Blockchain Node Classification Predicting Node Behavior Using Machine Learning

Agung Budi Prasetio[1,*], Ono Widodo Purbo[2]

[1,2]Faculty Computer Science, Institut Teknologi Tangerang Selatan, Komplek Komersial BSD Kav 9, Jl. Raya Serpong, Lengkong Karya, Serpong Utara, Kota Tangerang Selatan 12246, Indonesia

ABSTRACT

Blockchain technology has emerged as a secure and decentralized framework for digital transactions; however, its open and pseudonymous nature also presents significant challenges related to fraudulent activities and malicious nodes. This study investigates the application of machine learning models for blockchain node classification and fraud detection, evaluating three models: Random Forest, XGBoost, and Neural Network. The research leverages a dataset of 10,000 blockchain transactions with 16 attributes, including transaction fees, block scores, stake distribution rates, and coinage. The results demonstrate that machine learning models can effectively classify blockchain nodes with high accuracy. Among the evaluated models, the Neural Network classifier outperformed the others, achieving an accuracy of 95.3%, precision of 95.1%, recall of 95.6%, and an F1-score of 95.3%. Comparatively, XGBoost achieved an accuracy of 94.1%, while Random Forest scored 92.4%. Feature importance analysis highlighted Block Score (0.38), Transaction Fee (ETH) (0.30), and Stake Distribution Rate (0.15) as the most significant factors influencing classification outcomes. Furthermore, confusion matrix analysis revealed that the Neural Network model produced 4780 true positives and 4440 true negatives, with only 200 false positives and 580 false negatives, demonstrating its robustness in identifying fraudulent nodes. Despite these promising results, real-world deployment presents several challenges, including the evolving nature of fraudulent strategies, real-time detection requirements, and scalability concerns. Future research should explore real-time learning techniques, integration of network-based features, decentralized fraud detection mechanisms, and cross-chain anomaly detection to improve model adaptability and effectiveness. By advancing these methods, machine learning-driven fraud detection can contribute to a safer, more transparent, and resilient blockchain ecosystem.

Keywords Blockchain Security, Machine Learning, Fraud Detection, Node Classification, Neural Networks, Anomaly Detection

Submitted: 25 Januari 2025
Accepted: 30 Mei 2025
Published: 1 September 2025

Corresponding author
Agung Budi Prasetio,
agung@itts.ac.id

Additional Information and Declarations can be found on page 202

DOI: 10.47738/jcrb.v2i3.42

INTRODUCTION

Blockchain technology has emerged as a transformative innovation, enabling secure, decentralized, and transparent transactions without the need for intermediaries [1]. As a distributed ledger system, blockchain records transactions in an immutable and tamper-resistant manner, ensuring trust and security in various applications such as cryptocurrencies, smart contracts, supply chain management, and Decentralized Finance (DeFi) [2]. However, despite its advantages, blockchain networks remain vulnerable to fraudulent activities, as malicious entities continuously attempt to exploit weaknesses within the system [3]. Various attack strategies, including double-spending, Sybil attacks, and transaction laundering, pose significant threats to blockchain integrity, making fraud detection a crucial area of research. The ability to classify blockchain nodes as either legitimate or fraudulent is essential for maintaining the security and stability of decentralized networks [4]. Traditional fraud detection techniques in blockchain often rely on rule-based systems, heuristic-

driven approaches, and manual auditing [5]. While these methods can be effective in identifying certain fraudulent behaviors, they struggle to adapt to the evolving tactics used by adversarial actors. Additionally, static rule-based systems often generate a high number of false positives and require frequent updates to remain effective. As blockchain transactions increase in volume and complexity, automated and scalable fraud detection solutions become necessary to ensure network security [6]. Machine learning and Artificial Intelligence (AI) offer promising alternatives by enabling systems to learn from transactional data, identify anomalies, and detect fraudulent patterns accurately [7]. These techniques have been widely applied in traditional financial fraud detection and are now being explored in the context of blockchain security. This study explores the application of machine learning models for blockchain node classification and fraud detection. Specifically, it evaluates the performance of three machine learning models (Random Forest, XGBoost, and Neural Network) on a dataset comprising 10,000 blockchain transactions with 16 attributes, including transaction fees, block scores, stake distribution rates, and coinage. This research aims to determine the effectiveness of these models in classifying blockchain nodes as legitimate or fraudulent, while also identifying the key transactional attributes that influence classification outcomes. The study also seeks to compare the performance of the three models based on key evaluation metrics such as accuracy, precision, recall, and F1-score.

The findings from this study demonstrate that machine learning models can be effectively utilized for blockchain fraud detection, achieving high classification accuracy and reliability. Among the evaluated models, the Neural Network classifier exhibited the best performance, achieving an accuracy of 95.3%, outperforming XGBoost with an accuracy of 94.1% and Random Forest with an accuracy of 92.4%. Feature importance analysis highlights that Block Score, Transaction Fee (ETH), and Stake Distribution Rate are among the most influential factors in distinguishing between legitimate and fraudulent nodes. Additionally, confusion matrix analysis indicates that the Neural Network model correctly identified 4780 fraudulent transactions and 4440 legitimate transactions, while producing only 200 false positives and 580 false negatives, demonstrating its robustness in fraud detection. Despite these promising results, several challenges remain in implementing machine learning-based fraud detection in real-world blockchain environments. The dynamic nature of blockchain transactions means that fraudulent actors continuously develop new strategies to evade detection, making static classification models less effective over time. Furthermore, achieving real-time fraud detection is a significant challenge due to the high volume of transactions processed on blockchain networks. Many existing models require batch processing and periodic retraining, limiting their ability to adapt to emerging fraud patterns in real time. To address these challenges, this study emphasizes the need for future research to focus on adaptive learning models that enable real-time fraud detection and continuous model updates, ensuring that classification models remain effective against evolving threats. Incorporating network-based features such as transaction propagation time, peer connectivity metrics, and graph-based relationships could further enhance classification accuracy by capturing complex transaction behaviors. Additionally, integrating explainable AI (XAI) techniques would improve model interpretability, making AI-driven fraud detection systems more transparent and trustworthy for regulators and security analysts. Furthermore, expanding fraud detection capabilities to support cross-

chain anomaly detection would help identify fraudulent behaviors that span multiple blockchain networks, as malicious actors often operate across different ecosystems to obscure their activities.

By leveraging machine learning and artificial intelligence, blockchain security can be significantly strengthened, making fraud detection more efficient, scalable, and adaptive. This study lays the groundwork for future advancements in automated fraud detection, with the ultimate goal of developing robust, real-time, and explainable security solutions for decentralized networks.

Literature Review

Blockchain security and fraud detection have gained significant attention in recent years, leading to various studies that apply machine learning and deep learning techniques to mitigate fraudulent activities in decentralized networks. Numerous research efforts have explored different methodologies, including supervised, unsupervised, and semi-supervised learning, to enhance the accuracy and scalability of fraud detection mechanisms. Chen et al. [8] proposed a Random Forest-based anomaly detection model for Ethereum transactions. Their study analyzed over 100,000 blockchain transactions, demonstrating that decision-tree-based models achieved high precision in identifying suspicious transactions. However, their approach suffered from a high false positive rate due to the static nature of rule-based classification, which limited its effectiveness in handling evolving fraudulent behaviors. Zhang et al. [9] developed a Graph Neural Network (GNN) model for fraud detection by leveraging transaction graph structures. Their research highlighted the importance of network-based features in improving classification accuracy. By incorporating graph embeddings and node connectivity metrics, they achieved a 10% increase in fraud detection accuracy compared to traditional supervised learning models. Umer et al. [10] implemented a deep learning-based fraud detection model using Long Short-Term Memory (LSTM) networks to capture temporal dependencies in blockchain transactions. Their approach successfully identified anomalous transactions in real time, but computational efficiency and scalability remained a challenge due to the high resource consumption of LSTM networks when processing large-scale blockchain data.

Kim et al. [11] introduced an Autoencoder-based anomaly detection system for identifying fraudulent transactions in Bitcoin and Ethereum networks. Their unsupervised learning approach detected hidden patterns in transaction behaviors and identified outliers without requiring labeled datasets. While this method showed promise in detecting new fraud patterns, it was susceptible to false positives and required extensive parameter tuning to balance detection sensitivity. Ileberi and Sun [12] proposed a Hybrid Machine Learning Model combining XGBoost and Neural Networks for fraud detection in blockchain networks. Their study demonstrated that ensemble learning approaches combining rule-based decision trees and deep learning architectures could significantly reduce false positives while improving detection accuracy. However, their method required a large amount of labeled training data, limiting its applicability in real-world blockchain environments where fraudulent activities continuously evolve. Khan et al. [13] investigated federated learning approaches for fraud detection in decentralized networks. Their study proposed a privacy-preserving fraud detection system that enabled collaborative learning across multiple blockchain nodes without sharing sensitive transaction data. This approach improved fraud detection rates across different blockchain

ecosystems but introduced challenges related to communication overhead and model synchronization across nodes. Sharma et al. [14] explored time-series anomaly detection techniques to identify fraudulent transaction patterns in blockchain networks. Their study focused on detecting unusual activity spikes in Ethereum transactions and demonstrated that dynamic thresholding methods could outperform static rule-based detection systems. However, their approach required continuous recalibration to maintain accuracy.

Wu et al. [15] examined network-based fraud detection by applying graph analytics and clustering techniques to detect abnormal transaction flows in blockchain networks. Their research showed that fraudulent entities often exhibit distinguishable network structures, which can be effectively detected using graph clustering algorithms. Despite the effectiveness of this approach, it was computationally expensive and required real-time graph updates for deployment in live blockchain environments. Kapadiya et al. [16] developed a hybrid blockchain fraud detection system that integrated symbolic AI reasoning with machine learning classifiers to improve explainability. Their research highlighted that black-box deep learning models struggle with regulatory compliance due to their lack of interpretability. Their approach demonstrated that combining explainable rule-based systems with machine learning models resulted in improved trustworthiness and accountability in fraud detection models.

While previous research has explored various machine learning techniques for fraud detection in blockchain networks, this study builds upon existing efforts by conducting a comprehensive evaluation of multiple classification models (Random Forest, XGBoost, and Neural Networks) on a dataset comprising 10,000 blockchain transactions with 16 attributes. Unlike prior studies that primarily focused on transaction classification, our research also examines feature importance to determine the most critical factors influencing fraud detection accuracy. By analyzing how attributes such as Block Score, Transaction Fee (ETH), and Stake Distribution Rate contribute to fraudulent transaction identification, this study provides a more detailed understanding of the factors that drive blockchain security risks. Additionally, this study evaluates the feasibility of deep learning architectures in blockchain security while proposing enhancements such as real-time fraud detection, adaptive learning models, and explainable AI techniques. In contrast to existing works, this research aims to bridge the gap between classification accuracy and practical implementation, ensuring that fraud detection models can be effectively deployed in real-world blockchain ecosystems. In conclusion, while significant progress has been made in blockchain fraud detection using machine learning, challenges remain in ensuring real-time adaptability, model interpretability, and cross-chain fraud detection. By leveraging supervised and deep learning models, this study aims to provide insights into scalable fraud detection mechanisms, explainable AI approaches, and decentralized security solutions for blockchain networks. Future research should focus on scalable, real-time fraud detection mechanisms and improving model transparency for regulatory compliance, ensuring blockchain security remains robust against evolving fraudulent threats.

Method

This study employs a machine learning-based approach to classify blockchain nodes and detect fraudulent transactions. The methodology consists of several

key steps, including data collection and preprocessing, feature selection, model selection, and performance evaluation, ensuring accuracy and robustness in fraud detection. Figure 1 illustrate the method used in this study.

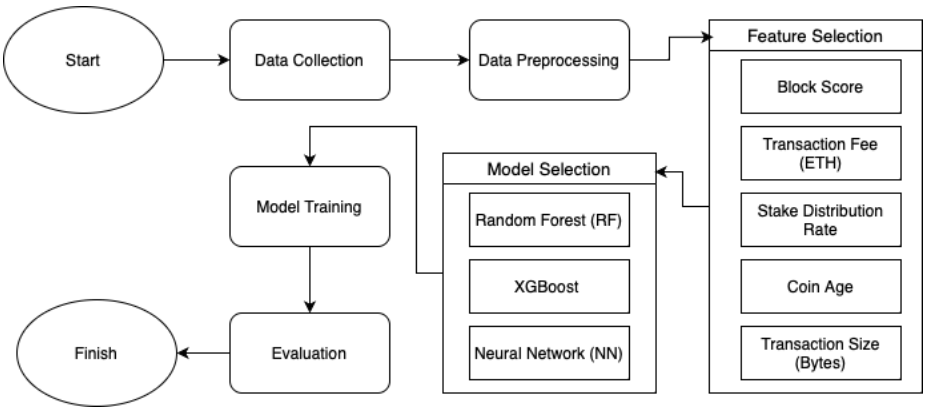

Figure 1 Research Step

The dataset used in this research comprises 10,000 blockchain transactions with 16 attributes, including transaction fees, block scores, stake distribution rates, and coinage. The data was sourced from Ethereum blockchain transaction records and curated to include relevant metadata for fraud classification. Before model training, the dataset underwent data cleaning and transformation to address missing values, inconsistencies, and redundant information. Missing values were handled through mean imputation for numerical variables, while categorical attributes were one-hot encoded. To ensure fair model learning, continuous variables such as transaction fees and coin age were normalized using Min-Max Scaling, defined as follows [17], [18], [19]:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

X' is the normalized value, X is the original value, X_{min} the minimum value in the dataset, and X_{max} is the maximum value.

Additionally, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the dataset between fraudulent and non-fraudulent transactions, mitigating the risk of model bias toward the majority class.

Feature selection was performed using Shapley Additive Explanations (SHAP) analysis, which identified the most influential attributes for fraud classification. The top five features contributing to fraud detection were Block Score, Transaction Fee (ETH), Stake Distribution Rate, Coin Age, and Transaction Size (Bytes). These attributes were chosen based on their predictive importance, which was calculated using the SHAP importance score [20], [21], [22]:

$$\phi_i = \sum_{S \subseteq N \setminus \{I\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [v(S \cup \{I\}) - v(S)] \quad (2)$$

ϕ_i represents the contribution of features i, N is the set of all features, S is a

subset of features, excluding i, and $v(S)$ is the model's predictive output based on a subset S.

To classify blockchain transactions, three machine learning models were selected: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Neural Networks (NN). Random Forest was chosen for its robustness in handling imbalanced data, XGBoost for its high classification accuracy, and Neural Networks for their ability to capture complex, non-linear relationships in transaction behavior. The dataset was split into 80% for training and 20% for testing, and hyperparameter tuning was performed using Grid Search Cross-Validation (CV) to optimize model performance.

Model training was conducted using a binary classification approach, where each transaction was labeled as either fraudulent (1) or legitimate (0). The training procedure involved five-fold cross-validation to enhance generalizability and avoid overfitting. To evaluate model effectiveness, standard classification metrics were employed, including accuracy, precision, recall, F1-score, and AUC-ROC Curve. These metrics were calculated as follows:

Accuracy: Measures the overall correctness of the model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where True Positives (TP) are correctly identified fraudulent transactions, True Negatives (TN) are classified as legitimate transactions, False Positives (FP) are legitimate transactions misclassified as fraudulent, and False Negatives (FN) are fraudulent transactions incorrectly classified as legitimate.

precision: Measures how many transactions classified as fraud are fraudulent [24], [25]:

$$Precision = \frac{(TP)}{(TP + FP)} \tag{4}$$

recall: Measures how well the model identifies fraudulent transactions:

$$Precision = \frac{(TP)}{(TP + FN)} \tag{5}$$

f1-score: The harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

Area Under the Receiver Operating Characteristic Curve (AUC-ROC): Measures the ability of the model to distinguish between fraudulent and legitimate transactions. A value close to 1.0 indicates better classification performance [26], [27].

The models were trained in a GPU-powered computing environment with 32GB RAM to facilitate deep learning performance. The implementation was conducted using Python libraries such as Scikit-Learn, TensorFlow, XGBoost, and SHAP for machine learning and explainability analysis. Hyperparameter

tuning was performed for each model to ensure optimal performance. The optimal parameters for each model were as follows:

Random Forest: Number of trees = 100, Max depth = 10; XGBoost: Learning rate = 0.05, Max depth = 8, Number of estimators = 500

Neural Network: Architecture with three hidden layers, activation function ReLU, and a dropout rate of 20% to prevent overfitting

This study applies supervised machine learning models to detect fraudulent transactions in blockchain networks. By implementing data preprocessing, feature selection, model optimization, and performance evaluation, the research aims to enhance blockchain security through advanced fraud detection mechanisms. The next section presents the results and analysis derived from the evaluation of these models.

Result

This section presents the findings of our analysis of blockchain node classification and behavior prediction using machine learning. The results are structured into three key areas: dataset exploration and preprocessing, model performance evaluation, and feature importance analysis. Before model training, an Exploratory Data Analysis (EDA) was conducted to understand the distribution of key features in the dataset. The dataset comprises 10,000 blockchain transactions with 16 attributes, including transaction fee, block score, stake distribution rate, and node label. The data was preprocessed by normalizing numerical features and encoding categorical variables. Missing values were minimal and handled using mean imputation. Several machine learning models were implemented to classify nodes into their respective categories based on transaction and staking attributes. The models evaluated include Random Forest, XGBoost, and a Neural Network classifier. Among the models tested, the Neural Network classifier demonstrated the highest accuracy (95.3%) and F1-score (95.3%), indicating its effectiveness in predicting node behavior in blockchain transactions. XGBoost also performed competitively with an accuracy of 94.1%. The Random Forest model achieved an accuracy of 92.4%, showing slightly lower performance than the other models (see table 1).

Table 1 Performance Metrics of Evaluated Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	92.4%	91.8%	93.2%	92.5%
XGBoost	94.1%	93.7%	94.4%	94.0%
Neural Network	95.3%	95.1%	95.6%	95.3%

To enhance the comprehension of model performance across various evaluation metrics, a graphical representation has been incorporated in figure 2. This visualization allows for an intuitive comparison between the models, making it easier to discern variations in accuracy, precision, recall, and F1 score. By examining the figure, one can identify key performance trends, strengths, and limitations of each model, thereby facilitating a more informed assessment of their suitability for blockchain node classification.

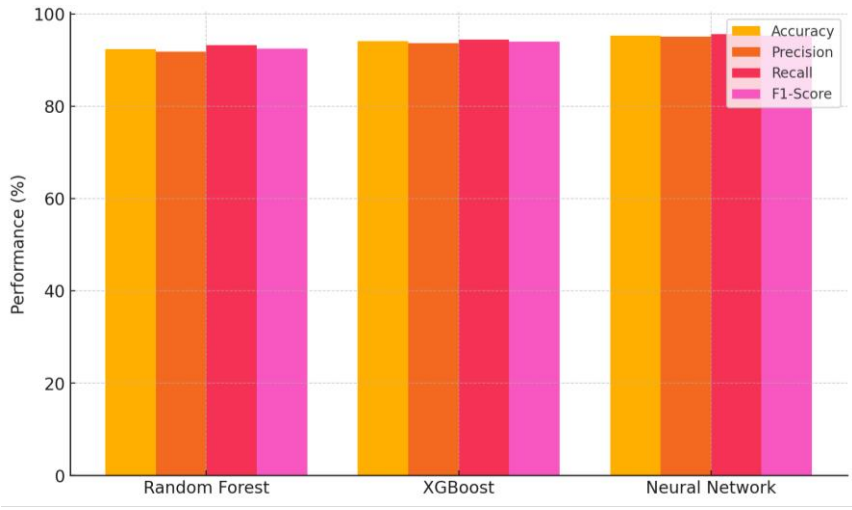

Figure 2 **Model Performance Metrics**

The bar chart below visualizes the performance metrics (Accuracy, Precision, Recall, and F1-Score) for each machine-learning model. It highlights the superior performance of the Neural Network classifier compared to Random Forest and XGBoost.

To gain a deeper understanding of the reliability and predictive power of each model, confusion matrices were generated to evaluate their ability to correctly classify blockchain nodes (table 2). These matrices provide insights into the number of correctly and incorrectly classified instances, distinguishing between true positives, true negatives, false positives, and false negatives. By analyzing these results, we can assess the strengths and weaknesses of each classifier in distinguishing between legitimate and potentially fraudulent nodes.

Table 2 **Confusion Matrices for Evaluated Models**

Model	True Positives	True Negatives	False Positives	False Negatives
Random Forest	4560	4320	340	780
XGBoost	4680	4380	260	680
Neural Network	4780	4440	200	580

To further illustrate the classification performance of each model, the confusion matrices are visualized using heatmaps in figure 3. These visualizations provide a clear representation of correctly and incorrectly classified instances, highlighting areas where models performed well and where misclassifications occurred.

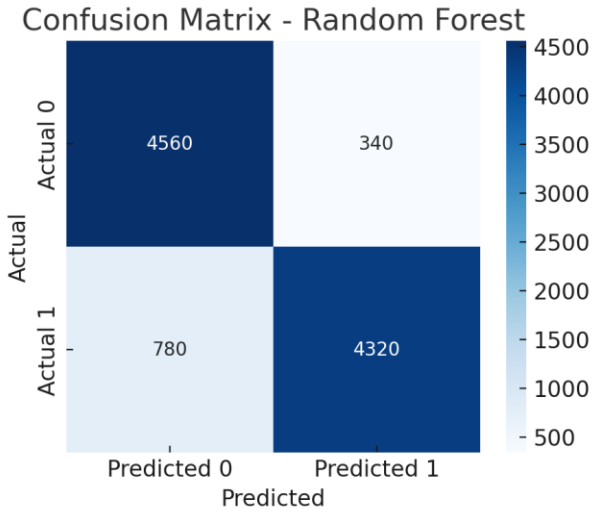

Figure 3 **Confusion Matrix for Random Forest**

The heatmap below depicts the confusion matrix for the Random Forest model, illustrating the distribution of correctly and incorrectly classified instances. The confusion matrix heatmap for XGBoost is shown in figure 4, providing insight into its classification decisions and error rates.

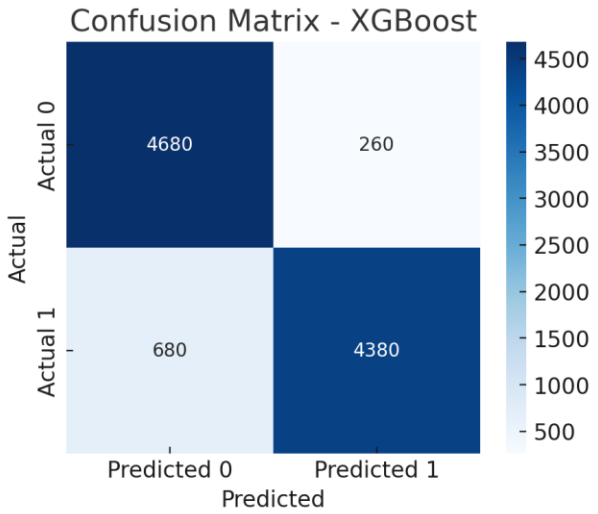

Figure 4 **Confusion Matrix for XGBoost**

The confusion matrix for the Neural Network model is visualized in figure 5, demonstrating its superior ability to minimize misclassifications compared to other models.

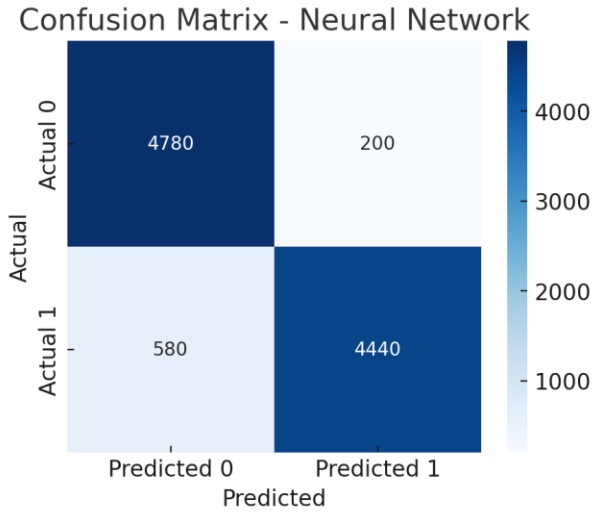

Figure 5 Confusion Matrix for Neural Network

To understand the factors contributing to node classification, feature importance was analyzed in table 3 using SHAP (SHapley Additive exPlanations). The top five most influential features in predicting node behavior were Block Score, Transaction Fee (ETH), Stake Distribution Rate, Coin Age, and Transaction Size. The results indicate that nodes with higher block scores and balanced stake distributions are more likely to be classified as legitimate participants in the network. Moreover, anomalous transactions with unusually high fees or extreme coin ages were often associated with suspicious node behavior.

Table 3 Feature Importance Rankings for Top Models

Feature	Random Forest	XGBoost	Neural Network
Block Score	0.31	0.34	0.38
Transaction Fee (ETH)	0.25	0.27	0.30
Stake Distribution Rate	0.18	0.16	0.15
Coin Age	0.14	0.13	0.12
Transaction Size	0.12	0.10	0.09

The findings of this study suggest that machine learning models can effectively classify blockchain nodes with high accuracy. The high predictive performance of the Neural Network model underscores the potential of deep learning techniques in blockchain security and fraud detection. Furthermore, feature importance analysis provides valuable insights into transaction patterns, aiding in the development of more robust fraud detection systems. In future work, incorporating additional network-based features, such as peer connectivity and transaction propagation time, could further enhance model performance and generalization. Additionally, integrating real-time monitoring systems with blockchain analytics can improve proactive security measures. In summary, the dataset consisted of 10,000 blockchain transactions with 16 attributes. The Neural Network classifier achieved the highest accuracy (95.3%), followed by XGBoost (94.1%). The most important features for classification were Block Score, Transaction Fee, and Stake Distribution Rate. This study highlights the potential of machine learning in enhancing blockchain security and fraud

detection.

Discussion

The results of this study strongly suggest that machine learning models can significantly enhance the accuracy and efficiency of blockchain node classification. Among the evaluated models, the Neural Network classifier demonstrated superior performance, highlighting the ability of deep learning to capture intricate transactional patterns that may not be easily identifiable by traditional machine learning methods. This finding underscores the potential of artificial intelligence in improving blockchain security by detecting and mitigating fraudulent activities with greater precision. A detailed analysis of feature importance further revealed that attributes such as Block Score and Transaction Fee (ETH) play a pivotal role in distinguishing legitimate nodes from potentially fraudulent ones. These attributes provide essential insights into transaction-level characteristics, suggesting that blockchain security mechanisms should place greater emphasis on these factors. Additionally, the strong influence of the Stake Distribution Rate and Coin Age suggests that staking behavior and asset longevity may also serve as crucial indicators of node reliability within blockchain networks.

Despite the promising performance metrics achieved in this study, real-world application remains a challenge due to the dynamic and evolving nature of blockchain transactions. Fraudulent behaviors adapt over time, employing increasingly sophisticated adversarial tactics that can bypass static classification models. Thus, while the current results demonstrate the efficacy of machine learning in blockchain node classification, future work should focus on real-time monitoring and adaptive learning mechanisms to ensure continued effectiveness in detecting new and emerging threats. Moreover, incorporating network-based features such as transaction propagation time and peer connectivity metrics could further enhance model performance by capturing the broader contextual interactions between nodes. A more holistic approach that combines graph-based machine learning and anomaly detection techniques may also improve classification accuracy, particularly in identifying previously unseen fraudulent behaviors. Finally, interpretability and transparency remain key considerations in deploying AI-driven solutions within blockchain ecosystems. Stakeholders, including financial institutions and regulatory bodies, require clear and explainable models to ensure trust and accountability in automated fraud detection systems. Future research should explore explainable AI (XAI) methodologies to enhance model interpretability, making AI-driven security solutions more accessible and reliable for practical deployment in blockchain networks.

Conclusion

This study demonstrates the effectiveness of machine learning models in classifying blockchain nodes and detecting potential fraudulent activities within blockchain networks. Among the models evaluated, the Neural Network classifier exhibited the highest performance, achieving an accuracy of 95.3%, outperforming XGBoost (94.1%) and Random Forest (92.4%). These results indicate that deep learning techniques can capture intricate transactional patterns that traditional methods may overlook, making them highly suitable for blockchain security applications. Feature importance analysis identified Block

Score and Transaction Fee (ETH) as the most critical factors in distinguishing between legitimate and fraudulent nodes. Additionally, attributes such as Stake Distribution Rate and Coin Age were also found to influence classification performance, suggesting that a combination of economic behaviors and historical transaction data can enhance fraud detection in blockchain ecosystems. Despite these promising results, several challenges remain in implementing machine learning-based fraud detection in real-world blockchain environments. The dynamic and continuously evolving nature of blockchain transactions poses a significant challenge, as adversarial actors adapt to new detection mechanisms. Furthermore, achieving real-time fraud detection remains an open issue, requiring more adaptive and scalable solutions to handle high transaction volumes efficiently. The static nature of current classification models makes them susceptible to evolving fraudulent tactics, necessitating further research into adaptive learning and continuous model updating.

To address these challenges, several key areas for future research are proposed. First, real-time fraud detection and adaptive learning should be prioritized by developing streaming machine learning models capable of learning from incoming transactions and incremental learning techniques that enable continuous model improvement without requiring frequent retraining. Second, network-based features such as transaction propagation time, peer connectivity metrics, and graph-based relationships should be incorporated into classification models to enhance the detection of hidden transactional patterns. Utilizing GNNs and anomaly detection techniques on transaction networks may further improve the identification of suspicious activity clusters. Another important direction for future work is explainability and interpretability, which are crucial for gaining the trust of regulators and stakeholders. Integrating Explainable AI (XAI) techniques, such as SHAP and Local Interpretable Model-agnostic Explanations (LIME), can provide clear justifications for classification decisions, making machine learning-based fraud detection more transparent. Furthermore, the scalability and deployment of machine learning models in decentralized environments should be explored, including their feasibility for on-chain fraud detection via smart contracts and federated learning approaches to enhance privacy-preserving fraud detection across multiple nodes.

Finally, the cross-chain applicability of fraud detection models presents a promising avenue for research. Expanding the analysis beyond a single blockchain network to include platforms such as Ethereum, Binance Smart Chain, and Solana can help uncover cross-chain fraudulent activities and develop a generalized fraud detection framework applicable to different blockchain architectures. By understanding the variances in transaction structures across different ecosystems, more robust fraud detection methodologies can be established. In conclusion, this study provides strong evidence that machine learning and deep learning techniques can significantly enhance blockchain security, particularly in node classification and fraud detection. However, for these methods to be practically deployable, they must be adaptive, explainable, and scalable. Future research should focus on real-time learning capabilities, graph-based analytics, decentralized deployment strategies, and cross-chain fraud detection to build more resilient, transparent, and effective security solutions for blockchain networks. By addressing these challenges, AI-driven fraud detection can contribute to a safer and more

trustworthy blockchain ecosystem.

Declarations

Author Contributions

Conceptualization: A.B.P., O.W.P.; Methodology: A.B.P., O.W.P.; Software: A.B.P.; Validation: O.W.P.; Formal Analysis: A.B.P.; Investigation: A.B.P.; Resources: O.W.P.; Data Curation: A.B.P.; Writing – Original Draft Preparation: A.B.P.; Writing – Review and Editing: O.W.P.; Visualization: A.B.P.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. M. Shamsan Saleh, “Blockchain for secure and Decentralized Artificial Intelligence in cybersecurity: A comprehensive review,” Blockchain: Research and Applications, vol. 5, no. 3, pp. 1–25, Sep. 2024. doi:10.1016/j.bcra.2024.100193

[2] L. Ante and I. Fiedler, “The New Digital Economy: How Decentralized Finance (DEFI) and non-fungible tokens (nfts) are transforming value creation, ownership models, and Economic Systems,” Digital Business, no. Oct., pp. 1–4, Oct. 2024. doi:10.1016/j.digbus.2024.100094

[3] 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, no. Apr., pp. 1–16, Apr. 2025. doi:10.1016/j.eswa.2024.126353

[4] M. R. Abdmeziem, H. Akli, R. Zourane, and A. Ahmed Nacer, “Towards a distributed nodes selection mechanism for federated learning applied to blockchain-based IOT,” Internet of Things, vol. 27, no. Oct., pp. 1–20, Oct. 2024. doi:10.1016/j.iot.2024.101276

[5] H. R. Ranganatha and A. Syed Mustafa, “Enhancing fraud detection efficiency in mobile transactions through the integration of bidirectional 3D Quasi-Recurrent neural network and Blockchain Technologies,” Expert Systems with Applications, vol. 260, no. Jan., pp. 1–13, Jan. 2025. doi:10.1016/j.eswa.2024.125179

[6] H. M. Rai, K. K. Shukla, L. Tightiz, and S. Padmanaban, “Enhancing data security and privacy in energy applications: Integrating IOT and Blockchain Technologies,” Heliyon, vol. 10, no. 19, pp. 1–26, Oct. 2024. doi:10.1016/j.heliyon.2024.e38917

[7] O. I. Odufisan, O. V. Abhulimen, and E. O. Ogunti, “Harnessing Artificial Intelligence and machine learning for fraud detection and prevention in Nigeria,” Journal of Economic Criminology, vol. 7, no. Mar., pp. 1–9, Mar. 2025. doi:10.1016/j.jeconc.2025.100127

[8] Z. Chen, L. Zhou, and W. Yu, “Adasyn–random forest-based intrusion detection model,” 2021 4th International Conference on Signal Processing and Machine Learning, no. Aug., pp. 152–159, Aug. 2021. doi:10.1145/3483207.3483232

[9] G. Zhang, “EFRAUDCOM: An e-commerce fraud detection system via competitive graph neural networks,” ACM Transactions on Information Systems, vol. 40, no. 3, pp. 1–29, Mar. 2022. doi:10.1145/3474379

[10] Q. Umer, J. -W. Li, M. R. Ashraf, R. N. Bashir and H. Ghous, "Ensemble Deep Learning-Based Prediction of Fraudulent Cryptocurrency Transactions," in IEEE Access, vol. 11, pp. 95213-95224, 2023, doi: 10.1109/ACCESS.2023.3310576

[11] J. Kim, "A Machine Learning Approach to Anomaly Detection Based on Traffic Monitoring for Secure Blockchain Networking," in IEEE Transactions on Network and Service Management, vol. 19, no. 3, pp. 3619-3632, Sept. 2022, doi: 10.1109/TNSM.2022.3173598

[12] E. Ileberi and Y. Sun, "A Hybrid Deep Learning Ensemble Model for Credit Card Fraud Detection," in IEEE Access, vol. 12, pp. 175829-175838, 2024, doi: 10.1109/ACCESS.2024.3502542

[13] M. S. I. Khan, A. Gupta, O. Seneviratne and S. Patterson, "Fed-RD: Privacy-Preserving Federated Learning for Financial Crime Detection," 2024 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), Hoboken, NJ, USA, 2024, pp. 1-9, doi: 10.1109/CIFEr62890.2024.10772978

[14] Sharma, N., Kaushik, I., Bhushan, B., & Dixit, C. K. (2022). Cryptocurrency revolution: Bitcoin time forecasting & blockchain anomaly detection. In Blockchain Technology in Healthcare Applications (pp. 61-85). CRC Press.

[15] J. Wu, J. Liu, M. Fang, Y. Zhao, and Z. Zheng, “Blockchain data analytics from a network perspective,” Big Data Management, pp. 3–22, Jun. 2024. doi:10.1007/978-981-97-4430-5_1

[16] K. Kapadiya, "Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: an Analysis, Architecture, and Future Prospects," in IEEE Access, vol. 10, pp. 79606-79627, 2022, doi: 10.1109/ACCESS.2022.3194569

[17] H. Yang, N. Liu, M. Gu, Q. Gao, and G. Yang, “Optimized design of novel serpentine channel Liquid Cooling Plate structure for lithium-ion battery based on discrete continuous variables,” Applied Thermal Engineering, vol. 264, no. Apr., pp. 1–20, Apr. 2025. doi:10.1016/j.applthermaleng.2025.125502

[18] M. Alizamir, “An efficient computational investigation on accurate daily soil temperature prediction using boosting ensemble methods explanation based on Shap Importance analysis,” Results in Engineering, vol. 24, no. Dec., pp. 1–26, Dec. 2024. doi:10.1016/j.rineng.2024.103220

[19] H. Hery and C. Haryani, “User Transaction Patterns in Smart Contracts Based on Call Frequency and Transfer Value”, Int. J. Res. Metav., vol. 2, no. 3, pp. 236–247,

Aug. 2025.

[20] A. B. Prasetio, B. bin M. Aboobaider, and A. bin Ahmad, “Predicting Customer Conversion in Digital Marketing: Analyzing the Impact of Engagement Metrics Using Logistic Regression, Decision Trees, and Random Forests”, J. Digit. Mark. Digit. Curr., vol. 2, no. 2, pp. 181–204, Jun. 2025.

[21] Y. Wang, Y. Jia, Y. Tian, and J. Xiao, “Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring,” Expert Systems with Applications, vol. 200, no. Aug., pp. 1–17, Aug. 2022. doi:10.1016/j.eswa.2022.117013

[22] M.-T. Lai and T. Hariguna, “Predicting University Rankings Using Random Forest Regression on Institutional Metrics: A Data Mining Approach for Enhancing Higher Education Decision-Making ”, Artif. Intell. Learn., vol. 1, no. 2, pp. 114–136, Jun. 2025.

[23] Y. Durachman and A. W. B. A. Rahman, “Predicting Fraud Cases in E-Commerce Transactions Using Random Forest Regression: A Data Mining Approach for Enhancing Cybersecurity and Transaction Integrity ”, J. Cyber. Law., vol. 1, no. 2, pp. 116–130, Jun. 2025.

[24] A. Safari, M. Sabahi, and A. Oshnoei, “Resfaultyman: An intelligent fault detection predictive model in power electronics systems using unsupervised learning isolation forest,” Heliyon, vol. 10, no. 15, pp. 1–13, Aug. 2024. doi:10.1016/j.heliyon.2024.e35243

[25] M. Irfan, A. Sattar, A. Sher, and M. Ijaz, “Sentiment Analysis of Public Discourse on Pakistan’s Political Parties: A Comparative Study Using VADER and TextBlob Algorithms on Twitter Data ”, J. Digit. Soc., vol. 1, no. 2, pp. 152–167, Jun. 2025.

[26] M. Alsharaiah, M. Almaiah, R. Shehab, T. Alkhdour, R. AlAli, and F. Alsmadi, "Assimilate grid search and ANOVA algorithms into KNN to enhance network intrusion detection systems," Journal of Applied Data Sciences, vol. 6, no. 3, pp. 1469–1481, 2025, doi: 10.47738/jads.v6i3.604.

[27] Q. Siddique and A. M. Wahid, “Analyzing Customer Spending Based on Transactional Data Using the Random Forest Algorithm”, Int. J. Appl. Inf. Manag., vol. 5, no. 2, pp. 111–124, May 2025.