electronics

Article

ARCADE—Adversarially Robust Cost-Sensitive Anomaly Detection in Blockchain Using Explainable Artificial Intelligence

Muhammad Kamran [1,2], Muhammad Maaz Rehan [1,3], Wasif Nisar [1] and Muhammad Waqas Rehan [4,*]

[1] Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah 47040, Pakistan; kamran.uow@gmail.com or muhammad.kamran@aack.au.edu.pk (M.K.); maazrehan@gmail.com or maazrehan@ciitwah.edu.pk or mmrehan@uclan.ac.uk (M.M.R.); wasif@ciitwah.edu.pk (W.N.)
[2] Department of Computer Science, Air University Islamabad, Aerospace and Aviation Campus Kamra, Attock 43600, Pakistan
[3] Department of Computer Science, School of Engineering and Computing, University of Lancashire (Formerly University of Central Lancashire), Preston PR1 2HE, UK
[4] Institute for Software Engineering and Programming Languages (ISP), University of Lübeck, 23562 Lübeck, Germany
* Correspondence: waqas.rehan@ymail.com or waqas.rehan@isp.uni-luebeck.de

Abstract: Blockchain technology is increasingly being adopted across critical domains, such as healthcare and finance, yet it remains susceptible to anomalies and malicious attacks. Hence, robust anomaly detection is essential in these decentralized systems to maintain integrity, trust, and reliability. However, anomaly detection is still challenging due to data imbalances, adversarial resilience, and the lack of explanation in existing approaches. This work presents ARCADE, a novel approach for adversarially resilient anomaly detection in blockchain networks that leverages an optimized cost-sensitive stacking ensemble learning combined with explainable artificial intelligence (XAI) techniques. Firstly, the proposed approach uses cost-sensitive learning to address the data imbalance problem by optimizing class weights that are integrated with stacking ensemble learning to enhance detection accuracy. Secondly, along with this, newly engineered features are employed to strengthen the resilience of the model against malicious perturbations. Lastly, XAI techniques are applied to provide comprehensive insights and explanations for model prediction. To evaluate ARCADE, the Ethereum network transactions dataset is utilized to ensure a realistic case study. The experimental results show the superiority of the ARCADE in several aspects, achieving a high accuracy of 99.65%; strong resilience against adversarial perturbations, achieving an accuracy of 99.38% for low-intensity attacks, 91.04% for moderate attacks, and over 78% for extreme attacks; and surpassing existing techniques while also providing explainability for domain users.

Keywords: blockchain; anomaly detection; stacking; cost sensitive; adversarial robustness; XAI

Academic Editor: Aryya Gangopadhyay

Received: 12 February 2025
Revised: 4 April 2025
Accepted: 15 April 2025
Published: 18 April 2025

Citation: Kamran, M.; Rehan, M.M.; Nisar, W.; Rehan, M.W. ARCADE—Adversarially Robust Cost-Sensitive Anomaly Detection in Blockchain Using Explainable Artificial Intelligence. *Electronics* **2025**, *14*, 1648. https://doi.org/10.3390/electronics14081648

1. Introduction

In recent years, blockchain technology has witnessed significant proliferation across various critical domains, including healthcare, finance, supply chain management, e-governance, and the Internet of Things (IoT). Blockchain is inherently decentralized, transparent, and immutable, making it a pivotal technology for building trust, reliability, and security in the modern digital ecosystem [1]. Subsequently, this expansion has driven significant market growth, with the blockchain market valued at approximately USD 17.57 billion in 2023 [2]. It is projected to reach USD 825.93 billion by 2032, exhibiting an impressive compound annual growth rate (CAGR) of 52.8% during the forecast period. Despite its

growing popularity and widespread adoption, blockchain technology remains susceptible to several challenges, i.e., anomalous transactions, security breaches, and cybersecurity attacks [3,4]. Consequently, these challenges pose significant constraints on the blockchain system's integrity, reliability, and overall trustworthiness.

The blockchain architecture comprises multiple layers, where each layer is susceptible to specific types of anomalies [5]. The data layer is vulnerable to double-spending and market manipulation anomalies. The network layer is vulnerable to user anomalies, i.e., 51%, and Sybil attacks. Likewise, the incentive layer is prone to risks such as selfish mining and gas price manipulation. Finally, the contract layer is vulnerable to malicious smart contracts. In this context, detecting anomalous transactions is of paramount importance to mitigate the potential risks in blockchain systems. Therefore, the situation necessitates cross-layer and cross-anomaly detection strategies to ensure the overall integrity and security of the blockchain.

Several AI/ML techniques, including supervised learning [3,6–9] and unsupervised learning [10–13], have been used for anomaly detection in blockchain networks. Additionally, ensemble-learning-based approaches have been employed to develop a more robust anomaly detection mechanism [14,15]. Nevertheless, anomaly detection in the blockchain is a nontrivial task due to challenges such as data imbalances, the need for adversarial resilience, and the lack of explainability of the existing approaches. Data imbalance arises when the dataset is highly skewed toward the majority class (normal transactions) rather than the minority class (anomalous transactions). Hence, it hinders the model's ability to learn the minority class (anomalous transactions) features and makes it difficult to train a reliable model. The data imbalance problem is alleviated using sampling and cost-sensitive learning techniques. Sampling methods adjust the distribution of classes, either replicating the minority class (oversampling) or reducing the majority class (undersampling), which causes overfitting or risks discarding valuable information critical for accurate detection [16,17]. Cost-sensitive learning places higher penalties on the model for misclassifying the minority class and hence encourages the model to pay more attention to the rare class (anomalous instances) [18] without overfitting or loss of valuable information. These approaches enhance the model's performance in detecting anomalies. However, existing machine learning methods are prone to adversarial attacks, where even small malicious perturbations can deceive the model [19]. Therefore, in the context of the blockchain, adversarial resilience is imperative to make models robust against deceitful transactions while maintaining high detection accuracy [20].

Notwithstanding the advancements in machine learning techniques for anomaly detection, a crucial facet is the confidence in black-box models and how to create trust in real-world scenarios. Explainable AI (XAI) has emerged as an integral component of ML models to enhance the transparency and interpretability of AI systems [21]. XAI plays an important role in situations where transparency and interpretability are required, such as anomaly detection in blockchain systems. The Shapley Additive Explanations (SHAP) is the most common method of XAI, which was utilized in this study to obtain information about the feature contribution in model prediction. The utilization of an XAI in blockchain anomaly detection enhances the interpretability, transparency, and trustworthiness of detection models. It also improves the reliability of a decentralized blockchain environment [22].

This study proposes ARCADE, which provides an analogy of a secure and structured AI framework or model for blockchain anomaly detection. ARCADE is a novel approach that utilizes optimized cost-sensitive stacking ensemble learning techniques, adversarial resilience, and explainable AI method to address the challenges facing anomaly detection in blockchain systems. The cost-sensitive learning method is used to deal with the class

imbalance issue that is inherent in the blockchain anomaly dataset. A genetic algorithm [23] is used to optimize the class weights, which enables the model to be more accurate in detecting the minority class (anomalous transactions). Tree-based stacking ensemble learning is used to combine multiple base learners with a meta-learner to boost anomaly detection performance. The tree-based stacking ensemble learning with newly extracted features strengthens the model against malicious perturbations. An XAI technique, SHAP analysis, is applied to validate the ARCADE model's predictions by obtaining information about the feature contribution in the decision. This study evaluated the proposed method using the Ethereum network transactions dataset. To validate the effectiveness of a proposed method in terms of anomaly detection accuracy, adversarial resilience, and explainability, extensive experiments were performed. The main contributions of this study can be summarized as follows:

- Handling imbalanced data: we introduce genetic-algorithm-based cost-sensitive stacking ensemble learning to address the data imbalance problem in blockchain datasets, improving the detection of minority-class anomalies.
- Cross-layer and cross-anomaly detection: we detect both transaction- and user-level anomalies across the data layer and network layer of blockchain architecture.
- Adversarial resilience: we develop a model resilient to adversarial attacks, enhancing blockchain anomaly detection by defending against intentional input data manipulations.
- Explainable AI integration: we use SHAP analysis to explain the model predictions, offering transparent insights into the features influencing decision making and enhancing trustworthiness.
- Comprehensive evaluation: we exhibit the efficiency of the proposed approach through extensive experiments on the Ethereum network transaction dataset.

This paper is composed as follows: Section 2 reviews existing ensemble-based anomaly detection methods, highlighting their strengths and limitations. Section 3 describes the proposed methodology. Section 4 presents an evaluation of the proposed model, discussing its results and performance metrics. In Section 5, a comparative analysis is presented. Section 6 outlines ARCADE interpretability using explainable AI. Section 7 concludes this paper, summarizing the findings. Finally, Section 8 suggests directions for future scope and extensions. A list of the acronyms used in this paper and their descriptions is presented in Table 1.

Table 1. List of acronyms and descriptions.

Acronym	Description
AI	Artificial Intelligence
XAI	Explainable Artificial Intelligence, Explainable AI
CAGR	Compound Annual Growth Rate
ARCADE	Adversarially Robust Cost-Sensitive Anomaly Detection
ML	Machine Learning
SHAP	Shapley Additive Explanations
DT	Decision Tree
RF	Random Forest
ETs	Extra Trees
GA	Genetic Algorithm
SFS	Sequential Forward Selection

Table 1. *Cont.*

Acronym	Description
IoT	Internet of Things
SMOTE	Synthetic Minority Oversampling Technique
ADASYN	Adaptive Synthetic Sampling
RUS	Random Undersampling
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
Bi-LSTM	Bidirectional Long Short-Term Memory
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic

2. Literature Review

Recent studies have explored several techniques to detect anomalies in blockchain networks. This literature review delves into state-of-the-art ensemble-based anomaly detection methods for blockchain, examining the strengths and limitations of the existing methods. An efficient blockchain addresses classification [24] through a cascading ensemble learning approach with bitcoin transaction analysis and user behavior classification using machine learning techniques for deanonymization and user-type identification. However, the method has issues like data scarcity, bitcoin users' partial anonymity, and limited focus on specific anomalous behaviors, potentially reducing generalizability. This model lacks cross-layer detection capabilities, model interpretability, and adversarial resilience.

Another method, the BCEAD bagging ensemble-based method by X. Yang et al. [25], focuses on improving wireless sensor network security using a blockchain-based ensemble anomaly detection system. It integrates the blockchain with the isolation forest algorithm for distributed anomaly detection, ensuring trust and resistance to internal attacks. It achieves 94–96% accuracy and reduces storage overhead. However, the system lacks cross-layer detection capabilities and does not address interpretability and adversarial resilience. Further, EnLFADE [26] addresses the detection of fraudulent accounts on the Ethereum blockchain using ensemble learning techniques. It balances the dataset with SMOTE, selects key features using correlation analysis, and compares individual and ensemble classifiers, achieving 99.2% accuracy. However, limitations persist, including reliance of SMOTE for data balancing, which can introduce synthetic data bias as well as the lack of cross-layer detection capabilities, model interpretability, and adversarial resilience.

R. Saxena et al. [27] classified bitcoin transactions as normal and anomalous using ensemble learning. They used a dataset of 4 million transactions and achieved 98.45% accuracy through models like XGBoost and random forest, with hyperparameter tuning and class balancing performed using SMOTE. However, limitations arise due to the dependency on SMOTE, which can introduce synthetic data bias and model complexity as well as lacks cross-layer detection capabilities, model interpretability, and adversarial resilience. S. Hisham et al. [28] proposed anomaly detection in smart contracts based on optimal relevance hybrid feature analysis in the Ethereum blockchain, focusing on detecting anomalies in Ethereum smart contracts using a hybrid feature approach. It combines opcode, ABI code, and transaction data, reducing feature size with the SULOV and MRMR methods. The proposed ensemble model achieves 92.99% accuracy. However, the method faces limitations such as reliance on small datasets. Additionally, the model lacks cross-layer detection capabilities, interpretability, and resilience against adversarial attacks.

The boosting ensemble approach by Q. Umer et al. [29] detects fraud within cryptocurrency ecosystems by combining CNN and LSTM models to handle sequential and structural data. The model achieves an accuracy of 96.4%. However, it faces several limitations, in-

cluding data scarcity and inefficient data-balancing techniques. The lack of cross-layer detection capabilities and interpretability methods also reduces its generalizability and transparency. Further, the authors did not consider model adversarial resilience, making it vulnerable to intentional data perturbations. In another paper, C. Jatoth et al. [30] classified risky and non-risky blockchain blocks. The authors used correlation-based feature selection and combined classifiers through boosting and stacking. The authors claimed a 2–3% improvement in accuracy and a 7–8% increase in F1 score. This method lacks cross-layer detection capabilities, is not resilient to adversarial attacks, and is not interpretable.

N. Nayyer et al. [31] detects anomalous bitcoin transactions using stacking-based ensemble model. The model achieved 97% accuracy. The imbalanced bitcoin heist ransomware dataset was used in this study, and ADASYN-TL resampling technique was used to deal with the class imbalance problem. They also utilized the random search and Bayesian optimization methods to tune the hyperparameters of the model. However, their method has several limitations, such as dependency on data-balancing techniques, which may introduce synthetic data bias. This model also lacks cross-layer detection capabilities, resilience against adversarial attacks, and interpretability. A. Q. Md et al. [32] proposed fraud detection in Ethereum transactions using a stacking-based ensemble model. In their ensemble model, they combined multinomial naive Bayes and random forest as base learners with logistic regression as the meta-learner. Their proposed model achieved an accuracy of 97.18% and an F1 score of 97.02%. However, their proposed method is not interpretable and not resilient to adversarial attacks. Their method also lacks cross-layer detection capabilities.

A study [33] proposed a method of analyzing fraud detection in the Ethereum blockchain using machine learning and ensemble methods. The proposed hard voting ensemble model achieved 99% accuracy. The study also integrated XAI to enhance transparency and trust in the detection process. However, the method is limited by dependency on dataset quality and algorithm-specific constraints. Additionally, the model lacks cross-layer detection capabilities and adversarial resilience. AHEAD [20] proposed an ensemble learning model with stratified random sampling to enhance resilience against adversarial attacks. They used hybrid feature selection and ADASYN for data balancing. However, the study addressed only low-intensity adversarial attacks and did not explain the model. In another study, R. O. Ogundokun et al. [34] proposed a deep-learning-based framework to detect phishing attacks in Ethereum blockchain transactions. The authors use dLSTM, Bi-LSTM, and CNN-LSTM models combined through ensemble voting techniques. However, it has high computational complexity, which restricts its scalability. Additionally, the model lacks cross-layer detection capabilities and does not incorporate adversarial resilience or interpretability methods, limiting its robustness and transparency in real-world applications. A summary of the existing work is presented in Table 2.

It has been observed from the literature review that most of the research work carried out focuses on a single layer and single anomaly detection, limiting the scope of anomaly detection. Furthermore, most models have relied on traditional data-balancing techniques like SMOTE and ADASYN, which can lead to overfitting or loss of important information [16,17,20,26,32,33]. Furthermore, the absence of XAI techniques in many existing approaches reduces transparency and trust in decision-making processes, which is critical for real-world adoption [21,22]. Lastly, the existing machine learning models for anomaly detection remain susceptible to adversarial attacks, which can exploit their vulnerabilities and undermine the overall security and integrity of blockchain networks [24,25,28,29,34].

Table 2. Summary of existing works on anomaly detection using ensemble methods.

Existing Work	Year	Dataset	Ensemble Method	Cross-Anomaly	XAI	Imbalanced Data Handling	Feature Selection	Adversarial Defense
BCEAD [25]	2021	KDD CUP'99	Bagging	×	×	×	Feature extraction	×
Detection of Fraudulent Transactions with XAI [33]	2024	Ethereum Fraud Detection	Voting	×	✓	SMOTE & RUS	Manual	×
Ensemble Learning-Based Anomalous Transaction Detection [29]	2023	Ethereum Fraud Detection	Boosting + Bagging	×	×	×	Pearson correlation	×
Blockchain Addresses Classification [24]	2023	Blockchair, WalletExplorer	Boosting + Bagging	×	×	×	×	×
Improved Classification of Blockchain Transactions [30]	2022	Elliptic Dataset	Stacking + Boosting	×	×	×	Correlation-based	×
Fraud Detection in Bitcoin Transactions [31]	2023	Bitcoin Heist Ransomware	Stacking	×	✓	ADASYN-TL + SMOTE-ENN	Manual	×
Fraud in Ethereum Transactions Using Stacking [32]	2023	Ethereum (Kaggle)	Stacking	×	×	SMOTE	Correlation-based	×
Phishing Detection in Blockchain Transactions [34]	2023	Ethereum Darklist	Voting	×	×	×	Manual	×
EnLFADE [26]	2023	Ethereum Dataset	Bagging + Boosting	×	×	SMOTE	Pearson Correlation	×
Blockchain Transaction Deanonymization [27]	2024	Blockchair, WalletExplorer	Bagging	×	×	SMOTE	Manual	×
Ahead [20]	2024	Ethereum	Voting	✓	×	ADASYN	Hybrid	Partial
Anomaly Detection in Smart Contracts [28]	2023	Ponzi & Non-Ponzi (Etherscan)	Bagging + Boosting	×	×	×	Manual	×
Proposed ARCADE	2025	Ethereum	Stacking	✓	✓	Cost-Sensitive Learning	Sequential Forward Selection	✓

3. Proposed Methodology

This section presents the ARCADE methodology for anomaly detection in blockchain networks. The ARCADE methodology consists of three main components: (i) the preprocessing part introduces newly engineered features that enhance the performance of the model against both normal inputs and adversarially perturbed inputs; (ii) a novel ensemble approach that combines cost-sensitive learning with stacking and uses a genetic algorithm to optimize class weights for base learners that handle imbalanced dataset, further enhancing the model's performance; and, (iii) finally, by integrating XAI, ARCADE provides transparent and interpretable model predictions, giving actionable insights to domain experts for the decision-making process.

The ARCADE methodology is presented in Figure 1 and detailed in the following subsections.

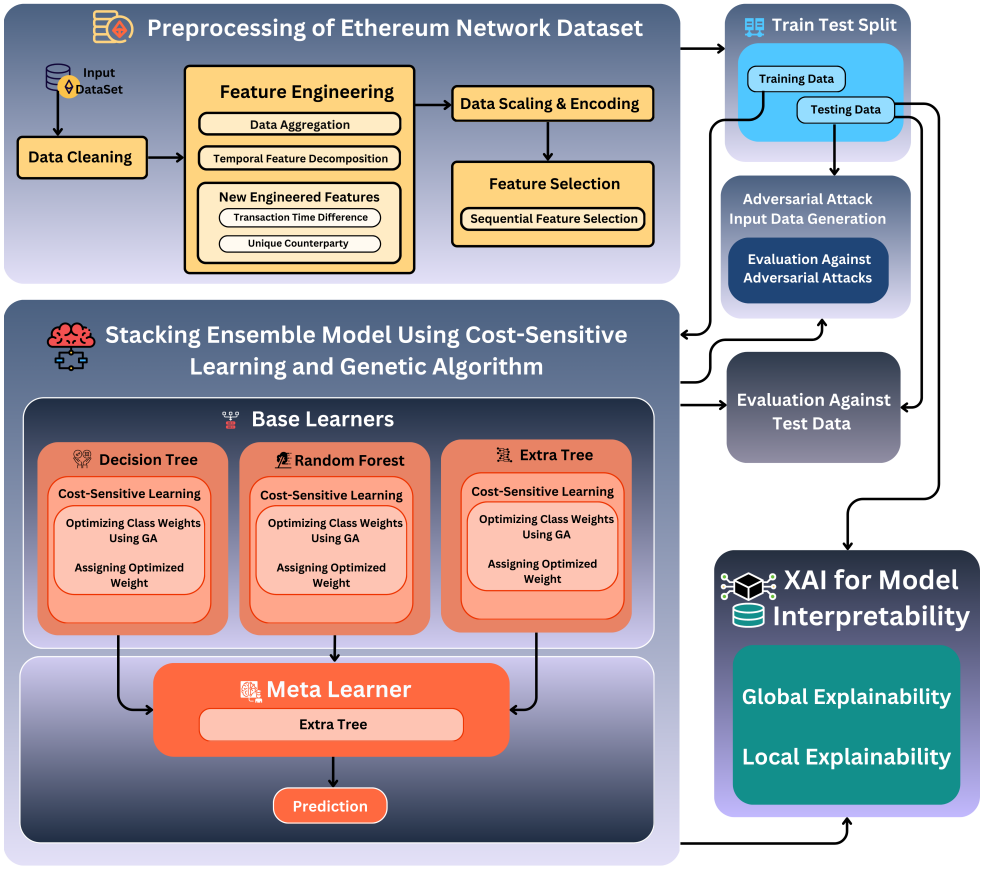

Figure 1. Proposed ARCADE methodology.

3.1. Dataset Preprocessing

The dataset used in our study was derived from the Ethereum network transaction [35] and consists of 71,250 transactions. Each transaction has 18 features. Among the transactions, 57,000 are normal, and 14,250 are anomalous. The dataset has transactional data that helps in detecting anomalies at the data layer and user-related data that helps in detecting anomalies at the network layer. The description of the dataset is presented in Table 3.

Table 3. Dataset features' description.

S. No.	Feature Name	Description
1	Hash	Transaction hash
2	Nonce	Sender's transaction count
3	Transaction_index	Transaction's index in block
4	From_address	Sender's address
5	To_address	Receiver's address
6	Value	Transaction value in Wei
7	Gas	Gas used by transaction
8	Gas_price	Gas price set by sender
9	Input	Transaction data payload
10	Receipt_cumulative_gas_used	Total gas used in block
11	Receipt_gas_used	Gas used by transaction
12	Block_timestamp	Block's timestamp
13	Block_number	Block's number
14	Block_hash	Block's hash
15	From_scam	Flag if sender is anomalous
16	To_scam	Flag if receiver is anomalous
17	From_category	Sender anomaly type
18	To_category	Receiver anomaly type

3.1.1. Data Cleaning

The dataset contains missing values in the 'from_category' and 'to_category' columns. These columns were too noisy and redundant for our model's objectives. We excluded them to maintain data integrity and minimize potential bias.

3.1.2. Feature Engineering

Firstly, data aggregation was performed strategically to enable the model to perform cross-layer and cross-anomaly detection. We introduced a new target attribute class derived from the 'to_scam' and 'from_scam' columns. This splits transactions into three classes:

Class 0: Normal transaction with a normal user.
Class 1: Anomalous transaction with an anomalous sender.
Class 2: Anomalous transaction with an anomalous receiver.

Class 0 represents a normal transaction with the normal user, while 1 and 2 represent an anomalous transaction with an anomalous sender and an anomalous transaction with an anomalous receiver, respectively. The 'from_scam' and 'to_scam' columns were excluded because their information was encapsulated in the derived target attribute, making them redundant.

Further temporal feature decomposition was performed. The 'block_timestamp' feature was decomposed into more detailed components: year, month, day, hour, minute, and second. These time-based features help identify potential temporal anomalies in transaction patterns and reveal recurring or abnormal transaction behaviors across different periods.

Lastly, we engineered several new features to enhance model accuracy further and strengthen resilience against adversarial attacks. First, we introduced transaction time difference features and the gas_to_value_ratio feature. Finally, we developed unique counterparty features to identify abnormal transaction patterns by analyzing repeated or large-scale interactions between specific blockchain addresses. The details of the newly engineered features are provided below:

i. Time Since Last Sender Transaction: This measures the time in hours since the last transaction for each sender, calculated using the difference between consecutive transaction timestamps for each sender, as shown in Equation (1). This feature helps to capture activity frequency and detect unusually long or short intervals that may indicate anomalous behavior.

$$\text{time\_since\_last\_sender\_transaction} = \frac{T_{\text{current}} - T_{\text{previous}}}{3600} \tag{1}$$

ii. Time Since Last Receiver Transaction: This feature captures the time since the last transaction for receiver, as shown in Equation (2).

$$\text{time\_since\_last\_receiver\_transaction} = \frac{T_{\text{current}} - T_{\text{previous}}}{3600} \tag{2}$$

iii. Gas to Value Ratio: It represents the ratio of gas used to the transaction value, as shown in Equation (3). This helps in detecting anomalies in transactions where the gas used is disproportionate to the transaction value. The formula applied is

$$\text{gas\_to\_value\_ratio} = \frac{\text{gas}}{\text{value}} \tag{3}$$

iv. Unique to Counterparties: This feature represents the cumulative count of distinct 'to_address' values that each sender 'from_address' has interacted with over time, as shown in Equation (4). This feature captures the diversity of recipients that a sender

interacts with, and a sudden change in its pattern may indicate anomalous behavior, such as a compromised account attempting multiple fraudulent transactions.

$$\text{unique\_to\_counterparties} = \text{count}(\text{unique}(\text{to\_address})) \tag{4}$$

v. Unique from Counterparties: similarly, this feature captures the cumulative count of distinct 'from_address' values that each receiver 'to_address' has engaged with.

$$\text{unique\_from\_counterparties} = \text{count}(\text{unique}(\text{from\_address})) \tag{5}$$

3.1.3. Encoding and Scaling

To address categorical features, i.e., 'from_address', 'to_address', and 'input', we used label encoding to convert string categories into numerical values. Further, normalization was performed using a standard scaler on all numerical features to ensure the features contribute to the anomaly detection model.

3.1.4. Feature Selection

We employed Sequential Forward Selection (SFS) for feature selection, an iterative method that begins with no features and incrementally adds the most significant ones based on their contribution to model performance. Feature importance was evaluated using an extra tree classifier. The dataset was split into 70% training and 30% testing sets. SFS identified an optimal combination of 12 features, validated through 5-fold cross-validation to ensure robustness and prevent overfitting, achieving an accuracy of 99.21%. The selected features included nonce, to_address, gas, gas_price, input, receipt_gas_used, month, day, hours_since_last_sender_transaction, gas_value_ratio, unique_to_counterparties, and unique_from_counterparties.

3.2. Stacking Ensemble with Cost-Sensitive Learning and Optimization Using Genetic Algorithm

We integrated stacking ensemble learning with cost-sensitive learning, optimizing class weights using a genetic algorithm (GA) to address class imbalance by assigning different misclassification costs. The stacking ensemble includes three base learners: decision tree (DT), random forest (RF), and extra trees (ET). Classifiers were selected through extensive experiments to obtain optimal performance. The meta-learner is an extra tree classifier, chosen for its ability to handle complex data distributions and mitigate overfitting. Let $h_1(x), h_2(x), h_3(x)$ represent the base learners, and let $h_{\text{meta}}(x)$ represent the meta-learner. The stacking model can be represented by Equation (6):

$$\hat{y} = h_{\text{meta}}(h_1(x), h_2(x), h_3(x)) \tag{6}$$

where h_{meta} takes the outputs of the base learners as input and produces the final prediction $\hat{y}$. Next, we explain how cost-sensitive learning and the genetic algorithm are applied in our approach.

3.2.1. Cost-Sensitive Learning in Handling Class-Imbalanced Data

In cost-sensitive learning, different penalties (costs) are assigned for misclassifying instances from different classes, ensuring the model prioritizes correctly classifying minority class instances (anomalies). For our problem, we define class weights w_0, w_1, w_2 for classes 0 (normal transactions), 1 (anomalous transactions and sender), and 2 (anomalous transactions and receiver), respectively. We aim to maximize the following cost-sensitive loss function:

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^{n} w_{y_i} \cdot l(y_i, \hat{y}_i) \tag{7}$$

where y_i is the true class label for the i-th sample, $\hat{y}_i$ is the predicted class label, w_{y_i} is the cost assigned to class y_i, and $\ell(y_i, \hat{y}_i)$ represents the individual cross-entropy loss function. For each classifier in the stacking model, class weights are assigned using the following class weight vectors:

$$C\_W_{\mathrm{DT}} = \{0:1, 1:w_1^{(\mathrm{DT})}, 2:w_2^{(\mathrm{DT})}\};$$
$$C\_W_{\mathrm{RF}} = \{0:1, 1:w_1^{(\mathrm{RF})}, 2:w_2^{(\mathrm{RF})}\};$$
$$C\_W_{\mathrm{ET}} = \{0:1, 1:w_1^{(\mathrm{ET})}, 2:w_2^{(\mathrm{ET})}\}.$$

3.2.2. Genetic Algorithm for Optimizing Class Weights

Further, we use a genetic algorithm (GA) to optimize class weights for cost-sensitive learning. Based on the principles of natural selection, the GA evolves candidate solutions (individuals) over generations to maximize the weighted F1 score of the stacking ensemble model. The detailed process is outlined below:

i. Representation of individuals: Each individual (I) in the GA represents a candidate solution, defined as a set of class weights for the stacking ensemble model. Specifically, each individual is represented as a vector of six real-valued numbers, as presented in (8):
$$\mathrm{I} = [w_1^{(DT)}, w_2^{(DT)}, w_1^{(RF)}, w_2^{(RF)}, w_1^{(ET)}, w_2^{(ET)}] \quad (8)$$
where $w_1^{(DT)}$ and $w_2^{(DT)}$ are the class weights for the decision tree, while $w_1^{(RF)}$, $w_2^{(RF)}$ and $w_1^{(ET)}$, $w_2^{(ET)}$ represent the class weights for the random forest and extra tree classifier, respectively.

ii. Fitness function evaluates the performance of each individual by computing the weighted F1 score, weighted according to class frequencies as in (9):
$$F1_{\mathrm{weighted}} = \frac{\sum_{c\in\{0,1,2\}} w_c \cdot F1_c}{\sum_{c\in\{0,1,2\}} w_c} \quad (9)$$
where $F1_c$ is the F1 score for class c, and w_c is the corresponding class weight.

iii. Genetic algorithm operations: The genetic algorithm evolves a population using three operations: selection, performed via tournament selection to choose the fittest individuals; crossover, which combines parent class weights using blend crossover with a blending parameter α; and mutation, introducing diversity by adding Gaussian perturbations with mean 0 and standard deviation σ to the class weights.

iv. Early stopping and optimization results: the genetic algorithm terminates if the weighted F1 score shows no significant improvement over five consecutive generations or reaches a maximum of 30 generations. The optimal class weights obtained are

$$C\_W_{\mathrm{DT}} = \{0:1, 1:-10.2719, 2:11.0265\},$$
$$C\_W_{\mathrm{RF}} = \{0:1, 1:82.1375, 2:46.9380\},$$
$$C\_W_{\mathrm{ET}} = \{0:1, 1:14.2535, 2:88.0345\}$$

3.3. Adversarial Robustness Evaluation Methodology

To evaluate the adversarial robustness of the anomaly detection model in blockchain systems, we analyzed the impact of varying the input feature perturbations, simulating adversarial attacks. These perturbations, applied as multiplicative factors to a random subset of 1 to 7 features, tested the model's sensitivity to minor and significant changes. Perturbation ranges and their characteristics are presented in Table 4.

Table 4. Perturbation ranges and their characteristics.

S. No.	Perturbation Range	Change Percentage (Max)	Perturbation Level	Adversarial Attack Strength
1	(0.95, 1.05)	±5%	Mild	Low
2	(0.85, 1.15)	±15%	Moderate	Moderate
3	(0.75, 1.25)	±25%	Moderate	Moderate
4	(0.65, 1.35)	±35%	Strong	High
5	(0.55, 1.45)	±45%	Strong	High
6	(0.45, 1.55)	±55%	Very Strong	Very High
7	(0.35, 1.65)	±65%	Very Strong	Very High
8	(0.25, 1.75)	±75%	Extreme	Extreme

The perturbed test data were evaluated using the trained stacking model, with performance measured by accuracy, precision, recall, and F1 score. We identified thresholds where the model's robustness degraded by analyzing performance across different perturbation ranges. These findings highlight the importance of designing models with strong adversarial resilience for effective anomaly detection in blockchain applications.

3.4. Explainable AI for Model Interpretability

To ensure transparency and interpretability in the proposed anomaly detection framework ARCADE, we integrated the XAI technique SHAP to explain the predictions made by the stacking ensemble model. After training the stacking ensemble on the blockchain transaction dataset, we analyzed both global and local model behaviors. Global interpretability was achieved through SHAP summary plots, which rank features based on their average impact on predictions, highlighting the most influential factors for anomaly detection. Local interpretability is provided through SHAP force plots, which offer detailed insights into individual predictions by visualizing the feature's positive and negative contributions. They ensure the interpretability of a model and make the model more reliable for real-time deployment in decentralized environments.

4. Evaluation of the ARCADE Model

This study evaluated the proposed ARCADE model using the two key areas. First, this study evaluated the model's ability for cross-layer and cross-anomaly detection to detect transactional and user anomalies at the data and network layers. Secondly, this study evaluated the model's resilience against adversarial attacks. The performance of the model was evaluated using the following metrics:

- Accuracy: measures the overall correctness of the model.
- Precision: evaluates the proportion of correctly identified anomalies out of all predicted anomalies.
- Recall: measures the model's ability to detect all actual anomalies.
- F1 Score: the harmonic mean of precision and recall, providing a balanced measure.
- AUC: quantifies the model's ability to distinguish among different classes.
- ROC: presents the trade-off between the true and false positive rates, which shows the model's classification performance across different decision boundaries.

4.1. Cross-Layer and Cross-Anomaly Detection Evaluation

The model used a train–test split method with 30% of the data were used for testing, while the remaining 70% were used for training. The experimental result showed that the model achieved near 99.65% accuracy. The proposed method achieved high performance by utilizing advanced preprocessing techniques, feature engineering, sequential forward

feature selection, and cost-sensitive learning. These steps enhanced the prediction accuracy of the learning model.

The confusion matrix, shown in Figure 2, demonstrates the effectiveness of the model in accurately detecting anomalies while minimizing false positives and negatives. Table 5 presents the class-wise AUC score, precision, recall, and F1 score, providing a detailed evaluation of the model's performance for each class. This analysis underscores the robustness and reliability of ARCADE in detecting anomalies across different classes.

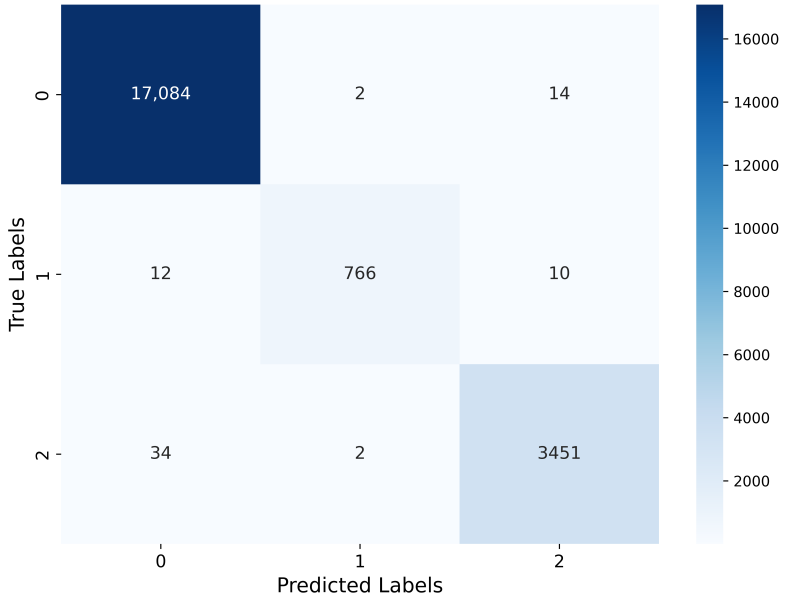

Figure 2. Confusion matrix for cross-layer and cross-anomaly detection model.

Table 5. ARCADE class-wise AUC, precision, recall, and F1 score.

Class	AUC	Precision (%)	Recall (%)	F1 Score (%)
0	0.998	99.73	99.91	99.82
1	0.997	99.48	97.21	98.33
2	0.997	99.31	98.97	99.14

Figure 3 presents the ROC curve for ARCADE across three classes. The AUC values for all classes are 1.00, indicating optimal classification performance. The separation from the random guessing baseline (dashed line) indicates ARCADE's ability to distinguish between classes effectively. The cost-sensitive learning strategy with a genetic algorithm in ARCADE ensured that the classifier was well trained to the imbalanced nature of the dataset, indicating its effectiveness in blockchain anomaly detection, making it a reliable tool for fraud detection, security monitoring, and risk assessment in decentralized environments.

Furthermore, Figure 4 presents a cost-annotated confusion matrix where the values in each cell represent the penalty for misclassification rather than the raw counts. For instance, misclassifying an anomalous receiver as normal incurs a penalty of 170 (the highest cost), the most critical. During training, ARCADE prioritizes minimizing these critical errors by optimizing class weights through genetic algorithms, ensuring that the model learns to avoid costly misclassifications. As a result, the cost-sensitive approach enhances ARCADE's ability to handle imbalanced data and detect high-risk anomalies effectively.

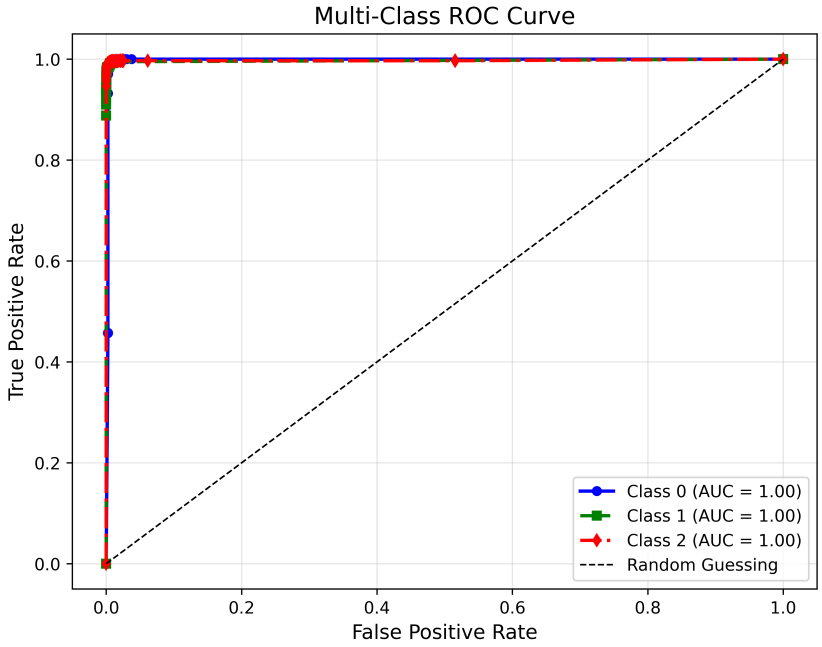

Figure 3. Multi-class ROC curve For ARCADE.

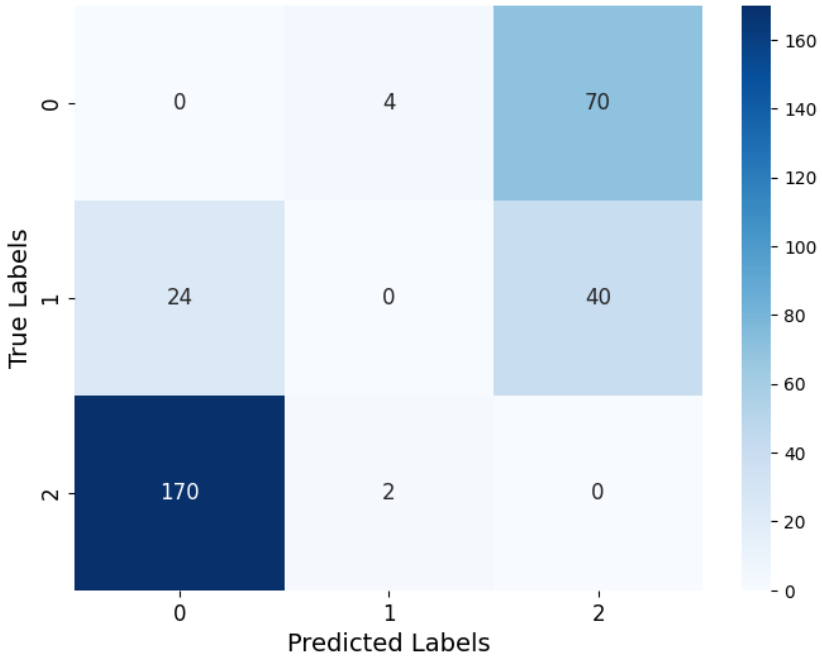

Figure 4. Cost-annotated confusion matrix.

4.2. Adversarial Robustness Evaluation

We assessed the adversarial robustness of our blockchain anomaly detection model by applying controlled perturbations to the input features at varying levels of attack intensity. These perturbations, implemented as multiplicative factors, ranged from minimal noise (0.95–1.05) to severe manipulations (0.25–1.75). Additionally, the number of modified features was progressively increased from one to seven, simulating different levels of adversarial attacks to evaluate the model's resilience.

The results demonstrate the model's strong resilience to adversarial attacks, with performance evaluated across varying attack intensities, as detailed in Table 4. For low-intensity attacks, accuracy remains between 99.38% (one feature modified) to 97.24% (seven features modified), effectively handling minor perturbations. For moderate-intensity attacks, accuracy ranges from 98.61% to 91.04%. Under high-intensity attacks, accuracy drops from 97.52% to 83.80%, and for very high-intensity attacks, it ranges from 97.04% to 80.32%.

As the intensity of the attack and the number of modified features increase, the model remains robust, particularly in detecting anomalies. Even under the extreme perturbation range of 0.25–1.75 and with seven features modified, the model maintains an accuracy of 78.02% and a stable recall of 80.35%. This shows its strong ability to identify anomalies in challenging adversarial scenarios.

The model's adversarial robustness comes from newly engineered features for detecting subtle patterns along a stacking ensemble framework that enhances performance. Table 6 presents the significant improvement in adversarial resilience with engineered features in ARCADE. This demonstrates the effectiveness of the engineered features in improving the robustness of the model against adversarial attacks, making a more reliable anomaly detection system.

Table 6. Adversarial resilience performance comparison: impact without vs. with engineered features (ARCADE).

Noise Range	One Feature Modified		Four Features Modified		Seven Features Modified	
	Without Engineered Features	With Engineered Features (ARCADE)	Without Engineered Features	With Engineered Features (ARCADE)	Without Engineered Features	With Engineered Features (ARCADE)
(0.95, 1.05)	98.55	99.38	96.69	98.00	94.56	96.75
(0.85, 1.15)	96.89	98.43	89.40	94.74	81.11	91.04
(0.75, 1.25)	95.66	97.87	84.88	92.57	74.01	86.85
(0.65, 1.35)	95.01	97.52	82.03	90.76	70.02	84.14
(0.55, 1.45)	94.55	97.38	81.02	89.91	67.22	82.20
(0.45, 1.55)	94.36	97.12	79.27	88.70	64.77	81.16
(0.35, 1.65)	93.89	96.93	77.71	88.00	63.35	79.51
(0.25, 1.75)	92.33	96.65	77.10	87.49	62.52	79.00

Note: all values in the table represent accuracy percentages (%).

The heatmaps of accuracy in Figure 5, precision in Figure 6, recall in Figure 7, and F1 score in Figure 8 present the detailed performance trends across feature ranges and perturbation levels. Overall, the results show that the model is highly robust to low- and moderate-intensity adversarial attacks and remains effective under extreme conditions, highlighting its reliability for real-world applications involving adversarial scenarios.

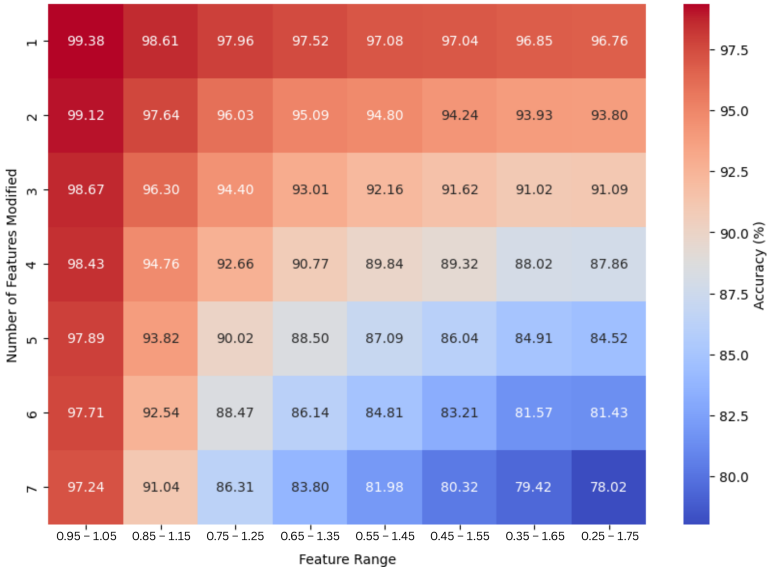

Figure 5. Accuracy: impact of feature range and number of features modified.

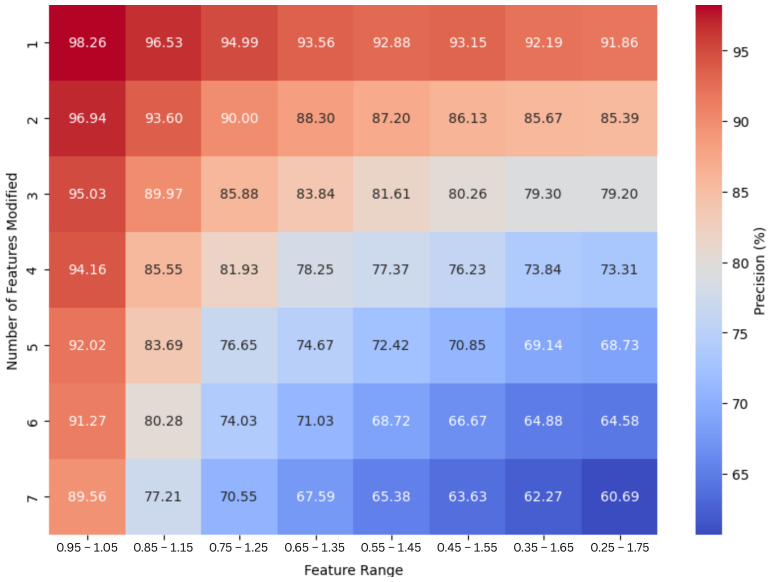

Figure 6. Precision: impact of feature range and number of features modified.

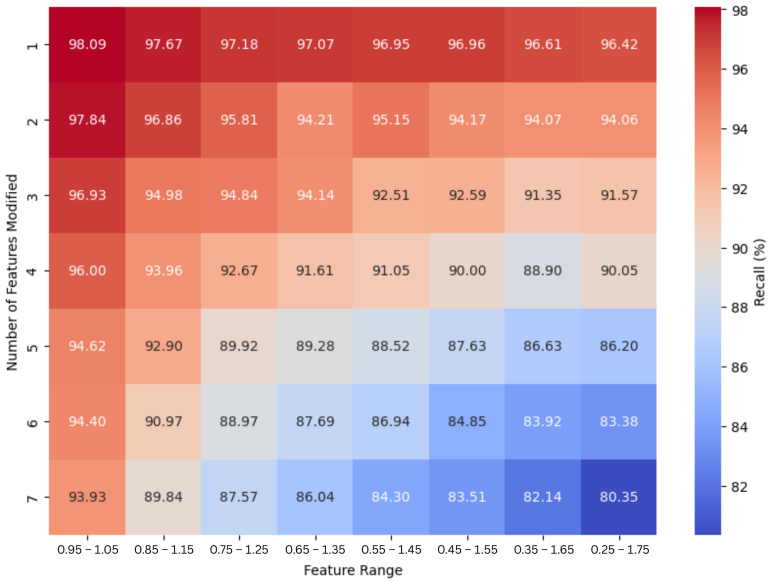

Figure 7. Recall: impact of feature range and number of features modified.

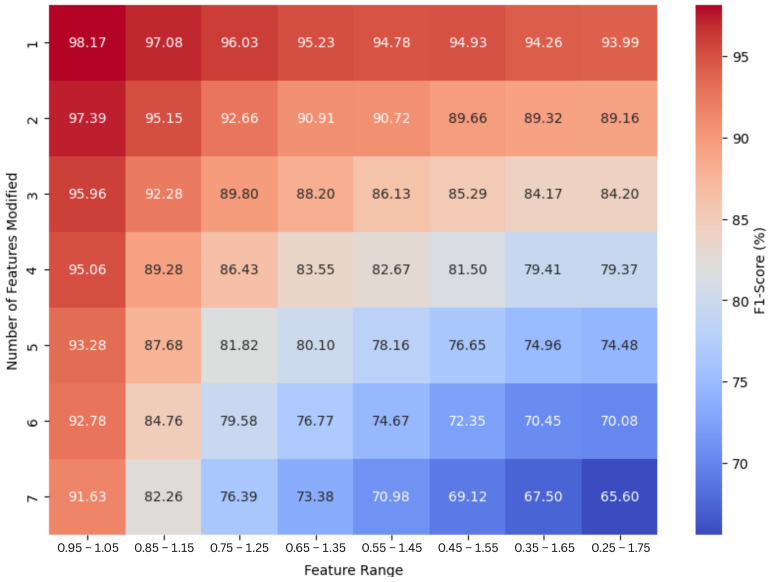

Figure 8. F1 Score: impact of feature range and number of features modified.

5. Comparison of ARCADE Model with Existing Approaches

In this section, we first compare ARCADE with prevalent ensemble learning techniques to evaluate its overall performance. Then, we compare the adversarial robustness of ARCADE with that of ML models.

5.1. Comparing ARCADE with Other Ensemble Learning Techniques

The proposed ARCADE model outperforms existing ensemble learning approaches, achieving 99.65% accuracy. This enhanced performance is driven by several key innovations. Firstly, ARCADE addresses imbalanced data using cost-sensitive learning, optimized with a genetic algorithm that improves minority-class detection without overfitting and bias due to resampling methods that create synthetic data. Secondly, advanced feature engineering includes the temporal decomposition of the transaction timestamps and the creation of new insightful features, such as transaction time differences and unique counterparty features, that improve the model's ability to detect complex anomaly patterns even under adversarial perturbations, a critical requirement for real-world blockchain security.

For instance, AHEAD [20] achieves 98.85% accuracy, falling short by 0.80%. AHEAD is not interpreted using XAI techniques and relies on ADASYN for data balancing, which introduces synthetic data bias. Furthermore, the improved classification of transactions [30], which employs stacking, achieves an accuracy of 98%, falling short by 1.65%. Similarly, the accuracy in detecting fraud in Ethereum transactions [32] is 97.18%, which is 2.47% lower than ARCADE. Both of these techniques are not adversarial-resilient, dependent on SMOTE for imbalanced data handling, and are not explainable. Lastly, the EnLFADE [26] model, which uses bagging, achieves 99.2%; ARCADE surpasses it by 0.45%. EnLFADE also lacks adversarial resilience and model interpretability. A detailed comparison of ARCADE with existing state-of-the-art ensemble learning techniques is presented in Table 7.

ARCADE's performance further stems from its stacking ensemble framework, combining decision tree, random forest, and extra tree classifiers as base learners with extra tree as a meta-learner, and improves accuracy by leveraging diverse perspectives. Lastly, the integration of XAI techniques further enhances transparency and interpretability, providing actionable insights to domain users for an actionable decision-making process. This comparison validates ARCADE as a state-of-the-art advancement in blockchain anomaly detection.

Table 7. Comparison of ARCADE with existing methods.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Adversarial-Resilient	XAI	Imbalanced Data Handling
AHEAD [20]	98.85	98.85	98.85	98.85	Partial	×	ADASYN
EnLFADE [26]	99.20	99.00	99.00	99.00	×	×	SMOTE
Fraud in Ethereum Transactions [32]	97.18	97.03	97.04	97.02	×	×	SMOTE
Improved Classification of Transactions [30]	98.00	98.00	93.00	95.00	×	×	×
ARCADE (Ours)	99.65	99.65	99.65	99.65	✓	✓	Cost-Sensitive Learning

5.2. Adversarial Robustness Analysis: Comparison with ML Models

ARCADE demonstrates resilience against adversarial attacks compared to traditional models. Adversarial robustness testing involved applying controlled perturbations across varying intensity ranges (e.g., 0.95–1.05 to 0.25–1.75) and progressively modifying up to seven features in the test dataset. We present three figures: Figure 9 illustrates a performance comparison for mild to extreme adversarial attack intensity ranges with one feature altered; Figure 10 shows the performance comparison with four features altered. Finally, Figure 11 highlights the performance comparison with seven features altered.

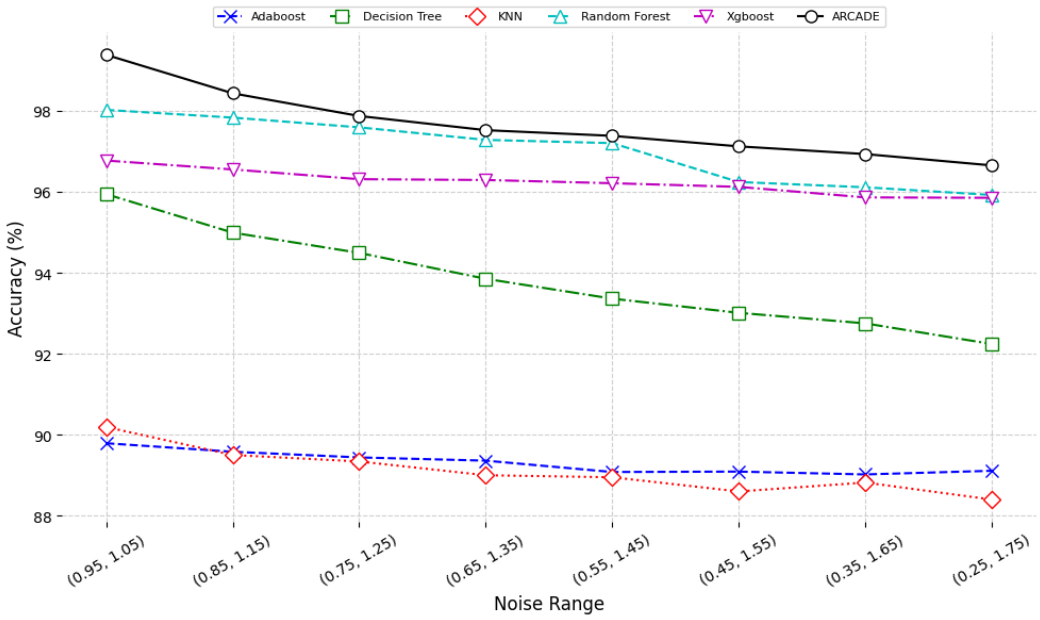

Figure 9. Performance comparison across range of adversarial attacks with one feature altered.

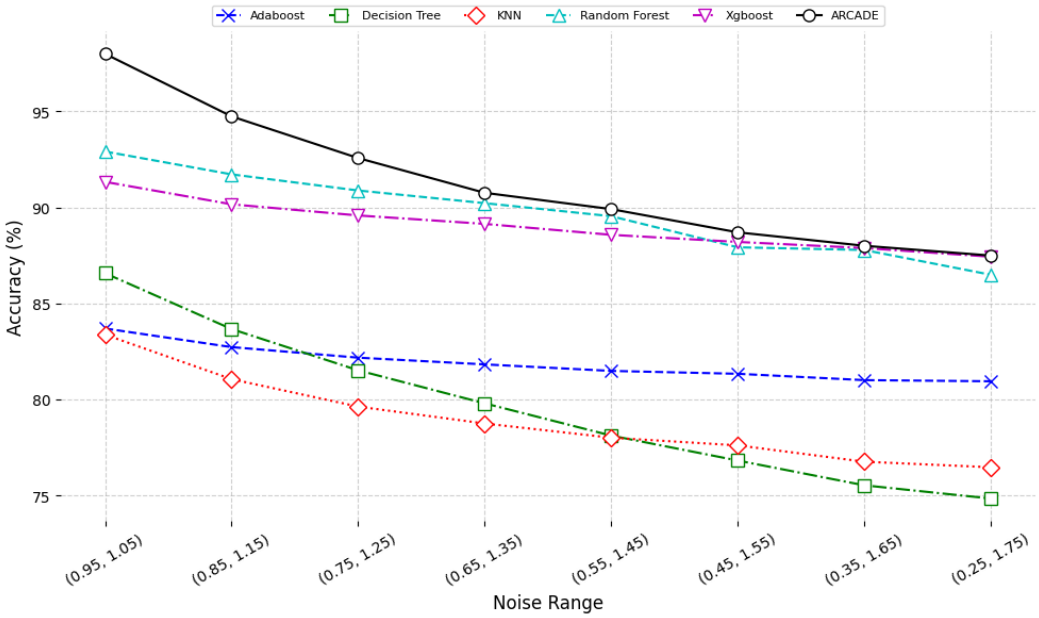

Figure 10. Performance comparison across range of adversarial attacks with four features altered.

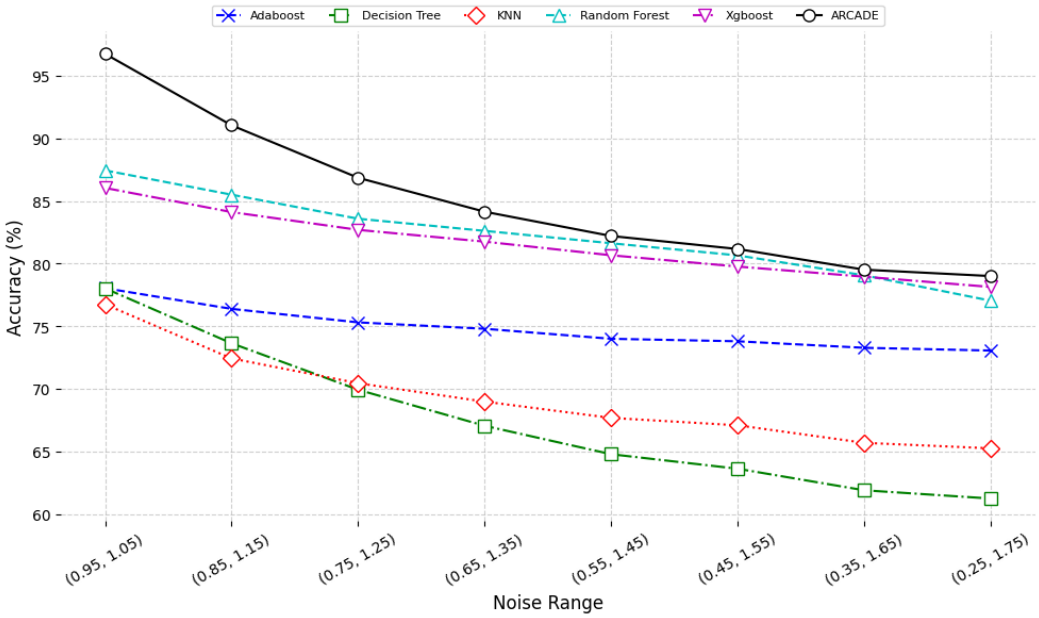

Figure 11. Performance comparison across range of adversarial attacks with seven features altered.

Figures 9–11 reveal interesting performance trends under varying adversarial noise levels and numbers of altered features. Figure 9 shows that ARCADE demonstrates superior performance in the low-, medium-, and high-noise ranges when only one feature is altered, while random forest comes second and maintains relatively comparable performance to ARCADE in the medium-noise ranges of (0.75, 1.25), (0.65, 1.35), and (0.55, 1.45). This is due to its ensemble nature. Xgboost performs relatively worse than ARCADE and random forest but is stable in all noise ranges, followed by decision tree. KNN and Adaboost have low performance across all noise ranges. Figure 10 shows that ARCADE outperforms the other methods in all noise ranges when four features are altered. Random forest and Xgboot have the next-best performance. Decision tree performs well in the low-noise range, but performance declines as noise ranges increase. Adaboost has stable performance across all noise ranges, and KNN comes last. Figure 11 shows that ARCADE also outperforms the other methods across the entire noise range when seven features are altered, followed

by random forest and Xgboost. The next best-performing Adaboost model has stable performance across all noise ranges, while the KNN and decision tree have the same performance at low noise but decrease as the noise range increases.

The comparison performed in this section shows that traditional algorithms, such as KNN and decision tree are not resilient to adversarial attacks and that ensemble-natured algorithms like random forest perform better than traditional algorithms. ARCADE outperforms the others by integrating cost-sensitive learning with a genetic algorithm for optimized weights, new engineered features, and an advanced ensemble strategy. This highlights the effectiveness of ARCADE in handling adversarial noise, ensuring reliable anomaly detection in challenging scenarios.

6. ARCADE Model Interpretability Using Explainable AI

This section describes the global and local interpretability of the ARCADE model using the XAI SHAP technique.

6.1. Global Interpretability of ARCADE Model

The SHAP summary plot provides a detailed ranking of features based on their average impact on the model's predictions, offering valuable insights into the factors driving anomaly detection across different classes. For Class 0 (normal transactions), the top five features include 'month', 'unique_from_counterparties', 'nonce', 'unique_to_counterparties', and 'receipt_gas_used'. These features highlight the importance of temporal patterns and transactional behaviors, such as the diversity of counterparties and gas usage, in distinguishing normal blockchain activity.

For Class 1 (anomalous sender transactions), key features include 'nonce', 'unique_to _counterparties', 'unique_from_counterparties', and 'hours_since_last_sender_transaction'. This indicates that irregularities in sender activity, such as unusually frequent or infrequent transactions and interactions with diverse counterparties, are critical indicators of anomalies.

For Class 2 (anomalous receiver transactions), the most significant features are 'month', 'unique_from_counterparties', 'nonce', 'unique_to _counterparties', 'day', and 'to_address'. These features emphasize the role of temporal patterns, receiver-specific behaviors, and address-level interactions in identifying suspicious activity.

The insights provided by these rankings, visualized in Figures 12–14, are highly actionable for various stakeholders in blockchain-integrated domains. These results provide valuable insights for designing more robust anomaly detection systems by emphasizing impactful features like transaction diversity and temporal patterns. The findings can support the development of advanced fraud detection tools tailored to specific blockchain layers, enabling the effective monitoring of sender and receiver behavior to prevent malicious activities such as double-spending or phishing. Additionally, these insights can be used for innovative feature engineering techniques and the refinement of existing models to improve anomaly detection in decentralized systems. They also highlight high-risk behaviors, offering a foundation for creating strategies and interventions to mitigate fraud and enhance trust and security in blockchain networks.

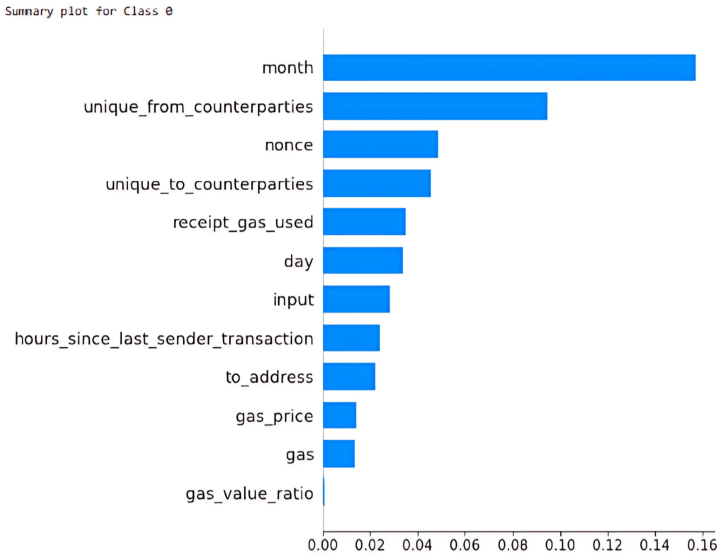

Figure 12. Global interpretation for Class 0.

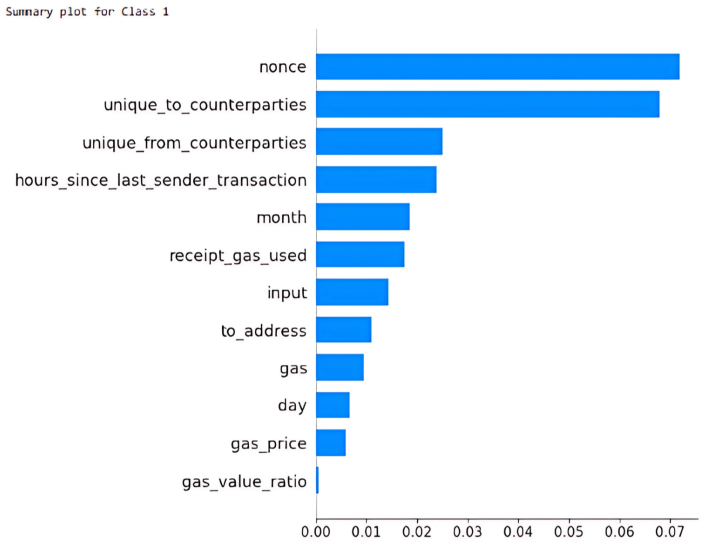

Figure 13. Global interpretation for Class 1.

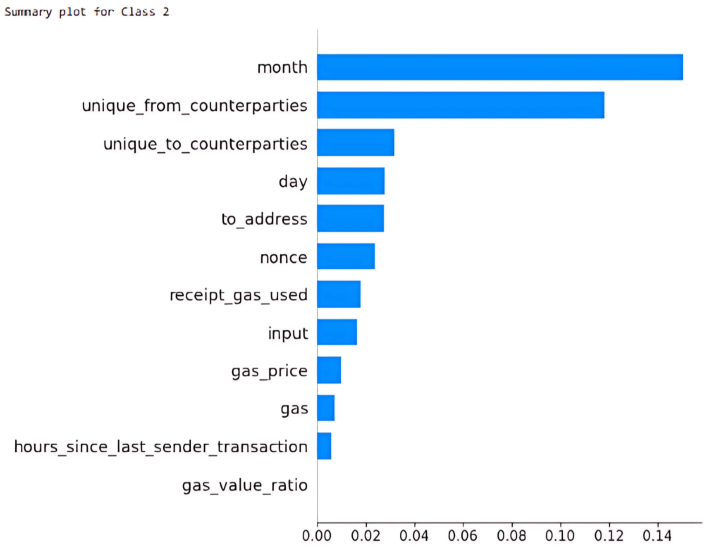

Figure 14. Global interpretation for Class 2.

6.2. Local Interpretability of ARCADE Model

We use force plots to present features that influence individual predictions for local interpretability. The red arrows indicate positive contributions, and the blue arrows show negative ones. The arrow lengths represent the magnitude of each feature's impact, and the final prediction is the sum of the base value and all contributions.

Figure 15 presents the force plot for Class 0 (normal transactions), where the 'month' feature significantly reduces the model output, followed by 'unique_from_counterparties' and 'input'. Minor contributions from 'receipt_gas_used' and 'day' further refine the prediction. This indicates that temporal patterns and transactional diversity play a critical role in identifying normal behavior, which can help developers and analysts optimize blockchain systems for routine operations.

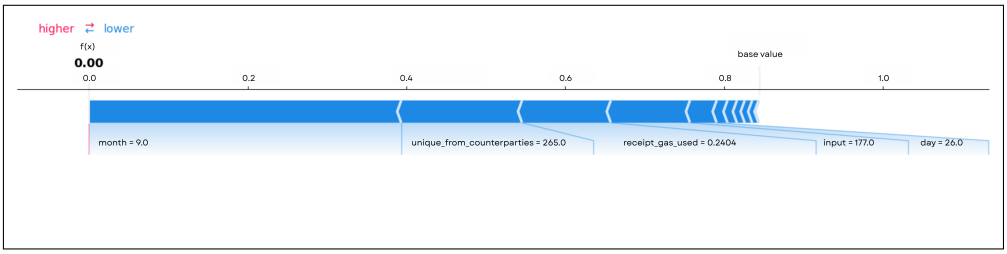

Figure 15. Local interpretation for Class 0.

Figure 16 presents the force plot for Class 1 (anomalous sender transactions), showing a balance of positive and negative contributions. Features like 'nonce' and 'hours_since_last _sender_transaction' reduce the output, while 'unique_to_counterparties' and 'unique_from _counterparties' provide smaller positive contributions, resulting in a prediction close to 0.00. This indicates the importance of monitoring sender activity patterns and transactional diversity to detect anomalies, which is particularly useful for fraud detection and compliance monitoring.

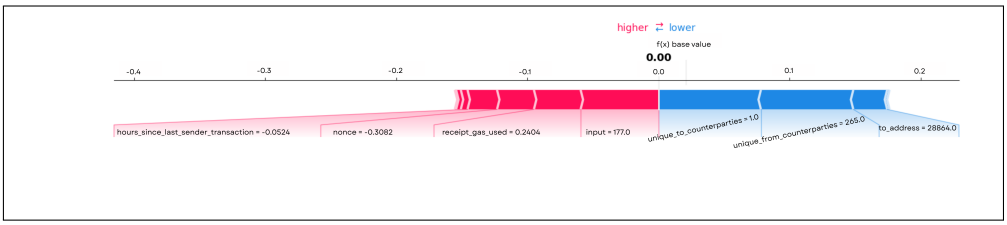

Figure 16. Local interpretation for Class 1.

Figure 17 highlights the force plot for Class 2 (anomalous receiver transactions), where positive contributions from 'to_address', 'receipt_gas_used', 'unique_to_counterparties', and 'input' dominate, pushing the prediction to 1.00. The 'month' feature offers minimal opposing influence. These results indicate the significance of receiver-specific behaviors and gas usage patterns in identifying suspicious transactions, providing valuable insights for designing targeted anomaly detection mechanisms.

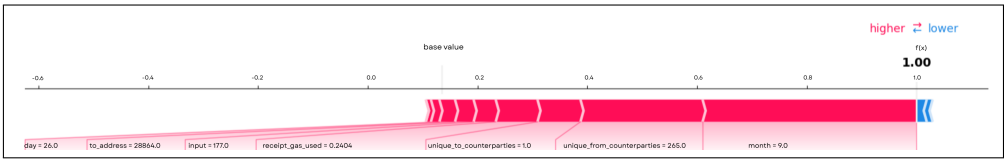

Figure 17. Local interpretation for Class 2.

These insights are critical for improving blockchain security, enabling stakeholders to identify and address specific patterns of anomalous behavior. For researchers, this

interpretability can guide the development of more effective anomaly detection models, while companies and developers can use these findings to enhance fraud detection systems and ensure the reliability of blockchain networks.

Furthermore, the XAI results highlight the strong influence of newly engineered features, such as 'unique_to_counterparties', 'unique_from _counterparties', and 'hours_since _last_sender_transaction', in anomaly detection across all classes. These features help to capture critical patterns and improve detection accuracy, which show the practical impact of this research in the detection of anomalies in the blockchain.

7. Conclusions

This study presents an ARCADE, a novel approach to anomaly detection in blockchain, combining stacking ensemble learning, adversarial resilience, and XAI techniques. This research adopts a multifaceted strategy to address key challenges in anomaly detection, including data imbalance, adversarial resilience, model interpretability and cross-layer and cross-anomaly detection. By employing stacking ensemble architecture combined with cost-sensitive learning optimized via the genetic algorithm, the ARCADE enhances the prediction accuracy up to 99.65% in identifying anomalous transactions and users. Some existing ensemble-based classifiers achieve an accuracy comparable to ARCADE because of their ensemble nature and dependence on synthetic data resampling techniques (e.g., SMOTE/ADASYN), introducing biases. In addition, they fail against adversarial attacks. In ARCADE, adversarial training combined with newly engineered features significantly reinforced the model's robustness against malicious perturbations. Subsequently, when experimentally evaluated on the Ethereum network transactions dataset, the model achieved accuracies of 99.38%, 91.04% and 78% for low-intensity attacks, moderate attacks and extreme attacks, respectively. In addition, the amalgamation of SHAP analysis plays a pivotal role in quantifying individual characteristic contributions to model predictions.

8. Future Scope and Extension

ARCADE demonstrated high performance and adversarial resilience in detecting anomalies using advanced feature engineering, cost-sensitive learning, and explainable AI. Future research may extend ARCADE to multiple blockchain platforms, such as Hyperledger Fabric, Solana, and emerging distributed ledger technologies, to address platform-specific vulnerabilities and validate generalizability. Additionally, we can utilize tree-based and statistics-based feature selection methods to select the relevant features from the data. Traditional deep learning and deep tabular models can also be utilized to enhance anomaly detection. Advanced adversarial training techniques, including graph-based neural networks and reinforcement learning, can further enhance the adversarial resilience of ARCADE. The detection of previously unseen or evolving anomalies remains an open challenge. To address this, future work may integrate unsupervised or semi-supervised learning approaches to enhance existing supervised methods. In the blockchain network, data are transmitted in real time. In the future, we will utilize the continual or lifelong learning method to learn new data anomalies by fine-tuning the ARCADE model. Finally, expanding the interpretability of the model through alternative explainable AI techniques, such as LIME or counterfactual explanation, can also provide deeper insights into decision-making processes. These directions would enhance the robustness and utility of blockchain anomaly detection systems in dynamic and real-world scenarios.

Author Contributions: Conceptualization, M.K.; methodology, M.K., M.M.R. and M.W.R.; software, M.K.; validation, M.K., M.M.R., M.W.R. and W.N.; formal analysis, M.K., M.M.R. and M.W.R.; investigation, M.K., M.M.R. and M.W.R.; resources, M.K., M.M.R., M.W.R. and W.N.; data curation, M.K.; writing—original draft preparation, M.K., M.M.R. and M.W.R.; writing—review and editing,

M.M.R., M.W.R. and W.N.; visualization, M.K., M.M.R. and M.W.R.; supervision, M.M.R., W.N. and M.W.R. All authors have read and agreed to the published version of this manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Hassan, M.U.; Rehmani, M.H.; Chen, J. Anomaly detection in blockchain networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2022**, *25*, 289–318. [CrossRef]
2. Fortune Business Insights. Blockchain Market Size, Share and Global Trend. 2024. Available online: https://www.fortunebusinessinsights.com/industry-reports/blockchain-market-100072 (accessed on 24 December 2024).
3. Hasan, M.; Rahman, M.S.; Janicke, H.; Sarker, I.H. Detecting anomalies in blockchain transactions using machine learning classifiers and explainability analysis. *Blockchain Res. Appl.* **2024**, *5*, 100207. [CrossRef]
4. Ehsan, A.; Iqbal, Z.; Abuowaida, S.; Aljaidi, M.; Zia, H.U.; Alshdaifat, N.; Alshammry, N.K. Enhanced Anomaly Detection in Ethereum: Unveiling and Classifying Threats with Machine Learning. *IEEE Access* **2024**, *12*, 176440–176456. [CrossRef]
5. Mollajafari, S.; Bechkoum, K. Blockchain technology and related security risks: Towards a seven-layer perspective and taxonomy. *Sustainability* **2023**, *15*, 13401. [CrossRef]
6. Sallam, A.; Rassem, T.; Abdu, H.; Abdulkareem, H.; Saif, N.; Abdullah, S. Fraudulent account detection in the Ethereum's network using various machine learning techniques. *Int. J. Softw. Eng. Comput. Syst.* **2022**, *8*, 43–50. [CrossRef]
7. Aziz, R.M.; Baluch, M.F.; Patel, S.; Kumar, P. A machine learning based approach to detect the Ethereum fraud transactions with limited attributes. *Karbala Int. J. Mod. Sci.* **2022**, *8*, 139–151. [CrossRef]
8. Chen, B.; Wei, F.; Gu, C. Bitcoin theft detection based on supervised machine learning algorithms. *Secur. Commun. Netw.* **2021**, *2021*, 6643763. [CrossRef]
9. Alarab, I.; Prakoonwit, S.; Nacer, M.I. Comparative analysis using supervised learning methods for anti-money laundering in bitcoin. In Proceedings of the 2020 5th International Conference on Machine Learning Technologies, Virtual, 19–21 June 2020; pp. 11–17.
10. Sayadi, S.; Rejeb, S.B.; Choukair, Z. Anomaly detection model over blockchain electronic transactions. In Proceedings of the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 895–900.
11. Arya, G.D.; Harika, K.V.S.; Rahul, D.V.; Narasimhan, S.; Ashok, A. Analysis of unsupervised learning algorithms for anomaly mining with bitcoin. In Proceedings of the Machine Intelligence and Smart Systems: Proceedings of MISS 2020, Gwalior, India, 24–25 September 2020; Springer: Singapore, 2021; pp. 365–373.
12. Adam, T.; Babič, F. Anomaly Detection on Distributed Ledger Using Unsupervised Machine Learning. In Proceedings of the 2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS), Berlin, Germany, 23–25 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–4.
13. Kabla, A.H.H.; Anbar, M.; Manickam, S.; Karupayah, S. Eth-PSD: A machine learning-based phishing scam detection approach in ethereum. *IEEE Access* **2022**, *10*, 118043–118057. [CrossRef]
14. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [CrossRef]
15. Hisham, S.; Makhtar, M.; Aziz, A.A. Combining multiple classifiers using ensemble method for anomaly detection in blockchain networks: A comprehensive review. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 404–422. [CrossRef]
16. Alkhawaldeh, I.M.; Albalkhi, I.; Naswhan, A.J. Challenges and limitations of synthetic minority oversampling techniques in machine learning. *World J. Methodol.* **2023**, *13*, 373. [CrossRef] [PubMed]
17. Xie, S.; Zhang, J. Handling highly imbalanced data for classifying fatality of auto collisions using machine learning techniques. *J. Manag. Anal.* **2024**, *11*, 317–357. [CrossRef]
18. Feng, F.; Li, K.C.; Shen, J.; Zhou, Q.; Yang, X. Using cost-sensitive learning and feature selection algorithms to improve the performance of imbalanced classification. *IEEE Access* **2020**, *8*, 69979–69996. [CrossRef]
19. Alshahrani, E.; Alghazzawi, D.; Alotaibi, R.; Rabie, O. Adversarial attacks against supervised machine learning based network intrusion detection systems. *PLoS ONE* **2022**, *17*, e0275971. [CrossRef]

20. Kamran, M.; Rehan, M.M.; Nisar, W.; Rehan, M.W. AHEAD: A Novel Technique Combining Anti-Adversarial Hierarchical Ensemble Learning with Multi-Layer Multi-Anomaly Detection for Blockchain Systems. *Big Data Cogn. Comput.* **2024**, *8*, 103. [CrossRef]
21. Minh, D.; Wang, H.X.; Li, Y.F.; Nguyen, T.N. Explainable artificial intelligence: A comprehensive review. *Artif. Intell. Rev.* **2022**, *55*, 3503–3568. [CrossRef]
22. Sharma, C.; Sharma, S.; Sharma, K.; Sethi, G.K.; Chen, H.Y. Exploring explainable AI: A bibliometric analysis. *Discov. Appl. Sci.* **2024**, *6*, 615. [CrossRef]
23. Albadr, M.A.; Tiun, S.; Ayob, M.; Al-Dhief, F. Genetic algorithm based on natural selection theory for optimization problems. *Symmetry* **2020**, *12*, 1758. [CrossRef]
24. Saxena, R.; Arora, D.; Nagar, V. Efficient blockchain addresses classification through cascading ensemble learning approach. *Int. J. Electron. Secur. Digit. Forensics* **2023**, *15*, 195–210. [CrossRef]
25. Yang, X.; Chen, Y.; Qian, X.; Li, T.; Lv, X. BCEAD: A Blockchain-Empowered Ensemble Anomaly Detection for Wireless Sensor Network via Isolation Forest. *Secur. Commun. Netw.* **2021**, *2021*, 9430132. [CrossRef]
26. Pahuja, L.; Kamal, A. Enlfade: Ensemble learning based fake account detection on Ethereum blockchain. *SSRN Electron. J.* **2022**. [CrossRef]
27. Saxena, R.; Arora, D.; Nagar, V.; Chaurasia, B.K. Blockchain transaction deanonymization using ensemble learning. *Multimed. Tools Appl.* **2024**, *83*, 84589–84618. [CrossRef]
28. Hisham, S.; Makhtar, M.; Aziz, A.A. Anomaly detection in smart contracts based on optimal relevance hybrid features analysis in the Ethereum blockchain employing ensemble learning. *Transactions* **2023**, *3*, 5.
29. Umer, Q.; Li, J.W.; Ashraf, M.R.; Bashir, R.N.; Ghous, H. Ensemble deep learning based prediction of fraudulent Cryptocurrency transactions. *IEEE Access* **2023**, *11*, 95213–95224. [CrossRef]
30. Jatoth, C.; Jain, R.; Fiore, U.; Chatharasupalli, S. Improved classification of blockchain transactions using feature engineering and ensemble learning. *Future Internet* **2021**, *14*, 16. [CrossRef]
31. Nayyer, N.; Javaid, N.; Akbar, M.; Aldegheishem, A.; Alrajeh, N.; Jamil, M. A new framework for fraud detection in bitcoin transactions through ensemble stacking model in smart cities. *IEEE Access* **2023**, *11*, 90916–90938. [CrossRef]
32. Md, A.Q.; Narayanan, S.S.S.; Sabireen, H.; Sivaraman, A.K.; Tee, K.F. A novel approach to detect fraud in Ethereum transactions using stacking. *Expert Syst.* **2023**, *40*, e13255. [CrossRef]
33. Taher, S.S.; Ameen, S.Y.; Ahmed, J.A. Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 12822–12830. [CrossRef]
34. Ogundokun, R.O.; Arowolo, M.O.; Damaševičius, R.; Misra, S. Phishing detection in blockchain transaction networks using ensemble learning. *Telecom* **2023**, *4*, 279–297. [CrossRef]
35. Al-E'mari, S.; Anbar, M.; Sanjalawe, Y.; Manickam, S. A labeled transactions-based dataset on the ethereum network. In Proceedings of the International Conference on Advances in Cyber Security, Penang, Malaysia, 8–9 December 2020; Springer: Singapore, 2020; pp. 61–79.