.org/10.55041/IJSREM39195

Larik, A. S., & Haider, S. (2011). Clustering based anomalous transaction reporting. *Procedia Computer Science*, 3, 606–610. https://doi.org/10.1016/j.procs.2010.12.101

Li, M., Sun, M., Liu, Q., & Zhang, Y. (2021). Fraud Detection Based on Graph Neural Networks with Self-attention. *2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, 349–353. https://doi.org/10.1109/AINIT54228.2021.00075

Li, X., Liu, S., Li, Z., Han, X., Shi, C., Hooi, B., Huang, H., & Cheng, X. (2020). FlowScope: Spotting Money Laundering Based on Graphs. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(04), 4731–4738. https://doi.org/10.1609/aaai.v34i04.5906

Lin, Z., Chen, J., Wu, J., Zhang, W., Wang, Y., & Zheng, Z. (2024). CRPWarner: Warning the Risk of Contract-Related Rug Pull in DeFi Smart Contracts. *IEEE Transactions on Software Engineering, 50*(6), 1534–1547. https://doi.org/10.1109/TSE.2024.3392451

Lu, H., & Wang, H. (2024). Graph Contrastive Pre-training for Anti-money Laundering. *International Journal of Computational Intelligence Systems, 17*(1), 307. https://doi.org/10.1007/s44196-024-00720-4

Makarov, I., & Schoar, A. (2022). Cryptocurrencies and Decentralized Finance (DeFi). *Brookings Papers on Economic Activity, 2022*(1), 141–215. https://doi.org/10.1353/eca.2022.0014

Manda, V. K., Abukari, A. M., Gupta, V., & Bharathi, M. J. (2024). Revolutionizing Finance with Decentralized Finance (DeFi) (pp. 127–150). https://doi.org/10.4018/979-8-3693-1532-3.ch006

Mo, Y., Chen, J., Wang, Y., & Zheng, Z. (2023). Toward Automated Detecting Unanticipated Price Feed in Smart Contract. *Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis*, 1257–1268. https://doi.org/10.1145/3597926.3598133

Nadia, P. (2020). The Open Legal Challenges of Pursuing AML/CFT Accountability within Privacy-Enhanced IoM Ecosystems. *Proceedings of the 3rd Distributed Ledger Technology Workshop*, 1–15. https://hdl.handle.net/11585/758077

Ozcan, R. (2021). Decentralized Finance. In In: Hacioglu, U., Aksoy, T. (eds) Financial Ecosystem and Strategy in the Digital Era (pp. 57–75). *Springer, Cham*. https://doi.org/10.1007/978-3-030-72624-9_4

Phyu, T. H., & Uttama, S. (2023). Improving Classification Performance of Money Laundering Transactions Using Typological Features. *2023 7th International Conference on Information Technology (InCIT),* 520–525. https://doi.org/10.1109/InCIT60207.2023.10413155

Ren, S., He, L., Tu, T., Wu, D., Liu, J., Ren, K., & Chen, C. (2025). LookAhead: Preventing DeFi Attacks via Unveiling Adversarial Contracts. *Proceedings of the ACM on Software Engineering, 2*(FSE), 1847–1869. https://doi.org/10.1145/3729353

Salami, I. (2021). Challenges and Approaches to Regulating Decentralized Finance. *AJIL Unbound, 115*, 425–429. https://doi.org/10.1017/aju.2021.66

Schmidt, J., Pasadakis, D., Sathe, M., & Schenk, O. (2024). GAMLNet: a graph based framework for the detection of money laundering. *2024 11th IEEE Swiss Conference on Data Science (SDS),* 241–245. https://doi.org/10.1109/SDS60720.2024.00043

Sechting, C., & Raschke, P. (2024). A Taxonomy of Anti-Fraud Measures within Token Economy: Insights from Rug Pull Schemes. *2024 6th Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS),* 1–9. https://doi.org/10.1109/BRAINS63024.2024.10732222

Sihombing, M. A. Y., Stephanie, S., Khaw, B., Hendra, H., & Geovedi, F. (2025). Peran Decentralized Finance (DeFi) dalam meningkatkan inklusi keuangan di negara berkembang. *Hexatech: Jurnal Ilmiah Teknik, 4*(1), 1–11. https://doi.org/10.55904/hexatech.v4i1.1317

Silva, Í. D. G., Correia, L. H. A., & Maziero, E. G. (2023). Graph Neural Networks Applied to Money Laundering Detection in Intelligent Information Systems. *Proceedings of the XIX Brazilian Symposium on Information Systems*, 252–259. https://doi.org/10.1145/3592813.3592912

Szabo, J., Bernard, C., & Philip, L. (2024). Legal Implications and Challenges of Blockchain Technology and Smart Contracts. *Computer Life, 12*(2), 6–10. https://doi.org/10.54097/ztn2w848

Trozze, A., Davies, T., & Kleinberg, B. (2023). Of degens and defrauders: Using open-source investigative tools to investigate decentralized finance frauds and money laundering. *Forensic Science International: Digital Investigation, 46*(September), 301575. https://doi.org/10.1016/j.fsidi.2023.301575

Udeh, E. O., Amajuoyi, P., Adeusi, K. B., & Scott, A. O. (2024). The role of big data in detecting and preventing financial fraud in digital transactions. *World Journal of Advanced Research and Reviews, 22*(2), 1746–1760. https://doi.org/10.30574/wjarr.2024.22.2.1575

Uzougbo, N. S., Ikegwu, C. G., & Adewusi, A. O. (2024). Regulatory Frameworks for Decentralized Finance (DeFi): Challenges and opportunities. *GSC Advanced Research and Reviews, 19*(2), 116–129. https://doi.org/10.30574/gscarr.2024.19.2.0170

Venčkauskas, A., Grigaliūnas, Š., Pocius, L., Brūzgienė, R., & Romanovs, A. (2025). Machine Learning in Money Laundering Detection Over Blockchain Technology. *IEEE Access*, 13, 7555–7573. https://doi.org/10.1109/ACCESS.2024.3452003

Wu, K. W., & Tech, V. (2025). Strengthening DeFi Security: A Static Analysis Approach to Flash Loan Vulnerabilities. *In Computer Science Cornell University*. http://arxiv.org/abs/2411.01230

Yu, Q. (2024). Enhancing Anti-Money Laundering Systems Using Knowledge Graphs and Graph Neural Networks. *Advances in Economics, Management and Political Sciences, 118*(1), 280–288. https://doi.org/10.54254/2754-1169/2024.18627

Yu, Q., Wang, S., & Tao, Y. (2025). Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks. *SHS Web of Conferences, 213*(25), 01016. https://doi.org/10.1051/shsconf/202521301016

Zetzsche, D. A., Arner, D. W., & Buckley, R. P. (2020). Decentralized Finance. *Journal of Financial Regulation, 6*(2), 172–203. https://doi.org/10.1093/jfr/fjaa010