Explainable Artificial Intelligence Framework for Blockchain Transaction Risk Prediction Using Behavioral Indicators and Transaction Patterns

Maram Merlam Allem[1,*]

[1]University of Business and Technology, Jeddah, Saudi Arabia

ABSTRACT

Blockchain technology has become a fundamental infrastructure for decentralized financial systems and digital asset transactions. However, the increasing volume and complexity of blockchain transactions have introduced significant security challenges, including the identification of potentially risky or suspicious activities. This study proposes an Explainable Artificial Intelligence (XAI) framework for blockchain transaction risk prediction using behavioral indicators and transaction patterns. The proposed framework integrates Extreme Gradient Boosting (XGBoost) for classification, Synthetic Minority Oversampling Technique (SMOTE) for addressing class imbalance, and SHapley Additive exPlanations (SHAP) for model interpretability. The experiments were conducted on a blockchain transaction risk dataset containing 9,000 transaction records with behavioral features such as transaction amount, transaction fee, wallet age, transaction velocity, mixing service usage, cross-border activity, and darknet association indicators. The results show that SMOTE effectively balanced the dataset and improved the detection of minority-class transactions. The proposed model achieved an accuracy of 68.17%, precision of 18.45%, recall of 30.82%, F1-score of 23.09%, and ROC-AUC of 52.67%. Although the predictive performance remains limited, the explainability analysis provides valuable insights into the factors influencing transaction risk. Feature importance analysis identified mixing_service_used and cross_border_flag as the most influential variables, while SHAP analysis highlighted unusual_pattern_score, avg_wallet_transaction_value, and total_wallet_transactions as the most impactful predictors. The findings demonstrate the usefulness of explainable artificial intelligence for understanding blockchain transaction risk and identifying key behavioral indicators associated with suspicious activities. The proposed framework contributes to blockchain security research by enhancing transparency in machine learning-based risk assessment and providing insights for future development of intelligent blockchain monitoring systems.

Keywords Explainable Artificial Intelligence (XAI), Blockchain Transaction Risk Prediction, XGBoost, SHAP, Behavioral Indicators

Submitted 5 March 2026
Accepted 18 May 2026
Published 31 August 2026

Corresponding author
Maram Merlam Allem,
m.alem@ubt.edu.sa

Additional Information and Declarations can be found on page 196

DOI: 10.47738/jcrb.v3i3.77

INTRODUCTION

Blockchain technology has emerged as a transformative innovation that enables decentralized, transparent, and secure transaction processing without the need for centralized authorities [1]. The widespread adoption of blockchain platforms in cryptocurrencies, decentralized finance (DeFi), digital payments, and asset management has significantly increased transaction volumes and network complexity [2]. Although blockchain systems provide immutability and transparency, they remain vulnerable to various security threats, including fraudulent transactions, money laundering, darknet-related activities, and other

How to cite this article: M. M. Allem, "Explainable Artificial Intelligence Framework for Blockchain Transaction Risk Prediction Using Behavioral Indicators and Transaction Patterns," *J. Curr. Res. Blockchain,* vol. 3, no. 3, pp. 184-199, 2026.

suspicious financial behaviors [3], [4]. As blockchain ecosystems continue to expand, the ability to identify and assess transaction risks has become a critical challenge for researchers, regulators, and financial institutions.

Traditional rule-based monitoring systems are often inadequate for detecting increasingly sophisticated transaction patterns because they rely on predefined rules and limited contextual information [5]. Consequently, machine learning techniques have gained significant attention as effective tools for blockchain transaction analysis and risk prediction. Machine learning models are capable of processing large-scale transaction datasets and discovering hidden relationships among behavioral indicators that may signal suspicious activities [6]. Several studies have reported promising results using algorithms such as Random Forest, Support Vector Machine, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) for fraud detection and blockchain transaction classification [7], [8]. These approaches have demonstrated the potential of artificial intelligence in enhancing blockchain security and improving automated risk assessment.

Despite these advances, several limitations remain in existing blockchain transaction risk prediction studies. One of the major challenges is the highly imbalanced nature of transaction datasets, where high-risk transactions typically constitute only a small proportion of the overall data [9]. This imbalance often causes machine learning models to favor majority classes and overlook minority-class transactions that may contain critical security threats. Furthermore, many studies focus primarily on improving predictive accuracy while paying limited attention to the interpretability of model decisions [10]. In security-sensitive environments such as blockchain monitoring and financial compliance, understanding why a transaction is classified as risky is often as important as the prediction itself.

Recent developments in Explainable Artificial Intelligence (XAI) have provided new opportunities to address this limitation. Explainability techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) have been successfully employed to enhance transparency and trustworthiness in machine learning applications across cybersecurity, financial risk assessment, and anomaly detection domains [11], [12]. These methods enable researchers and practitioners to quantify the contribution of individual features to prediction outcomes and gain a deeper understanding of the behavioral factors associated with risk-related activities.

Current state-of-the-art research in blockchain security has increasingly incorporated advanced machine learning and deep learning techniques to improve transaction classification and fraud detection performance. Ensemble learning methods, particularly XGBoost, have demonstrated strong predictive capabilities for handling complex financial datasets and nonlinear relationships among transaction features [8], [13]. Other studies have explored graph-based learning and network-oriented approaches to capture blockchain transaction relationships and interaction patterns more effectively [14]. In parallel, explainable artificial intelligence techniques have been introduced to improve model transparency and facilitate decision-making processes in risk assessment tasks [11], [12]. While these studies have advanced the field considerably, most of them emphasize predictive performance and fraud detection accuracy rather than investigating how specific behavioral indicators contribute to transaction risk.

This situation reveals an important research gap in the existing literature. Although machine learning has been widely applied for blockchain transaction analysis, limited attention has been devoted to systematically examining the influence of behavioral indicators and transaction patterns on risk prediction through an explainable artificial intelligence perspective. Moreover, explainability is often treated as an additional analytical component rather than being integrated into the core framework for risk assessment. As a result, there remains a need for a transparent and interpretable framework capable of identifying the behavioral characteristics that drive blockchain transaction risk while simultaneously providing meaningful predictive insights.

To address this gap, this study proposes an Explainable Artificial Intelligence Framework for Blockchain Transaction Risk Prediction using behavioral indicators and transaction patterns. The framework integrates Extreme Gradient Boosting (XGBoost) for classification, Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance, and SHapley Additive exPlanations (SHAP) for model interpretability. By combining predictive modeling with explainability analysis, this study seeks not only to classify transaction risks but also to identify and interpret the behavioral factors that influence risk-related decisions. The findings are expected to contribute to blockchain security research by providing a transparent risk prediction framework and offering insights into the behavioral indicators associated with suspicious blockchain transactions.

Literature Review

Blockchain technology has transformed digital transaction systems by providing decentralized, transparent, and immutable records of financial activities [15]. Despite these advantages, blockchain networks remain susceptible to various security threats, including money laundering, fraudulent transactions, ransomware payments, darknet-related activities, and illicit financial transfers [16]. The pseudonymous nature of blockchain transactions allows users to conceal their real identities, making risk assessment and transaction monitoring increasingly challenging [17].

To mitigate these risks, researchers have investigated various transaction attributes that may indicate suspicious behavior. Features such as transaction amount, transaction frequency, transaction velocity, wallet age, and cross-border activities have been identified as important indicators for assessing transaction risk [18]. Behavioral analysis has become particularly important because malicious activities often exhibit distinct transaction patterns that differ from those of legitimate users [19].

Machine learning has emerged as an effective approach for blockchain transaction analysis due to its ability to process large-scale datasets and identify hidden relationships among transaction features [20]. Traditional machine learning algorithms, including Logistic Regression, Decision Trees, Support Vector Machines, and Random Forests, have been widely applied for fraud detection and transaction classification tasks [21].

Among various machine learning approaches, ensemble learning methods have demonstrated superior predictive performance because they combine multiple weak learners into a stronger predictive model [22]. Extreme Gradient Boosting (XGBoost) has gained significant attention due to its efficiency, scalability, and ability to model complex nonlinear relationships among features [23]. Previous studies have shown that XGBoost can outperform conventional machine learning algorithms in financial risk assessment, anomaly detection, and fraud detection applications [24]. Consequently, XGBoost has become one of the most widely adopted algorithms for blockchain transaction risk prediction.

One of the major challenges in blockchain transaction risk prediction is class imbalance, where risky transactions typically represent only a small proportion of the overall dataset [25]. This imbalance often causes machine learning models to become biased toward majority classes, resulting in poor detection of high-risk transactions. To address this issue, oversampling techniques have been widely adopted to improve class representation during model training.

The Synthetic Minority Oversampling Technique (SMOTE) is one of the most commonly used methods for handling imbalanced datasets. SMOTE generates synthetic samples for minority classes by interpolating neighboring observations, thereby increasing class balance while preserving the underlying data distribution [26]. Previous studies have demonstrated that SMOTE can improve minority-class detection and enhance the robustness of machine learning models in fraud detection and cybersecurity applications [27].

Although machine learning algorithms are capable of achieving strong predictive performance, many advanced models suffer from limited interpretability, making it difficult to understand the reasoning behind prediction outcomes [28]. In blockchain security applications, transparency is essential because risk predictions often support critical decision-making processes related to fraud prevention, compliance monitoring, and security investigations [29].

Explainable Artificial Intelligence (XAI) has emerged as an important research area aimed at improving the transparency and interpretability of machine learning models [30]. Among the various explainability techniques, SHapley Additive exPlanations (SHAP) has become one of the most widely adopted approaches due to its strong theoretical foundation and ability to provide both local and global explanations [31]. SHAP quantifies the contribution of individual features to model predictions, allowing researchers to identify the most influential variables and understand how they affect decision outcomes [32].

In blockchain transaction analysis, SHAP can reveal the behavioral characteristics that contribute to transaction risk, such as unusual transaction patterns, transaction frequency, wallet characteristics, and suspicious activity indicators. This capability enables a more comprehensive understanding of machine learning predictions and supports the development of transparent blockchain security frameworks [33].

Method

Research Framework

This study proposes an Explainable Artificial Intelligence (XAI) framework for blockchain transaction risk prediction using behavioral indicators and transaction patterns. The framework integrates data preprocessing, class balancing, machine learning classification, and explainability analysis to identify influential factors associated with transaction risk. The overall research framework is illustrated in figure 1.

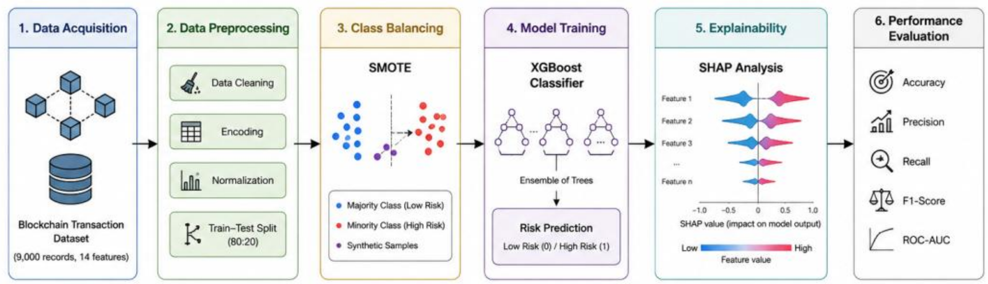

Figure 1 **Research Framework**

The framework begins with blockchain transaction data acquisition, followed by preprocessing and feature preparation. The Synthetic Minority Oversampling Technique (SMOTE) is then applied to address class imbalance in the training dataset. Subsequently, the Extreme Gradient Boosting (XGBoost) model is employed to classify transaction risk levels. Finally, SHapley Additive exPlanations (SHAP) are utilized to interpret model predictions and identify influential behavioral indicators.

Dataset Description

The experiments were conducted using a blockchain transaction risk dataset consisting of 9,000 transaction records and fourteen variables. The dataset contains transaction-related, wallet-related, and behavioral features that may influence transaction risk prediction. A detailed description of the dataset variables is presented in table 1.

Table 1 **Dataset Description**

Variable	Type	Description
transaction_amount	Numeric	Transaction amount
transaction_fee	Numeric	Transaction fee
block_confirmation_time	Numeric	Block confirmation delay
input_count	Numeric	Number of transaction inputs
output_count	Numeric	Number of transaction outputs
wallet_age_days	Numeric	Wallet age (days)
total_wallet_transactions	Numeric	Total historical transactions
avg_wallet_transaction_value	Numeric	Average transaction value
cross_border_flag	Binary	Cross-border transaction indicator
mixing_service_used	Binary	Mixing service usage indicator
darknet_association_score	Numeric	Darknet association score
transaction_velocity_24h	Numeric	Transaction frequency within 24 hours
unusual_pattern_score	Numeric	Behavioral anomaly score
risk_label	Binary	Transaction risk class

The target variable (risk_label) represents transaction risk, where class 0 indicates low-risk transactions and class 1 indicates high-risk transactions.

Data Distribution and Class Balancing

The original dataset exhibits an imbalanced class distribution, as shown in table 2.

Table 2 **Class Distribution**

Class	Count	Percentage (%)
Low Risk (0)	7,607	84.52

High Risk (1)	1,393	15.48
Total	9,000	100.00

The imbalance ratio between majority and minority classes can be expressed as:

$$IR = \frac{N_{majority}}{N_{minority}} \quad (1)$$

$N_{majority}$ denotes the number of low-risk transactions and $N_{minority}$ represents the number of high-risk transactions. To mitigate the impact of class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied. SMOTE generates synthetic minority samples by interpolating neighboring minority-class observations. The synthetic sample generation process can be expressed as:

$$x_{new} = x_i + \lambda(x_{nn} - x_i) \quad (2)$$

x_i is the original minority sample, x_{nn} is its nearest neighbor, and λ is a random value within the interval [0,1].

XGBoost Classification Model

The transaction risk prediction model was developed using Extreme Gradient Boosting (XGBoost), an ensemble learning algorithm based on gradient boosting decision trees. XGBoost combines multiple weak learners to produce a stronger predictive model capable of capturing complex nonlinear relationships among transaction features. The prediction function of XGBoost can be represented as:

$$\hat{y}i = \sum k = 1^K f_k(x_i) \quad (3)$$

$\hat{y}_i$ denotes the predicted output, f_k represents the k-th decision tree, and K is the total number of trees. The objective function optimized by XGBoost is defined as:

$$Obj = \sum_{i=1}^{n} l\,(y_i, \hat{y}i) + \sum k = 1^K\, \Omega(f_k) \quad (4)$$

l is the loss function and Ω is the regularization term used to reduce model complexity and prevent overfitting. The hyperparameter configuration used in this study is summarized in table 3.

Table 3 XGBoost Hyperparameter Configuration

Parameter	Value
n_estimators	500
max_depth	10
learning_rate	0.03
subsample	0.80
colsample_bytree	0.80
scale_pos_weight	5.46
objective	binary:logistic
eval_metric	auc

Explainable Artificial Intelligence Using SHAP

To improve model interpretability, SHapley Additive exPlanations (SHAP) were employed to explain the contribution of individual features to prediction outcomes. SHAP is based on cooperative game theory and assigns importance values to features according to their marginal contributions. The SHAP value is defined as:

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup i) - f(S)] \tag{5}$$

ϕ_i represents the contribution of feature i, S is a subset of features, and $f(\cdot)$ denotes the prediction function. SHAP provides both global and local explanations, enabling the identification of behavioral indicators that have the greatest influence on blockchain transaction risk predictions.

Performance Evaluation

The proposed framework was evaluated using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). These metrics were selected to provide a comprehensive assessment of classification performance, particularly in the presence of class imbalance. Accuracy measures the overall proportion of correctly classified transactions. Precision evaluates the reliability of positive predictions by measuring the proportion of predicted high-risk transactions that are actually classified correctly. Recall assesses the model's ability to identify high-risk transactions, which is particularly important in blockchain security applications where undetected risky activities may result in financial losses or security breaches.

The F1-Score provides a balanced evaluation by considering both Precision and Recall simultaneously, making it suitable for imbalanced classification problems. In addition, ROC-AUC was used to assess the model's ability to distinguish between low-risk and high-risk transactions across different classification thresholds. A higher ROC-AUC value indicates stronger discriminative capability. These evaluation metrics collectively provide a comprehensive assessment of the effectiveness of the proposed Explainable Artificial Intelligence framework for blockchain transaction risk prediction. Figure 2 illustrates the complete experimental workflow, beginning with data preprocessing and feature preparation, followed by class balancing using SMOTE, model training with XGBoost, explainability analysis using SHAP, and performance evaluation.

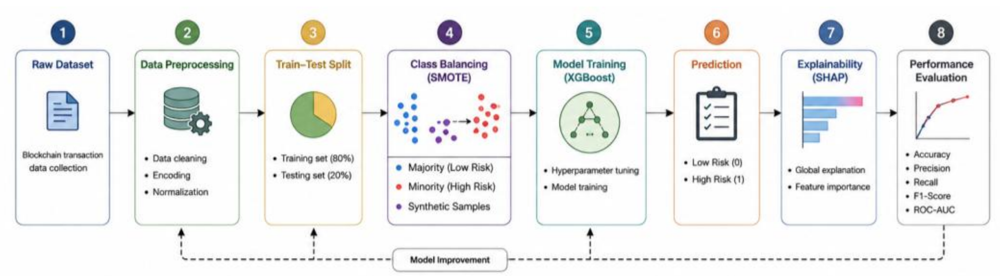

Figure 2 **Methodology Flowchart**

Result and Discussion

Class Distribution Analysis

The initial dataset consisted of 9,000 blockchain transaction records categorized into two risk classes. As illustrated in figure 3, the dataset exhibited a substantial class imbalance, where low-risk transactions (Class 0) accounted for 7,607 records (84.5%), while high-risk transactions (Class 1) represented only 1,393 records (15.5%).

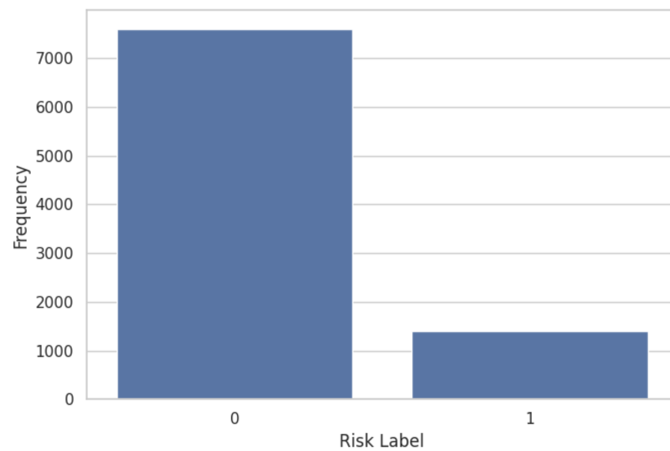

Figure 3 Class Distribution Before SMOTE

The imbalance may bias machine learning models toward the majority class, leading to poor detection of high-risk transactions. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training dataset. As shown in figure 4, SMOTE successfully balanced both classes, resulting in 6,086 samples for each class.

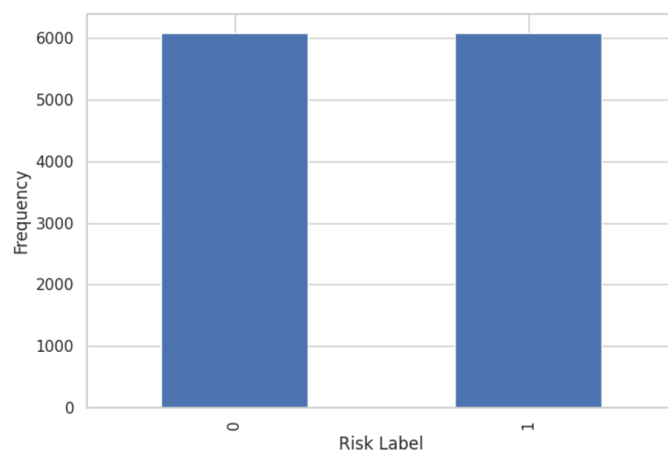

Figure 4 Class Distribution After SMOTE

The balanced distribution enables the classifier to learn patterns from minority-class transactions more effectively and reduces prediction bias toward low-risk transactions.

Classification Performance

The performance of the proposed Explainable Artificial Intelligence framework was evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics (table 4). The model achieved an overall accuracy of 68.17%. However, the precision (18.45%) and recall (30.82%) for the minority class remained relatively low. Although SMOTE improved the detection capability of high-risk transactions, the classifier still struggled to distinguish risk classes effectively. The ROC-AUC value of 0.5267 indicates limited discriminative capability, only slightly outperforming random classification. This suggests that behavioral indicators alone may not provide sufficient predictive information for accurately identifying high-risk blockchain transactions.

Table 4 Model Performance

Metric	Value
Accuracy	0.6817
Precision	0.1845
Recall	0.3082
F1 Score	0.2309
ROC-AUC	0.5267

Confusion Matrix Analysis

The confusion matrix presented in figure 5 provides further insight into the prediction behavior of the model.

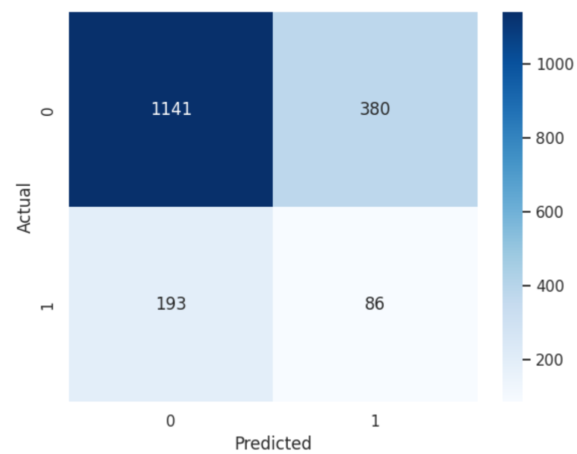

Figure 5 Confusion Matrix

The model correctly classified 1,141 low-risk transactions and 86 high-risk transactions. However, 193 high-risk transactions were incorrectly classified as low risk, indicating a relatively high false-negative rate. From a blockchain security perspective, false negatives are particularly concerning because risky transactions may remain undetected and potentially facilitate illicit activities such as money laundering, darknet transactions, or suspicious fund transfers. The results indicate that while class balancing improved minority-class recognition, further feature enhancement and additional blockchain behavioral indicators are required to improve risk detection performance.

Feature Importance Analysis

To identify the most influential variables affecting transaction risk prediction, feature importance analysis was conducted using the XGBoost model. The feature importance ranking is presented in table 5, which summarizes the five variables with the highest contribution to the model's prediction process.

Table 5 **Feature Importance Ranking**

Rank	Feature	Importance
1	mixing_service_used	0.2320
2	cross_border_flag	0.1952
3	output_count	0.0552
4	unusual_pattern_score	0.0547
5	input_count	0.0545

As shown in table 5, *mixing_service_used* and *cross_border_flag* were the two most influential features in the XGBoost model, with importance scores of 0.2320 and 0.1952, respectively. These results indicate that the use of transaction mixing services and cross-border transaction activities were the primary predictors contributing to the model's risk classification. The relative contribution of these features is further illustrated in figure 6, which presents the feature importance ranking generated by the XGBoost model.

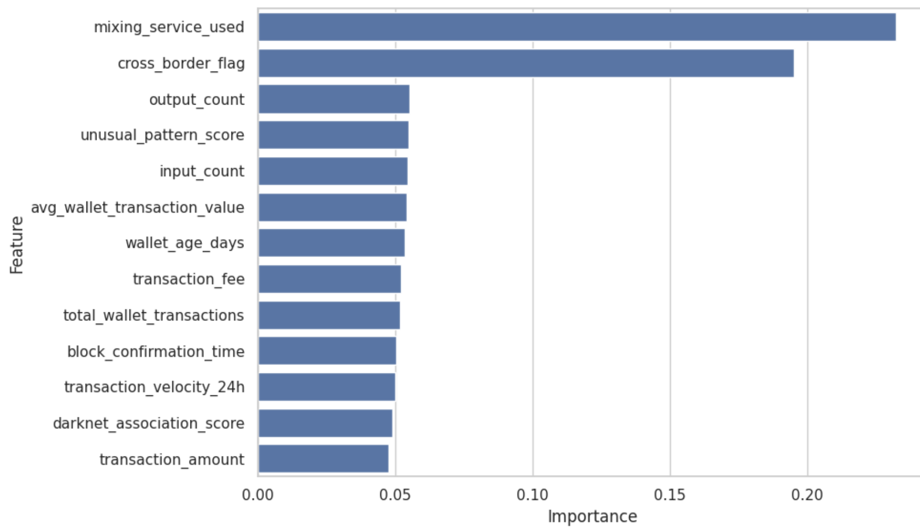

Figure 6 **Feature Importance Ranking**

As illustrated in figure 6, *mixing_service_used* has the highest feature importance, followed by *cross_border_flag*. Mixing services are commonly associated with privacy-enhancing techniques used to obscure transaction origins, while cross-border transactions may indicate complex fund movements that increase transaction risk. Therefore, these indicators appear to play an important role in the model's decision-making process.

Explainable Artificial Intelligence Analysis

To improve model transparency and interpretability, SHAP (SHapley Additive Explanations) was employed to analyze the contribution of individual features to the model's prediction outcomes. Figure 7 presents the SHAP summary plot, which provides a global explanation of feature contributions across the testing samples. In the plot, the horizontal axis represents the SHAP value, where positive values indicate a greater contribution toward the model's predicted risk outcome, while negative values indicate a contribution toward the opposite prediction. The color gradient represents the feature value, with red indicating higher feature values and blue indicating lower feature values. The SHAP feature ranking based on mean absolute SHAP values is presented in table 6.

As shown in table 6, *unusual_pattern_score* has the highest mean SHAP value of 0.4334, followed by *avg_wallet_transaction_value* (0.4202), *total_wallet_transactions* (0.4070), *transaction_fee* (0.3953), and *output_count* (0.3862). These results indicate that transaction behavior and transaction-related characteristics contribute substantially to the model's risk prediction. Figure 7 further illustrates the distribution and direction of these contributions across individual observations. In particular, the wider distribution of SHAP values for *unusual_pattern_score* indicates that this feature has a relatively strong influence on the model output across the evaluated samples.

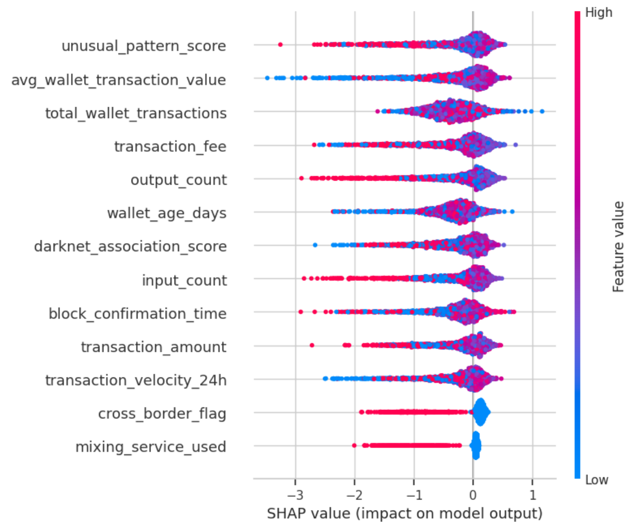

Figure 7 SHAP Summary Plot

The SHAP analysis also reveals differences between conventional XGBoost feature importance and SHAP-based interpretation. While *mixing_service_used* and *cross_border_flag* were among the most influential variables according to the conventional feature importance analysis, figure 7 shows that these features have comparatively narrower SHAP distributions and are ranked lower in terms of their overall contribution. This difference demonstrates that conventional feature importance and SHAP capture different aspects of feature influence. XGBoost feature importance reflects the contribution of variables to the construction of the decision trees, whereas SHAP evaluates how individual features contribute to the model output for each prediction.

Table 6 SHAP Feature Ranking

Rank	Feature	Mean SHAP
1	unusual_pattern_score	0.4334
2	avg_wallet_transaction_value	0.4202
3	total_wallet_transactions	0.4070
4	transaction_fee	0.3953
5	output_count	0.3862

Furthermore, the SHAP summary plot indicates that several behavioral variables, including *unusual_pattern_score*, *avg_wallet_transaction_value*, and *total_wallet_transactions*, exhibit broad distributions along the SHAP axis. This suggests that variations in transaction behavior can substantially affect the model's prediction outcomes. Therefore, the SHAP analysis provides a more detailed interpretation of how transaction-level behavioral characteristics influence blockchain transaction risk prediction, complementing the conventional feature importance analysis.

ROC Curve Evaluation

The receiver operating characteristic (ROC) curve is presented in figure 8. The ROC curve lies close to the diagonal baseline, confirming the relatively low ROC-AUC value of 0.5267. This finding indicates that the model has limited capability in distinguishing between low-risk and high-risk transactions. The results suggest that future studies should incorporate additional blockchain-related features, such as wallet interaction networks, smart contract interactions, transaction graph structures, and temporal behavioral patterns, to improve predictive performance. Overall, although the proposed Explainable Artificial Intelligence framework provides valuable insights into the behavioral factors associated with blockchain transaction risk, further enhancements are required to achieve robust risk prediction performance suitable for real-world blockchain security applications.

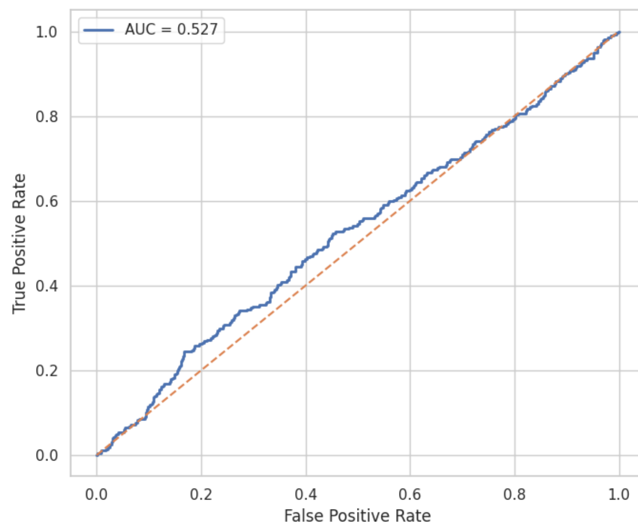

Figure 8 ROC Curve

Discussion

The experimental results indicate that behavioral indicators can provide useful information for blockchain transaction risk assessment; however, their predictive capability remains limited [7], [21], [30]. Although SMOTE successfully addressed the class imbalance problem, the model achieved a ROC-AUC value of only 0.5267, indicating a limited ability to distinguish between low-risk and high-risk transactions [4], [16], [28]. The confusion matrix revealed that a substantial number of high-risk transactions were misclassified as low-risk transactions [11], [25]. This finding suggests that the current behavioral indicators alone may not be sufficient for reliable risk detection in blockchain environments [2], [14], [33].

Feature importance analysis identified *mixing_service_used* and *cross_border_flag* as the most influential variables within the XGBoost model [9], [18], [27]. In contrast, SHAP analysis highlighted *unusual_pattern_score*, *avg_wallet_transaction_value*, and *total_wallet_transactions* as the most influential factors affecting prediction outcomes [1], [12], [23], [31]. This difference demonstrates the value of explainable artificial intelligence techniques in providing a more comprehensive understanding of model behavior [5], [20], [29].

The relatively low classification performance may be attributed to the limited information available in the dataset [8], [17], [26]. Since the dataset only contains transaction-level behavioral indicators and does not include wallet interaction networks or transaction graph structures, important contextual information may be missing [3], [15], [22], [32]. Future studies should incorporate graph-based blockchain features and temporal transaction patterns to improve risk prediction performance [6], [10], [13], [19], [24].

Conclusion

This study proposed an Explainable Artificial Intelligence Framework for Blockchain Transaction Risk Prediction using behavioral indicators and transaction patterns. The framework combined XGBoost classification, SMOTE-based class balancing, and SHAP explainability analysis to identify factors associated with blockchain transaction risk.

The experimental results demonstrated that SMOTE successfully mitigated the class imbalance problem and improved the detection of minority-class transactions. However, the overall predictive performance remained limited, achieving an accuracy of 68.17%, an F1-score of 23.09%, and a ROC-AUC of 52.67%. These results indicate that the behavioral indicators available in the dataset are insufficient for robust risk classification when used in isolation.

Feature importance analysis identified mixing_service_used and cross_border_flag as the most influential variables within the XGBoost model. Meanwhile, SHAP analysis revealed that unusual_pattern_score, avg_wallet_transaction_value, and total_wallet_transactions exerted the strongest impact on prediction outcomes. These findings demonstrate the value of explainable artificial intelligence techniques for understanding blockchain transaction risk beyond conventional machine learning interpretations.

Overall, the proposed framework provides useful insights into behavioral characteristics associated with blockchain transaction risk and highlights the importance of model interpretability in blockchain security applications. Nevertheless, the relatively low predictive performance suggests that future studies should incorporate richer blockchain-specific information, including transaction network structures, wallet interaction graphs, temporal behaviors, and smart contract activities. Such enhancements may significantly improve risk prediction accuracy and support the development of more effective blockchain security monitoring systems.

Declarations

Author Contributions

Conceptualization: M.M.A.; Methodology: M.M.A.; Software: M.M.A.; Validation: M.M.A.; Formal Analysis: M.M.A.; Investigation: M.M.A.; Resources: M.M.A.; Data Curation: M.M.A.; Writing Original Draft Preparation: M.M.A.; Writing Review and Editing: M.M.A.; Visualization: M.M.A.; The author has 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] S. Dong, K. Abbas, M. Li, and J. Kamruzzaman, "Blockchain technology and application: An overview," *PeerJ Comput. Sci.*, vol. 9, no. Nov., pp. 1–20, Nov. 2023, doi: 10.7717/peerj-cs.1705.

[2] R. Auer, B. Haslhofer, S. Kitzler, P. Saggese, and F. Victor, "The technology of decentralized finance (DeFi)," *Digital Finance*, vol. 6, no. 1, pp. 55–95, Aug. 2023, doi: 10.1007/s42521-023-00088-8.

[3] S. Navani and G. T. Cirella, "Cybercrimes in the cryptocurrency domain: Identifying types, understanding motives and techniques, and exploring future directions for technology and regulation," *J. Geogr., Politics Soc.*, vol. 14, no. 2, pp. 1–22, Dec. 2024, doi: 10.26881/jpgs.2024.2.01.

[4] A. Adel and M. Norouzifard, "Weaponization of the growing cybercrimes inside the dark net: The question of detection and application," *Big Data Cogn. Comput.*, vol. 8, no. 8, pp. 1–10, Aug. 2024, doi: 10.3390/bdcc8080091.

[5] D. Yuan, "Intelligent cross-border payment compliance risk detection using multi-modal deep learning: A framework for automated transaction monitoring," *Artificial Intelligence and Machine Learning Review*, vol. 2025, no. 2025, pp. 1–10, 2025, doi: 10.69987.

[6] J. Fan *et al.*, "Deep learning approaches for anti-money laundering on mobile transactions: Review, framework, and directions," *IEEE Internet Things J.*, vol. 13, no. 6, pp. 10407–10431, Mar. 2026, doi: 10.1109/JIOT.2026.3653434.

[7] T. Ashfaq *et al.*, "A machine learning and blockchain based efficient fraud detection mechanism," *Sensors*, vol. 22, no. 19, pp. 1–20, Sep. 2022, doi: 10.3390/s22197162.

[8] A. Ehsan *et al.*, "Enhanced anomaly detection in Ethereum: Unveiling and classifying threats with machine learning," *IEEE Access*, vol. 12, no. 2024, pp. 176440–176456, 2024, doi: 10.1109/ACCESS.2024.3504300.

[9] T. M. Alam *et al.*, "An investigation of credit card default prediction in the imbalanced datasets," *IEEE Access*, vol. 8, no. 2020, pp. 201173–201198, 2020, doi: 10.1109/ACCESS.2020.3033784.

[10] V. Hassija *et al.*, “Interpreting black-box models: A review on explainable artificial intelligence,” *Cogn. Comput.*, vol. 16, no. 1, pp. 45–74, Aug. 2023, doi: 10.1007/s12559-023-10179-8.

[11] S. R. Parisineni and M. Pal, “Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic SHAP explanations,” *Int. J. Data Sci. Anal.*, vol. 18, no. 4, pp. 457–466, Oct. 2023, doi: 10.1007/s41060-023-00458-w.

[12] M. Arunika, S. Saranya, S. Charulekha, S. Kabilarajan, and G. Kesavan, “A survey on explainable AI using machine learning algorithms SHAP and LIME,” *ICCCNT*, vol. 2024, no. Jul., pp. 1–6, Jul. 2024, doi: 10.1109/ICCCNT61001.2024.10725120.

[13] Z. Wang, “Adaptive ensemble learning framework with SHAP-based feature optimization for financial anomaly detection,” *Artificial Intelligence and Machine Learning Review*, vol. 5, no. 1, pp. 51–66, Jan. 2024, doi: 10.69987/aimlr.2024.50105.

[14] J. O. Guballo, “Network-based anomaly detection in blockchain transactions using graph neural network (GNN) and DBSCAN,” *J. Current Res. Blockchain*, vol. 3, no. 1, pp. 15–27, Jan. 2026, doi: 10.47738/jcrb.v3i1.55.

[15] Y. Chen and C. Bellavitis, “Blockchain disruption and decentralized finance: The rise of decentralized business models,” *J. Bus. Venturing Insights*, vol. 13, no. Jun., pp. 1–11, Jun. 2020, doi: 10.1016/j.jbvi.2019.e00151.

[16] S. Navani and G. T. Cirella, “Cybercrimes in the cryptocurrency domain: Identifying types, understanding motives and techniques, and exploring future directions for technology and regulation,” *J. Geogr., Politics Soc.*, vol. 14, no. 2, pp. 1–22, Dec. 2024, doi: 10.26881/jpgs.2024.2.01.

[17] W. Liang *et al.*, “On identity, transaction, and smart contract privacy on permissioned and permissionless blockchain: A comprehensive survey,” *ACM Comput. Surveys*, vol. 56, no. 12, pp. 1–35, Jul. 2024, doi: 10.1145/3676164.

[18] J. Wang, “Financial statecraft in the digital age: Fintech and the institutional shifts in China’s cross-border payment sector since 1993,” *J. Contemporary China*, vol. 2025, no. Aug., pp. 1–17, Aug. 2025, doi: 10.1080/10670564.2025.2544536.

[19] R. Ranjan and S. S. Kumar, “User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user,” *High-Confidence Comput.*, vol. 2, no. 1, pp. 1–10, Mar. 2022, doi: 10.1016/j.hcc.2021.100034.

[20] N. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, and S. Ferretti, “Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics,” *Electronic Markets*, vol. 33, no. 1, pp. 1–17, Jul. 2023, doi: 10.1007/s12525-023-00654-3.

[21] A. R. Khalid *et al.*, “Enhancing credit card fraud detection: An ensemble machine learning approach,” *Big Data Cogn. Comput.*, vol. 8, no. 1, pp. 1–10, Jan. 2024, doi: 10.3390/bdcc8010006.

[22] S. Abimannan, E.-S. M. El-Alfy, Y.-S. Chang, S. Hussain, S. Shukla, and D. Satheesh, “Ensemble multifeatured deep learning models and applications: A survey,” *IEEE Access*, vol. 11, no. 2023, pp. 107194–107217, 2023, doi: 10.1109/ACCESS.2023.3320042.

[23] A. C. Muhammad and R. F. Sari, "Extreme gradient boosting with XAI feature importance for energy prediction," *IAICT*, vol. 2025, no. Aug., pp. 491–497, Aug. 2025, doi: 10.1109/IAICT65714.2025.11101633.

[24] A. A. Ali, A. M. Khedr, M. El-Bannany, and S. Kanakkayil, "A powerful predicting model for financial statement fraud based on optimized XGBoost ensemble learning technique," *Applied Sciences*, vol. 13, no. 4, pp. 1–10, Feb. 2023, doi: 10.3390/app13042272.

[25] S. Liu, B. Cui, and W. Hou, "A survey on blockchain abnormal transaction detection," *Commun. Comput. Inf. Sci.*, vol. 2023, no. Nov., pp. 211–225, Nov. 2023, doi: 10.1007/978-981-99-8101-4_15.

[26] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023, doi: 10.1109/TNNLS.2021.3136503.

[27] T. Albalawi and S. Dardouri, "Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation," *Frontiers in Artificial Intelligence*, vol. 8, no. Oct., pp. 1–15, Oct. 2025, doi: 10.3389/frai.2025.1643292.

[28] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, pp. 1–15, Jul. 2019, doi: 10.3390/electronics8080832.

[29] H. O. Bello, C. Idemudia, and T. V. Iyelolu, "Integrating machine learning and blockchain: Conceptual frameworks for real-time fraud detection and prevention," *World J. Adv. Res. Rev.*, vol. 23, no. 1, pp. 56–68, Jul. 2024, doi: 10.30574/wjarr.2024.23.1.1985.

[30] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, no. 2018, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[31] A. M. Salih *et al.*, "A perspective on explainable artificial intelligence methods: SHAP and LIME," *Advanced Intelligent Systems*, vol. 7, no. 1, pp. 1–15, Jun. 2024, doi: 10.1002/aisy.202400304.

[32] A. V. Ponce-Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clin. Transl. Sci.*, vol. 17, no. 11, pp. 1–15, Oct. 2024, doi: 10.1111/cts.70056.

[33] R. Vijay Anand *et al.*, "Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions," *Scientific Reports*, vol. 15, no. 1, pp. 1–15, Jan. 2025, doi: 10.1038/s41598-024-83564-4.