Research on Blockchain Attack Detection and Defense Mechanism Based on Machine Learning

Yifan Zhou

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China

2022214413@stu.cqupt.edu.cn

Abstract. Blockchain technology, while revolutionizing industries through decentralization, faces escalating security threats such as 51% attacks, smart contract vulnerabilities, and DDoS attacks. Traditional defense mechanisms, including consensus protocols (e.g., PoW/PoS) and rule-based detection tools, exhibit limitations, such as susceptibility to double-spending and high false-positive rates of 15–20%. This paper systematically investigates machine learning (ML)-driven approaches for blockchain attack detection and defense, evaluating their performance through a multidimensional "attack type-defense level" framework. Supervised learning models show robust performance: LightGBM achieves 99.17% accuracy in Ethereum fraud detection using hybrid sampling and feature engineering, outperforming XGBoost (97.99%) and Random Forest (98.26%). Deep learning methods, such as CodeBERT combined with graph structural analysis, achieve an 84.78% F1-score in detecting reentrancy vulnerabilities, surpassing static graph-based approaches with 53.44% accuracy. Federated learning frameworks with dynamic reputation mechanisms attain 99.1% DDoS detection accuracy while preserving data privacy. However, challenges include adversarial attack vulnerabilities, ambiguous global transaction patterns, and computational overheads limiting real-time deployment. Structural constraints in graph-based vulnerability analysis and trade-offs between privacy and efficiency further complicate applications. Future directions include adversarial defense reinforcement, lightweight edge-computing architectures, and SDN-integrated cooperative defense systems to balance detection accuracy, privacy, and scalability in evolving threat landscapes.

INTRODUCTION

Blockchain technology, as a typical representative of distributed ledger systems, has triggered changes in finance, supply chain, and other fields while continuing to face new security threats such as 51% attacks, smart contract vulnerability exploits (e.g., the re-entry attack model proposed by Crisostomo et al.), and DDoS attacks and other novel security threats [1-3]. Traditional protection mechanisms have significant limitations in dealing with these attacks, PoW/PoS-based consensus protocols are susceptible to double-spending attacks triggered by the concentration of computational power, and rule-matching vulnerability detection tools have a false positive rate of up to 15%-20% [4, 5].

In this context, machine learning-driven security protection techniques show breakthrough potential. Supervised learning optimizes attack identification by extracting transaction features (e.g., Kilic et al. achieved 97.1% blacklist prediction accuracy by training SVMs with Ethereum address PageRank values) [6]. Deep learning breaks through the semantic analysis bottleneck (Choi et al. fusion of CodeBERT with graph structure to detect reentrant vulnerabilities with 84.78% F1-score, while Wang et al., 2022 bytecode control flow graph-based graph convolutional network only achieves 53.44% accuracy in detecting timestamp-dependent vulnerabilities [7]). Federated learning further balances privacy and efficiency (Saveetha et al. designing a reputation federation framework to against DDoS attacks with 99.1% accuracy [1]. In addition, Kim et al. uses differential privacy federation learning to improve the accuracy of anti-poisoning attack models by 12% [8]). However, the field still suffers from the following constraints. For example, many models are not optimized for real-time scenarios in large-scale blockchain networks and need to be tuned to ensure efficient operation without sacrificing speed or accuracy.

The incompleteness and noise of real-world data pose challenges for vulnerability detection, privacy-preserving techniques introduce significant computational overhead in blockchain networks leading to increased computational complexity, traditional machine learning methods fall short in detecting DoS attacks, it is difficult to satisfy the financial-grade real-time demand [3, 5, 8, 9].

This paper aims to systematically compare the performance boundaries and application scenarios of different machine learning techniques by constructing a multi-dimensional framework of "attack type-defense level" to provide theoretical support for breaking through the bottleneck of blockchain security protection.

METHOD

Blockchain Infrastructure

Blockchain is a decentralized distributed ledger technology whose core architecture consists of a data layer, a network layer, a consensus layer, an incentive layer, a contract layer, and an application layer [10]. As shown in Figure 1, the data layer ensures data tamper ability through hash chain storage structure and asymmetric encryption algorithms (e.g., ECDSA), in which the block header contains the root of the Merkel tree and the hash value of the previous block, which forms a chain validation relationship, the network layer adopts the P2P protocol to achieve distributed communication between nodes, and synchronizes transactions and block data through the broadcast mechanism; the consensus layer adopts Proof of Work (PoW) , Proof of Workload (PoW), Proof of Stake (PoS) and other algorithms to achieve decentralized verification and ensure that all nodes agree on the state of the ledger [4]. The incentive layer drives nodes to participate in the maintenance of the network through the issuance of tokens and the distribution mechanism (e.g., Bitcoin mining incentives), and the contract layer embeds the smart contract as a programmable logic into the system to support automated execution of transaction rules (e.g., ethereum Solidity). (e.g., ethereum Solidity scripts); and the application layer covers practical scenarios such as financial payments, supply chain traceability, and digital identity.

Blockchain security relies on the collaborative work of all layers, data layer encryption and consensus layer authentication form the base protection, but the complex business logic of the application layer (e.g., the DeFi protocol), and the node collaboration of the network layer are still exposed to multiple attack risks. For example, intelligence vulnerabilities at the contract layer can be exploited by reentry attacks, while consensus layer arithmetic imbalances may trigger 51% attacks and network layer P2P communication is vulnerable to DDoS attacks [1]. These threats are directly related to the types of attacks and defense mechanisms discussed subsequently.

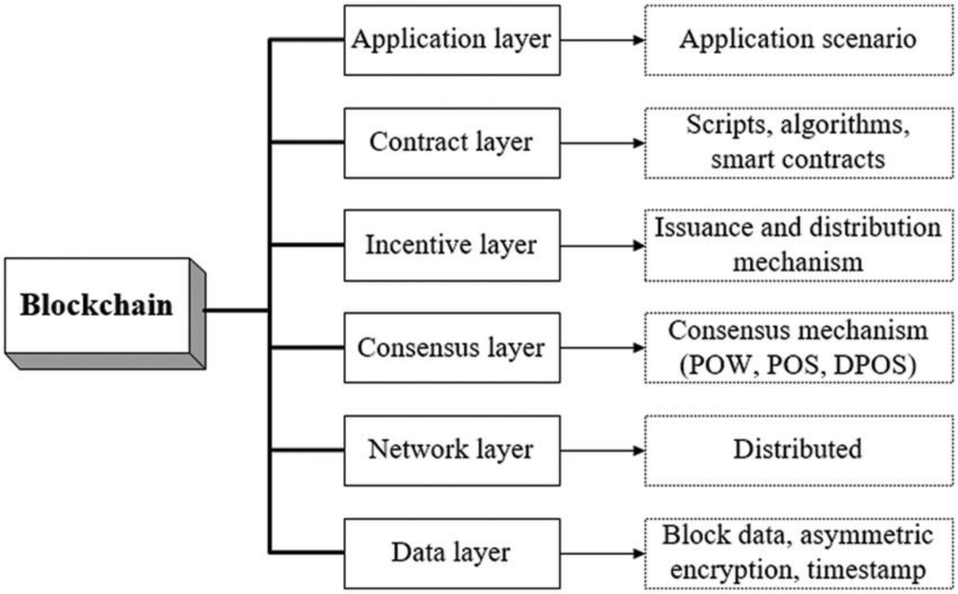
Blockchain
Application layer
Application scenario
Contract layer
Scripts, algorithms, smart contracts
Incentive layer
Issuance and distribution mechanism
Consensus layer
Consensus mechanism (POW, POS, DPOS)
Network layer
Distributed
Data layer
Block data, asymmetric encryption, timestamp

FIGURE 1. Blockchain infrastructure [11].

Attack Types and Machine Learning Models

Fraud Attack

Fraud attacks are one of the most prevalent security threats in blockchain networks, and their core tactics include forging transaction records, manipulating address identities (e.g., impersonating legitimate users), and utilizing transaction validation loopholes to implement Double-Spending and Sybil Attacks. Such attacks destroy the network trust mechanism through abnormal transaction patterns (e.g., high-frequency small transfers in a short period of time, abnormal address activity cycles), resulting in significant economic losses. For such attacks, traditional machine learning models are widely used due to their efficient feature learning capability and nonlinear pattern recognition advantages.

Ashfaq et al. proposed a hybrid machine learning framework based on the Random Forest and eXtreme Gradient Boosting (XGBoost) algorithms, which extracts transaction key metrics through multidimensional feature engineering, including network topology features (e.g., node outgoing/incoming degree averaging) and transaction pattern analysis (e.g., malicious transaction labeling) [12]. The study uses a publicly available dataset of 30,000 Bitcoin transactions, of which less than 1% are malicious samples, to solve the data imbalance problem by synthesizing minority class samples through Synthetic Minority Oversampling Technique (SMOTE), and finally the XGBoost model achieves a classification accuracy of about 90% on the test set.

Kilic et al. constructed a directed transaction graph containing 16.1 million nodes by collecting 141 million transaction data in the interval of 9,000,000 to 10,999,999 Ethereum block heights (November 2019 to October 2020), and combining public intelligence sources such as Etherscan and CryptoScamDB to 1,430 active blacklisted addresses are filtered to form the research dataset [6]. At the feature extraction level, the method integrates local and global features, local features cover direct behavioral metrics such as address access, transaction amount statistics (e.g., mean, extreme, total), and activetime; global features introduce PageRank algorithm to quantify node influence and map the global topological location of addresses through the connectivity subgraph identifier (con_comp_id). address's global topological location. Aiming at the class imbalance problem that the proportion of blacklisted addresses in the dataset is less than 0.01%, the study reconstructs the balanced training set using a hybrid strategy of random undersampling and SMOTE oversampling, and utilizes the extreme random tree to screen out the 10 key features, such as avg_out_amount, unique_indegree, pagerank, and so on, to input into the classification model.

The core contribution of lies in the improved (LGBM) LightGBM model, the optimization parameters are estimated via a Euclidean distance structure, and fraud classification is achieved by combining the transaction frequency and balance change features on an Ethereum dataset containing 4,000 tagged addresses [13]. The SMOTE oversampling technique is used to increase the fraud samples from 3,115 to 6,116, and finally achieves 99.17% test accuracy with 95.62% F1 score, which significantly outperforms comparative models such as XGBoost (97.99%) and Random Forest (98.26%).

Subsequently, Ravindranath et al. innovatively combined the K-Means-SMOTE oversampling technique with LightGBM to improve the model robustness by improving the sample distribution (the original fraud share of 22.14% was balanced to 50%) in a dataset of 6,000 fraudulent addresses [14]. Feature engineering focuses on traditional statistical indicators, ERC20 reception time difference, transaction amount average, address uniqueness, etc., and identifies key features such as “time difference between the first and the last transaction” through SHAP interpretive analysis.

DDoS Attack

DDoS attacks disrupt the availability of blockchain networks through distributed traffic flooding, protocol vulnerability exploitation, and smart contract resource exhaustion, and their detection requires a combination of network traffic characteristics and distributed defense mechanisms. Existing studies propose multi-layered defense strategies to deal with dynamic attack patterns through the co-optimization of blockchain architecture and machine learning models.

Banchhiwal et al. proposed a multilayered defense framework based on blockchain smart contracts and deep learning, which achieves dual protection of traffic filtering and anomaly detection through a layered architecture [3]. The framework redirects inbound traffic requests to the blockchain network, uses smart contracts for initial filtering (based on preset parameters such as request type, source address, packet size, etc.), and suspicious traffic enters the deep learning layer for further analysis. The blockchain storage layer uses encrypted chunks to store traffic data to ensure data integrity; the deep learning model identifies abnormal traffic patterns (e.g., TCP/UDP protocol anomalies, high-frequency request cycles) by analyzing features such as request types, resource access patterns, and embedded scripts. The experiments use real-time network traffic dataset (including core network and edge network nodes), and the robustness of the deep learning method under complex attack patterns is verified by comparing the performance of ANN and SVM models.

Saveetha et al. designed a detection framework integrating federated learning (FL) and blockchain to optimize the data quality of miner nodes through a dynamic reputation assessment mechanism and improve the robustness of the model against malicious traffic [1]. The study uses the CIC-DDoS2019 dataset (containing 12 types of attacks such as SYN, UDP, DNS, etc.) to construct a training set containing 80-dimensional traffic characteristics (e.g., target port, flow duration, and forward and backward packet statistics). In the federated learning framework, the miner nodes use the local traffic data to train Random Forest (RF), Multilayer Perceptron (MLP) with Logistic Regression (LR) models, and global model aggregation is achieved through the Flower framework. To cope with potential malicious node poisoning attacks in federated learning, a dynamic reputation evaluation mechanism is proposed, the reputation value is calculated based on the node's historical training accuracy (acc), pledge volume (data_stake) and the number of times of participation in training (training_count), and the high-reputation miners are screened to participate in training.

Defense Mechanisms and Deep Learning Models

Smart Contract Vulnerability Defense

The research in the field of smart contract vulnerability defense focuses on combining code semantic analysis and distributed architecture features to improve the comprehensiveness and dynamic adaptability of vulnerability detection through deep learning methods. By integrating natural language processing and graph structure analysis techniques, the existing research breaks through the limitations of traditional static rule detection and builds a hybrid detection framework that takes into account both code logic and execution paths, making significant progress in detection accuracy and interpretability.

Choi et al. proposed a smart contract vulnerability detection method that integrates large-scale language modeling (CodeBERT) and graph structural analysis and achieves collaborative analysis of code semantics and execution paths through a dual coding strategy [7]. The method firstly inputs the Solidity source code into CodeBERT model to generate 768-dimensional text embedding, and captures semantic features such as function naming, variable types, etc. Meanwhile, they constructed a code structure graph based on Abstract Syntax Tree (AST) and Control Flow Graph (CFG), and vectorize the sequence of operands by using Sent2Vec to extract critical path features through the triple analysis of Degree centrality, Katz centrality, and Proximity centrality. The critical path features are extracted by degree centrality, Katz centrality and proximity centrality triple analysis. The experiments use a dataset containing 30,000 Ethereum contracts and comparing CodeBERT detection alone (81.26% accuracy) with CodeBERT+AST combination (83.48% accuracy), the method achieves a combined accuracy of 86.70%, and the F1-score is improved to 84.46%.

Wang et al. proposed a bytecode-based graph convolutional network (GCN) detection framework to achieve vulnerability localization by constructing a control flow graph through reverse engineering [15]. The method first decompiles the smart contract bytecode into a sequence of opcodes, divides the base blocks by jump instructions and constructs a CFG, and uses a 3-layer GCN model for graph structure learning. The experiments use a dataset containing 1,420 contracts (472 with timestamp-dependent vulnerabilities), with adjacent matrix and unit matrix (not semantically processed) as node feature inputs, and nonlinear transformation by ReLU activation function.

Privacy and Attack Resistance

In the convergence of blockchain and machine learning, privacy protection and attack resistance are the core technical challenges. By combining differential privacy (DP), federated learning, and smart contracts, researchers have proposed a variety of defense mechanisms to cope with data leakage and malicious attacks. For example, Kim et al. designed a stochastic gradient descent method (DP-SGD) based on differential privacy for distributed machine learning (DML) scenarios in blockchain networks, which ensures that the data of a single participant cannot be inversely inferred by injecting Gaussian noise into the gradient update [8]. Meanwhile, its proposed error aggregation rule effectively defends against poisoning attacks by malicious nodes by filtering low error local gradients and filtering anomalous gradient paradigms.

Research in the other direction focuses on direct defense of machine learning models against blockchain attacks. For example, Latif et al. proposed a decentralized IoT security and privacy architecture based on the integration of blockchain and machine learning, aiming to address the core security issues such as single point of failure, data privacy leakage, and denial-of-service attacks in IoT networks [9]. The framework achieves device authentication through a dual registration mechanism between certification authorities and local nodes, uses blockchain to record access credential transactions to eliminate the risk of single point of failure, innovatively combines attribute-based encryption algorithms with blockchain to achieve key management and data privacy protection, and sets device resource thresholds through smart contracts to defend against DoS attacks.

DISCUSSION

Limitations and Challenges

The literature involved in this paper infers various limitations, and this paper divides it into three categories: detection accuracy, model robustness, and practical application, in order to clearly demonstrate its impact on the study.

Detection Accuracy Limit

- **Structural Representation Constraints：** Choi et al. noted that their graph centrality analysis struggled to effectively capture depth-related logical characteristics of function calls in CallDepth vulnerabilities, indicating limitations in handling layered execution dependencies through structural graph representations [7].
- **Semantic Feature Oversight：** Wang et al., the author acknowledged that their graph neural network approach had limitations in capturing semantic features of opcode nodes, as they did not integrate natural language processing for operand analysis, potentially constraining detection accuracy [15].

Model Robustness Limitations

- **Adversarial Attack Susceptibility：** The current state of research in blockchain fraud detection and related fields has made substantial progress, but several limitations and challenges remain. Ashfaq et al. identified a critical vulnerability in their machine learning-integrated blockchain fraud detection model, which is susceptible to adversarial attacks targeting ML classifiers [12].
- **Parameter Sensitivity：** Banchhiwal et al. highlighted that the accuracy of their deep learning-based detection model could be constrained by the need for cleaner datasets and further parameter refinement, limiting real-world adaptability when handling diverse attack patterns [3].
- **Systemic Deficiencies of Machine Learning in Dynamic Environments：** Traditional machine learning models rely on static datasets and fail to dynamically adapt to real-time evolving attack patterns in IoT environments, resulting in collinearity and data redundancy issues that degrade detection accuracy and real-time responsiveness [9].

Practical Application Limitations

- **Data Scale Dependency：** Aziz et al. noted that the LightGBM model requires significantly large datasets to perform effectively, posing a challenge when applied to small-scale Ethereum transaction data. This limitation restricts its adaptability in resource-constrained scenarios [13].
- **Computational Complexity:** Ravindranath et al. identified computational complexity as a key challenge in real-time fraud detection, particularly when processing high-frequency Ethereum transactions using ensemble models like LGBM and CATBoost [14].
- **Lack of Real-time Attack Mitigation Mechanisms:** Saveetha et al. made it clear that current research focuses on the attack detection level. Although the federated learning model stored through the blockchain

improves the detection reliability, it has not yet built a dynamic blocking system against the identified DDoS attacks [1].

Future Prospects

Based on the limitations and challenges identified in this study, the following are proposed future research directions to further enhance the effectiveness and applicability of machine learning in blockchain attack detection and defense mechanisms

- **Adversarial Defense Reinforcement:** Future work should prioritize enhancing the robustness of the integrated ML-blockchain framework against adversarial threats, such as data poisoning or evasion attacks. The authors explicitly state that mitigating these vulnerabilities is a key direction for further research, emphasizing the need to refine anomaly detection mechanisms to withstand sophisticated adversarial exploitation while maintaining blockchain's decentralized integrity [12].
- **Enhancing Dataset and Explainability:** To address the limitations of global features (e.g., connected component ID), in a study, the author proposed expanding the dataset size and integrating explainability algorithms to improve model transparency and reliability [6].
- **Computational Intelligence Optimization:** The integration of algorithms such as Elephant Herding Optimization (EHO), Monarch Butterfly Optimization (MBO), and Slime Mould Algorithm (SMA) could optimize feature selection and model robustness, particularly for data-intensive models like LightGBM [13].
- **Real-time Detection Frameworks:** To address computational bottlenecks in latency-sensitive environments, the study suggests developing optimized real-time detection systems that integrate lightweight model architectures with edge computing capabilities [14].
- **Data Refinement and Layered System Enhancement:** Future research should focus on integrating cleaner datasets and refining model parameters to improve detection accuracy, as noted in the paper's limitations. Additionally, advancing smart contract capabilities for real-time traffic analysis and expanding the blockchain-machine learning layered architecture could enhance adaptability to evolving DDoS attack patterns, while addressing computational efficiency constraints inherent in decentralized systems [3].
- **Cooperative Defense System Based on SDN:** Kilic et al. proposed extending the study to typical practical SDN network environment. It is suggested that the real-time traffic control module should be integrated into the software defined network architecture in the future [1].
- **Graph-based Vulnerability Analysis:** Advanced graph-based techniques (e.g., spectral analysis, community detection) should be explored to address depth-related vulnerabilities, while expanding dataset diversity improves generalizability [7].
- **Semantic Feature Extraction:** Enhancing semantic feature extraction through natural language processing of operand semantics and refining graph structure representations (e.g., differentiating edge types like conditional jumps) can improve model interpretability and detection accuracy [15].
- **Adaptive Models with Dynamic Threshold Optimization:** Developing reinforcement learning-based adaptive models that dynamically adjust feature weights and detection thresholds through real-time feedback mechanisms [9].

CONCLUSION

This study systematically compared different machine learning techniques and their application scenarios in blockchain security, highlighting both their strengths and limitations. While supervised learning models have demonstrated high accuracy in identifying fraud attacks, deep learning methods have made strides in semantic analysis and vulnerability detection. Federated learning has also shown promise in balancing privacy and efficiency. However, key challenges persist: models remain vulnerable to adversarial attacks and ambiguous transaction patterns, while data scale dependencies and computational bottlenecks limit real-world deployment. Future research should focus on three priorities: (1) adversarial defense frameworks combining blockchain-optimized attack simulations, (2) hybrid models integrating graph analysis with edge computing for real-time detection, and (3) adaptive privacy mechanisms coordinated with SDN-based defense orchestration. These aim to balance security, efficiency, and decentralization in evolving blockchain ecosystems. The continuous evolution of blockchain and machine learning technologies will undoubtedly provide new opportunities for improving security mechanisms, ensuring the integrity and reliability of decentralized systems.

REFERENCES

1. D. Saveetha, G. Maragatham, V. Ponnusamy, and N. Zdravković, “An Integrated Federated Machine Learning and Blockchain Framework With Optimal Miner Selection for Reliable DDOS Attack Detection,” IEEE Access, vol. 12, pp. 127903 – 127915, 2024.
2. J. Crisostomo, F. Bacao, and V. Lobo, “Machine learning methods for detecting smart contracts vulnerabilities within Ethereum blockchain — A review,” Expert Systems with Applications, vol. 268, p. 126353, Apr. 2025.
3. A. Banchhiwal, J. Bhardwaj, M. L. Sharma, V. K. Saini, and K. C. Tripathi, “DDoS Prevention on Distributed application using Blockchain Smart Contracts and Machine Learning,” The International Journal of Emerging Technologies and Innovative Research (JETIR), vol. 7, no. 4, 2020.
4. S. Kayikci and T. M. Khoshgoftaar, “Blockchain meets machine learning: a survey,” J Big Data, vol. 11, no. 1, p. 9, Jan. 2024.
5. M. Bresil, P. Prasad, M. S. Sayeed, and U. A. Bukar, “Deep Learning-Based Vulnerability Detection Solutions in Smart Contracts: A Comparative and Meta-Analysis of Existing Approaches,” IEEE Access, vol. 13, pp. 28894 – 28919, 2025.
6. B. Kilic, A. Sen, and C. Ozturan, “Fraud Detection in Blockchains using Machine Learning,” in 2022 Fourth International Conference on Blockchain Computing and Applications (BCCA), San Antonio, TX, USA: IEEE, Sep. 2022, pp. 214 – 218.
7. R.-Y. Choi, Y. Song, M. Jang, T. Kim, J. Ahn, and D.-H. Im, “Smart Contract Vulnerability Detection Using Large Language Models and Graph Structural Analysis,” Computers, Materials and Continua(CMC), vol. 83, no. 1, pp. 785 – 801, 2025.
8. H. Kim, S.-H. Kim, J. Y. Hwang, and C. Seo, “Efficient Privacy-Preserving Machine Learning for Blockchain Network,” IEEE Access, vol. 7, pp. 136481 – 136495, 2019.
9. S. Latif, M. S. B. Ilyas, A. Imran, H. A. Abosaq, A. Alzubaidi, and V. K. Jr., “Machine Learning Empowered Security and Privacy Architecture for IoT Networks with the Integration of Blockchain,” Intelligent Automation & Soft Computing(IASC), vol. 39, no. 2, pp. 353 – 379, 2024.
10. H. Taherdoost, “Blockchain and Machine Learning: A Critical Review on Security,” Information, vol. 14, no. 5, p. 295, May 2023.
11. Z. Xu and S. Cao, “Multi-Source Data Privacy Protection Method Based on Homomorphic Encryption and Blockchain,” Computer Modeling in Engineering & Sciences(CMES), vol. 136, no. 1, pp. 861 – 881, 2023.
12. T. Ashfaq et al., “A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism,” Sensors, vol. 22, no. 19, p. 7162, Sep. 2022.
13. R. M. Aziz, M. F. Baluch, S. Patel, and P. Kumar, “A Machine Learning based Approach to Detect the Ethereum Fraud Transactions with Limited Attributes,” Karbala International Journal of Modern Science, vol. 8, no. 2, pp. 139 – 151, May 2022.
14. V. Ravindranath, M. K. Nallakaruppan, M. L. Shri, B. Balusamy, and S. Bhattacharyya, “Evaluation of performance enhancement in Ethereum fraud detection using oversampling techniques,” Applied Soft Computing, vol. 161, p. 111698, Aug. 2024.
15. Z. Wang, W. Wu, C. Zeng, J. Yao, Y. Yang, and H. Xu, “Smart Contract Vulnerability Detection for Educational Blockchain Based on Graph Neural Networks,” in 2022 International Conference on Intelligent Education and Intelligent Research (IEIR), Wuhan, China: IEEE, Dec. 2022, pp. 8 – 14.