Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

EXPLAINABLE AI FOR DEFI FRAUD DETECTION: A COMPARATIVE STUDY WITH LARGE-SCALE TRANSACTION DATA

Muhammad Saqib[1], Qamas Gul Khan Safi[1], Muhammad Munwar Iqbal[1,*], Saleem Iqbal[2], Muhammad Farooq[3], and Muhammad Ibrahim[1]

[1]Department of Computer Science, University of Engineering and Technology Taxila, Pakistan.
[2]Department of Computer Science, Allama Iqbal Open University Islamabad, Pakistan.
[3]Department of Information Technology, University of the Punjab, Lahore

awansaqib342@gmail.com, qamas.gul@uettaxila.edu.pk, munawar.iq@uettaxila.edu.pk, saleem.iqbal@aiou.edu.pk, mfarooq@pucit.edu.pk, ukibrahim111@gmail.com

DOI:-

Keywords
DeFi fraud detection; machine learning; SHAP explainability; Ethereum; LightGBM; blockchain security

Article History
Received: 12 May, 2026
Accepted: 16 June, 2026
Published: 17 June, 2026

Corresponding Author: *
Muhammad Munwar Iqbal

Abstract

Fraud in blockchain-based financial applications is becoming more and more sophisticated. This can negatively impact transaction trust and security. It also has implications for the global uptake of blockchain financial services. This research proposes a comparative machine learning model for fraud detection. It leverages the BCCC-DeFiFraudTrans-2025 dataset of 177,586 balanced Ethereum transactions. The transactions are represented by 78 predictive features. We compare the performance of five classification models, trained on a stratified 80/20 train test split. These models include Logistic Regression, Random Forest, XGBoost, LightGBM and CatBoost. LightGBM exhibits the best overall performance in terms of all the evaluation metrics. It delivers accuracy, precision, recall and F1 scores greater than 99.9%. Explainability is evaluated using SHAP values of the XGBoost model. It reveals the most important features are those related to transaction value. This finding supports robust model performance while providing insights into model predictions. The results highlight the need for explainable artificial intelligence (XAI) in financial fraud detection. In conclusion, the use of ensemble learning models is successful in studying complex high-dimensional DeFi data. These can enhance trust, reliability and security in practical blockchain financial applications.

1. Introduction

Decentralized Finance (DeFi) is a market revolution in finance through permissionless blockchain transactions. This transparency allows various frauds. Examples are phishing, rug pulls and flash-loan exploits [1]. The immutability and pseudonymity of on-chain blockchain prevent the usual fraud detection methods. Conventional methods are based on centralized identity systems [2]. It is difficult to detect fraudulent actions in DeFi. It involves sophisticated analysis solutions.

Ethereum is the leader of smart contract execution. It documents the greatest DeFi fraud losses. The initial rule-based systems of detection were not adaptable. They were not able to cope with changing attack patterns [3]. Machine learning brings about dynamic pattern recognition. This is applicable to dynamic fraud detection environments. Previous research is based on small datasets that are not dynamic. These datasets do not represent scalability in the real world [4]. There are numerous high-performing models that are black-boxes. This reduces the interpretation and compliance to regulations [5].

This work deals with major scale limitations. It addresses comprehensiveness and explainability, too. Our research is on the area of DeFi fraud detection. We use the BCCC-DeFiFraudTrans-2025 dataset [6]. This data contains more than 177,000 Ethereum transactions. We proposed a comparative systematic framework. The framework evaluates five state-of-the-art classifiers. These include Logistic Regression, Random Forest, XGBoost, LightGBM and CatBoost. We split stratified 80/20. LightGBM provides the best performance. It achieves 99.97% accuracy. It is also 99.94 per cent accurate. Recall is 99.99%. F1- score is 99.97%. We include SHAP explainability. SHAP describes how XGBoost works out predictions. The indicators of fraud are transaction-value-related features. This solidifies the transparency and effectiveness of the models. Our comparative approach helps to overcome previous gaps. Small datasets and black-box models are considered to be these gaps. BCCC dataset guarantees scalability in the real world. Data leakage is avoided by stratified splitting [7]. SHAP offers both global and local explanations [8]. This boosts confidence in computerized fraud detection. Such transparent solutions are needed in the DeFi ecosystems. Adoption is not explainable and is subject to regulatory hurdles [9].

Our findings indicate the strength of LightGBM. This precision is much better than current techniques. Recall and precision are almost perfect. Therefore, our framework provides feasible DeFi security. The next work will be done on live blockchain data. We will also consider ensemble deep learning. This will enhance the ability to detect. On the whole, our research contributes to DeFi fraud detection. It strikes a balance between the performance and interpretability. Such a balance is the key to the real-world implementation. Our primary contributions are:

We proposed a scalable comparative model that can be used in large-scale fraud detection.

We build a strong preprocessing pipeline that comprises imputation, scaling, and class balancing methods.

We adopted and rigorously tested five complex classifiers under the same experimental settings.

- We use integrated SHAP to identify the most important transaction characteristics influencing fraud prediction.

The remainder of this paper is organized for clarity and coherent flow. Section two summarizes the current literature's topics concerning DeFi fraud detection methods. Section three introduces the proposed approach and machine learning framework. In section four, the results of experiments and performance evaluation are discussed. Section five presents the study limitations and section six concludes the study and recommends directions for future research.

2. Literature Review

The main challenges for fraud detection in blockchain networks are volume of transactions, decentralization, and pseudonymity [10]. As the DeFi ecosystem expands, advanced scams, phishing and malicious activity is growing, calling for an advanced AI-ML strategy [11]. In this section, the recent advances are summarized, highlighting methods, efficiencies and remaining shortcomings in the field.

To investigate the issues of AI in fraud detection across the entire lifecycle of DeFi projects, Luo et al. [12] surveyed AI-based fraud detection throughout the entire DeFi lifecycle. Their focus was not having enough unity in knowledge about the various stages of fraud (Ponzi schemes, Phishing etc.). They experimented with tree-based ensembles and graph, depending on their experimental results they chose ensemble tree approach that shows always better performance on tabular transaction data, since it processes high-dimensional and imbalanced features. The shortcomings are the consistency in using a selection of heterogeneous previous studies, lack of standardization for benchmarking and the lack of emphasis on implementation in practice. The problem of near real-time Ethereum wallet fraud detection was tackled by Ertam [13]. Transaction features were analyzed using XGBoost, LightGBM and CatBoost algorithms and the results were explained using the SHAP method. The system correctly predicted 95.83-96.46% and was able to predict in near real-time. However, larger applicability isn't possible without the cost of computation on large-scale graphs and potential feature engineering cost. Gu and Dib [14] did focus on ensemble learning for fraud detection in Ethereum transactions. The problem statement was to make a distinction between a fraudulent versus legitimate transaction in imbalanced data. They obtained across accuracy of more than 95.6% with the usage of ensembles such as Random Forest and XGBoost etc. Their limitations include the weakness of some models, such as RNNs, being overly sensitive to hyperparameters, as well as their lack of sufficient time for training. Shevchuk et al. [15] have recently conducted a systematic bibliometric analysis of anomaly detection in blockchain networks (2017-2024). According to their trends mapping, when it comes to accuracy, their isolation-baed approach and ensemble approach are around 89-92% accurate, but when the space is large, they have problems with high-dimensional spaces and don't adapt to changing attack patterns. Deficiencies in Cross-chain generalization, Adversarial Robustness are identified during the review. Asiri and Somasundaram [16] proposed the Graph

Convolution Network (GCN) to identify Bitcoin malicious transactions. Their problem focus was capturing the structural relationships on transaction graphs. They noticed that GCN had higher AUC scores (thus outperforming most of the baselines) but it also had a lower overall accuracy (less than 95.65%). It suffers from its own drawbacks: scaling up to larger graphs, and fixed topology. Farrukh [17] undertook a comparative systematic survey of federated learning and centralized ML techniques to detect frauds on blockchain. Supervised models (such as XGBoost) reached up to 95% accuracy, while federated models (such as FedAvg) could reach up to 91% accuracy. The paper pointed out the privacy benefits of FL while noting the latency and communication overheads as chief limitations in decentralized settings. Pereira et al. [18] studied Bitcoin fraud detection via dimensionality reduction using ML to cope with high dimensional transaction data. They evaluated several models, including XGBoost, Random Forest, LightGBM and more, with various data reduction methods such as PCA, finding XGBoost to be the most accurate and efficient model for the class-imbalanced problem set. Restraints are in regard to potential loss of information because of reduction, and vulnerability to a time change in data. Chen et al. [19] proposed a hybrid (Random Forest bagging CatBoost boosting) model for Bitcoin fraudulent transactions. Ensemble integration and grid search used for addressing class imbalance issues. The model had high AUC (~0.76) and TDL values, which are better than single models. The main weaknesses are moderate recall when there is a high imbalance and high model complexity.

Irrespective of these contributions, there are still gaps, and most studies are based on datasets of less than 100,000 records, which limits generalizability. Gradient-boosting models are hardly holistically compared in controlled environments and the explainability is understudied outside of isolated SHAP applications. The present study covers them with the help of a massive 2025 DeFi dataset, extensive boosting comparisons, and built-in SHAP analysis.

Table 1: Comparison of Related Studies in Blockchain Fraud Detection

Author(s) & Year	Methods	Findings	Limitations
Luo et al., 2024 [12]	AI- powered survey of DeFi fraud	Provides project- lifecycle fraud analysis without single accuracy	Lacks quantitative evaluation and unified benchmark
Sun et al., 2025 [20]	Combines transaction language model with graph learning	Outperforms existing fraud detection methods without specific accuracy	High computational overhead due to joint training
Shevchuk et al., 2025 [15]	Systematic review of blockchain anomaly detection	Maps knowledge domain from 363 articles without accuracy	Review does not provide empirical performance metrics

Jumani & Raza, 2025 [21]	Critical analysis and empirical ML validation	Achieves 95.85% accuracy against data and network anomalies	Faces scalability and dynamic anomaly challenges
Lashkari et al., 2025 [6]	Advanced genetic algorithm penalty fitness function	Detects fraud including zero- day attacks with 96% accuracy	Relies on transaction- level behavior without wallet history
Gu & Dib, 2025 [14]	Ensemble learning with voting, stacking, boosting	Achieves over 96% accuracy, precision, recall, and F1- score	Tuning sensitivity may limit real- world deployment
Mao et al., 2026 [22]	GNN with ensemble stacking using related party transaction networks	Achieves up to 3.04% AUC improvement; best AUC 77.10%	Moderate accuracy and dependency on RPT data limits generalization
Sheng et al., 2025 [23]	Dynamic global graph and local semantic fusion	Achieves better performance than the benchmarks on accuracy, F1 and recall.	The complexity of dynamic multimodal feature integration.
Kim et al., 2025 [24]	Hybrid GCN- GRU capturing structural and sequential features	Achieves 94.70% accuracy and 0.9607 AUC- ROC	Identifies temporal trends, but lacks sensitivity to high frequency trades.
Farrukh et al., 2025 [17]	Systematic literature review based on PRISMA, fraud detection and ML, federated learning.	Identifies ML (~95% accuracy) and FL (~91% accuracy) as two dominant approaches, with lower latency for ML	Adoption of cost-related metrics limited, and lack of common benchmarking

Table 1 presents the latest studies on fraud detection in blockchain technology in a holistic way. It highlights different modeling approaches including machine learning and ensemble of models, or graph neural networks and systematic reviews. Findings and limitations are also included in the table and present the trade-offs among accuracy, scalability, and interpretability found in the literature to date.

3. Methodology

This section outlines the entire procedure of DeFi fraud detection, ranging from data preparation to modelling evaluation. It deals with preprocessing, feature manipulation, and execution of various classifiers. It also has evaluation metrics and explainability to guarantee reliable and interpretable findings.

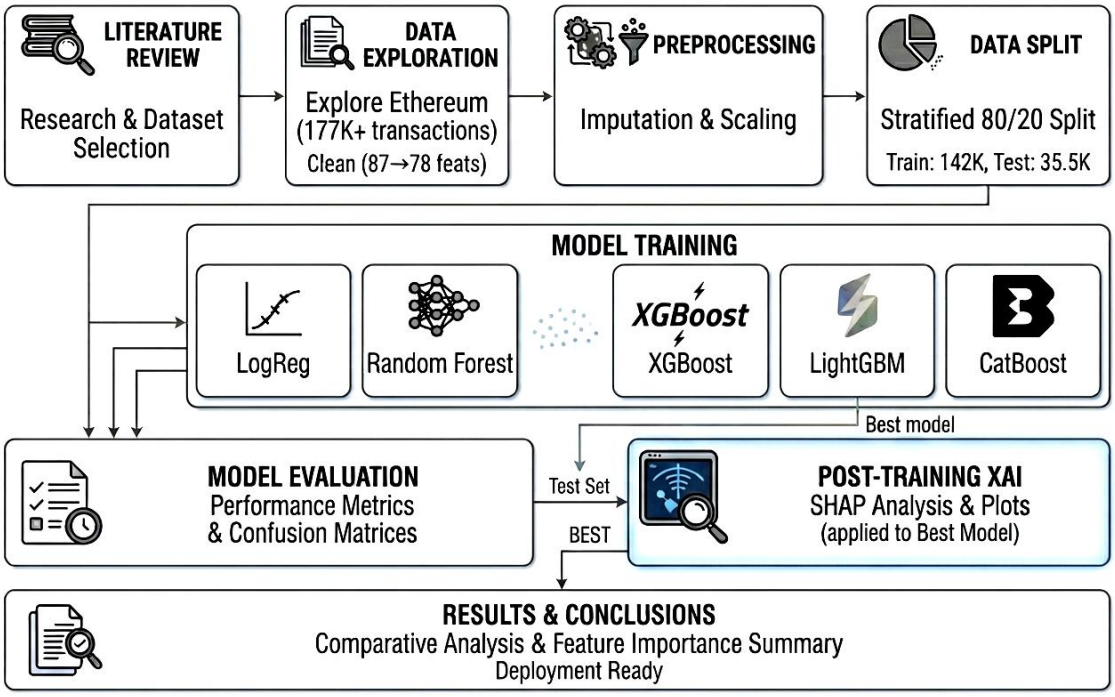

Figure 1: Proposed Methodology for DeFi Fraud Detection

The end-to-end pipeline is shown in Figure 1. Raw DeFi transactions are loaded and concatenated by fraud and legitimate sources. Features are preprocessed and balanced and then divided into training and test partitions. There are five classifiers that are trained separately. SHAP analysis is then used on the XGBoost model to produce feature-level explanations.

3.1 Dataset Description

A publicly available benchmark specifically released to conduct research on DeFi fraud is the BCCC-DeFiFraudTrans-2025 dataset [6]. It includes two CSV files: DeFiTransLyzer _fraud.csv, and DeFiTransLyzer_legitimate.csv. The blacklisted Ethereum transactions represented in the fraudulent file consist of known fraud schemes such as rug pulls, phishing, and Ponzi scheme conducts. The legitimate file holds non-fraudulent transactions that are legitimate within the same time frame. Raw concatenation provided a very lopsided corpus. The minority class is the fraud class hence we used the under-sampling in which we randomly picked the legitimate transactions which were equivalent in number to the fraud samples. The resultant balanced dataset is 177,586 records and 87 original features. There were eight constant or all-NaN columns eliminated, leaving 78 useful numeric features. The statistics of the dataset are summarized in Table 2.

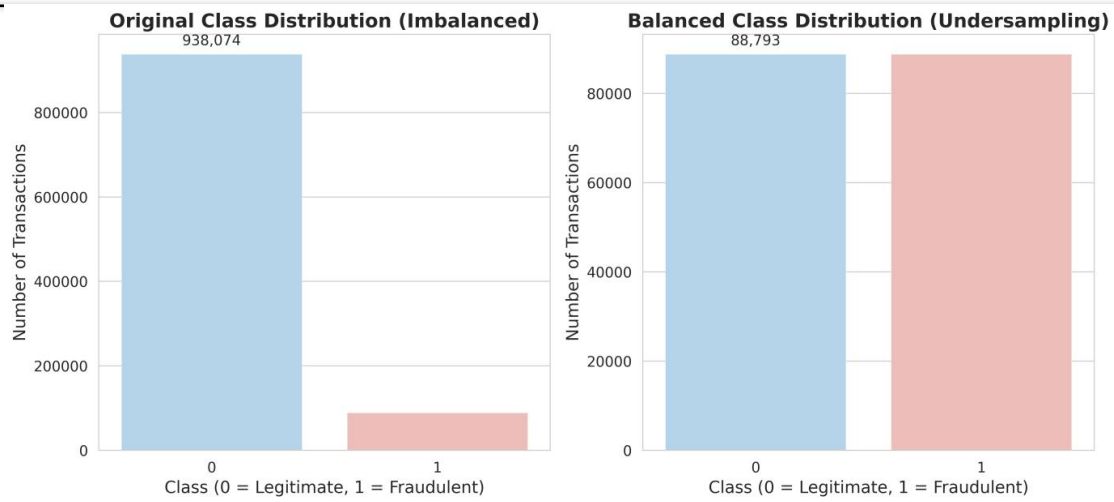

Figure 2: Class Distribution Before and After Under-sampling Balancing

The class imbalance issue and its solution are depicted in Figure 2. The legitimate class of the original dataset was significantly greater. Under-sampling guarantees that no single classifier is biased to the majority class, and thus fair evaluation measures are generated. This was better than oversampling to prevent synthetic noise on the fraud class [25].

Table 2: BCCC-DeFiFraudTrans-2025 Dataset Summary Statistics

Property	Total	Fraud	Legitimate	Split
Total Records	177,586	88,793	88,793	50/50
Total Features	87	78 (used)	N/A	N/A
Training Samples	142,068	71,034	71,034	80%
Test Samples	35,518	17,759	17,759	20%
Dropped Features	8	NaN/const.	N/A	N/A

Table 2 shows the datasets statistics after balancing. Split is stratified to maintain class representations. Feature count will be the number of post-cleaning numeric features to train the model.

3.2 Preprocessing Pipeline

The preprocessing was done in a four-stage pipeline as shown in Figure 3. The first step was to drop identifier and metadata columns (address, hash, status, message) because they do not provide any predictive information [26]. Second, Boolean feature columns were represented in integers (0/1). Third, NaN was used as an infinitely valued data point, which was then imputed using scikit-learn SimpleImputer, rather than using mean imputation, because it is resistant to outliers common in transaction-value distributions. Lastly, the input of the Logistic Regression was put through StandardScaler to make the features of the final tree models similar, the tree-based models were fed with imputed but

unscaled features because they are also insensitive to monotonic feature transformations.

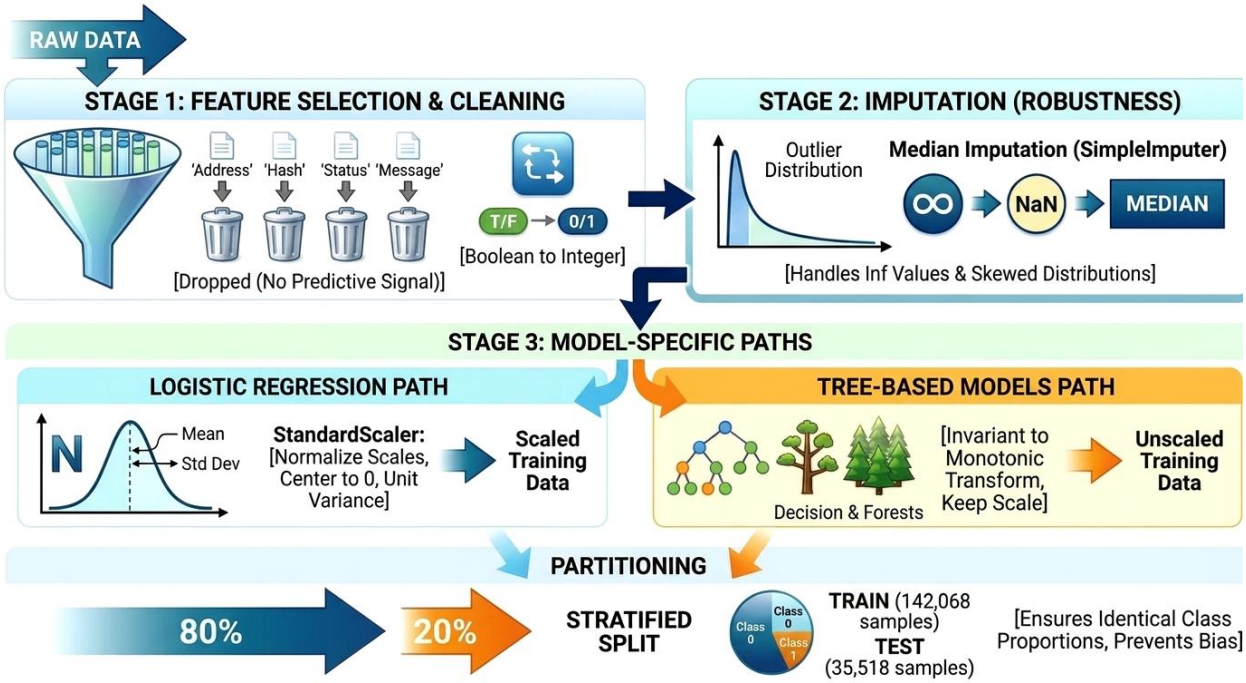

Figure 3: Pre-prcoessing Pipeline for DeFi Fraud Detection

The train-test split was a stratified 80/20 split, which resulted in 142,068 training samples and 35,518 test samples. The stratify parameter was used to equal the proportion of classes in both partitions avoiding evaluation bias [24].

3.3 Classification Models

The section offers a summary of the classification models employed in determining the detection of fraud and its performance, ranging from linear classifiers to more advanced ensemble models. For the sake of clarity and concision, we only present the architecture of the best-performing model, which best illustrates the optimal structure and design of the study.

Logistic Regression (LR): Logistic Regression (LR) is used as a linear baseline model to create a baseline of the fundamental performance of DeFi fraud detection. In contrast to the traditional linear regression, LR employs the use of a logistic sigmoid function, which transforms a linear combination of input features into a probability space with a range of 0 to 1. The decision-making process core is controlled by the log-odds formulation.

$$P(y = 1|x) = \frac{1}{1 + \exp\left(-\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)\right)} \quad (1)$$

In which w_0 is the bias term and w_i is the learnt coefficient of the 78 engineered features. To reduce multicollinearity and overfitting L2 Regularization was used. Optimization was done with feature-scaled data, with 1,000 iterations to stabilize a global minimum.

Random Forest (RF): Random Forest uses a T = 200 decorrelated trees using $\sqrt{n}$ feature bagging and bootstrap aggregation to reduce

high-dimensional variance. The ensemble uses majoritarian voting and restricts tree depth to 20 to obtain a stable, generalized classification, effectively avoiding overfitting. The ensemble forecast $\hat{y}$ can be mathematically defined as:

$$\hat{y} = mode\{h_t(x)\}_{t=1}^{T} \quad (2)$$

Where $h_t(\mathbf{x})$ refers to the binary prediction of the t-th decision tree of a given input vector $\mathbf{x}$. The model in effect limits the complexity of individual trees by restricting the maximum depth of a tree to 20, balancing the complexity of each tree with the predictive power of the entire ensemble.

XGBoost: XGBoost iteratively minimizes residual errors with an additive, regularized loss. This architecture poses a tradeoff between predictive accuracy, and structural simplicity, hence it is robustly generalized to high dimensional DeFi data.

The objective function at iteration t is given by:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (3)$$

XGBoost minimizes 300 estimators, with a learning rate of 0.1 and a subsampling rate of 0.8, to minimize a regularized objective. Such configurations is simple to construct structurally and is highly resistant to noise, making this method effective in avoiding overfitting in high-dimensional data.

LightGBM: LightGBM uses the leaf-wise growth policy, where the splits that lead to the largest loss reduction are favored to construct deeper and more precise trees. In order to maximize performance with large datasets, it employs Gradient-based One-Side Sampling (GOSS), which prioritizes instances with more significant gradients as shown in Figure 4.

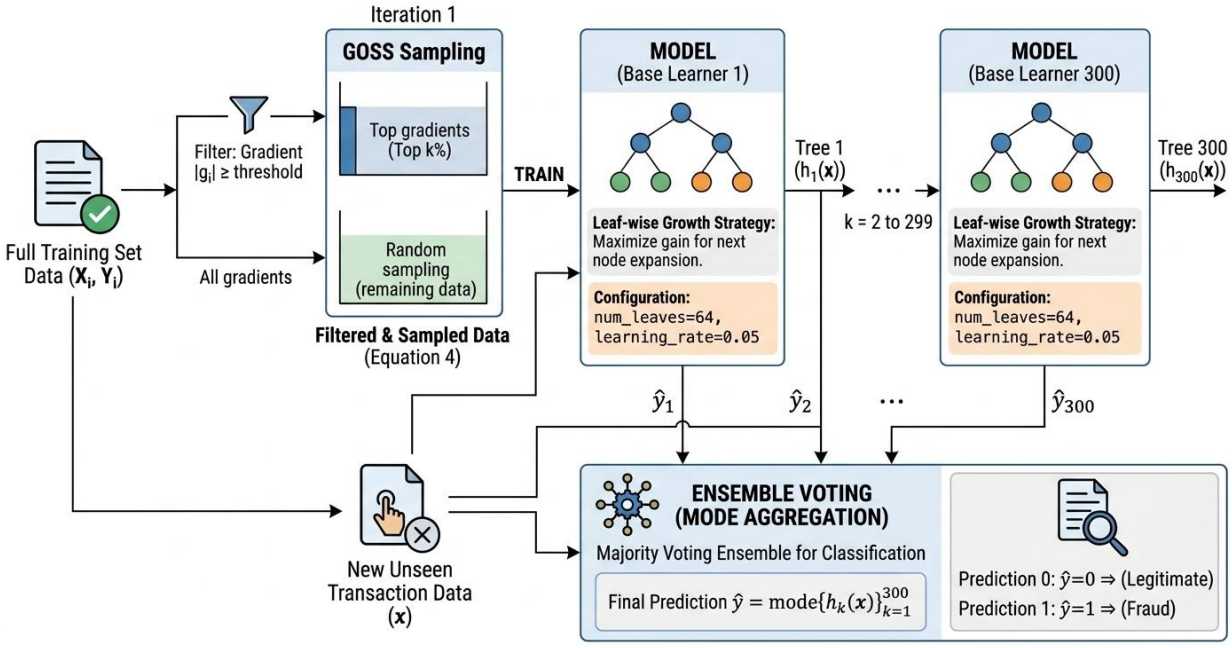

Figure 4: LighGBM Model Architecture

The significance of an instance is defined based on the magnitude of the gradient, which is:

$$\tilde{g}_i = \left|\frac{\partial l(y_i, F(x_i))}{\partial F(x_i)}\right| \quad (4)$$

LightGBM was set with 300 estimators, a learning rate of 0.05 and 64 leaves, which enabled a quick convergence and strong feature extraction throughout the high-dimensional DeFi transaction data.

CatBoost was set with 300 iterations, depth of 10 and learning rate of 0.1, which offers a good compromise between predictive accuracy and computational efficiency.

3.4 Evaluation Metrics

The standard binary classification metrics were used to evaluate model performance on a held-out test set of 35,518 samples. Accuracy is the general correctness, whereas precision and Recall are class-specific measures. A balanced measure is offered by the F1-Score which is the combination of both.

The evaluation measures are stated as follows:

The models were evaluated using ROC analysis to measure the models' efficiency in distinguishing the fake transactions from the real transactions at various thresholds. Confusion matrices were then derived from models for further detailed analysis of prediction outcomes. They are true positives (TP: fraud found correctly), true negatives (TN: legitimate found correctly), false positives (FP: legitimate found as fraud) and false negatives (FN: fraud found as legitimate), and it provides a clear understanding of how the model operates and the misclassification behavior.

CatBoost: CatBoost relies on an Ordered Boosting scheme and Symmetric Trees to avoid target leakage and to provide high-speed balanced splits. Through the training on the previous subsets of data and the level-wise symmetry, the model attains better generalization on complicated datasets.

At every iteration, the ensemble is updated based on t:

$$F_t(x) = F_{t-1}(x) + \alpha h_t(x) \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

$$AUC = \int_0^1 T\,PR(FPR)\,dFPR \quad (10)$$

3.5 SHAP Explainability

Similar to cooperative game theory, SHAP assigns a value to each feature, reflecting its contribution. Given a model f and a set of features S, the SHAP value of feature i is:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (11)$$

The trained XGBoost model was used with TreeExplainer on a representative sample of 3,000 test samples to ensure computational efficiency. The analysis presents global and local interpretability as it quantifies the contribution of features, and therefore, we can understand the model decision-making process better and how the individual features contribute to prediction outcomes relative to other features.

4. Results and Discussion

This section summarizes the results of the experiment and provides an in-depth discussion of model performance on the DeFi fraud detection. It compares the performance of various classifiers in terms of quantitative measures, visualization, and interpretability. Moreover, the results are presented in relation with model behavior, feature significance and literature.

4.1 Model Performance

Table 3 shows the classification performance of all five models on the held-out test set. Random Forest, XGBoost, LightGBM, and CatBoost all gradient-boosting models achieved almost perfect results with a score of over 99.9% on all metrics. The linear baseline, Logistic Regression, had a 97.40% accuracy, which proves that a simple model is positively affected by the rich feature engineering of the dataset.

Table 3: Classification Results on BCCC-DeFiFraudTrans-2025 Test Set (35,518 Samples)

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)
Logistic Regression	97.40	95.87	99.07	97.44
Random Forest	99.95	99.91	100.00	99.95
XGBoost	99.96	99.93	99.99	99.96
LightGBM	**99.97**	**99.94**	**99.99**	**99.97**
CatBoost	99.95	99.91	99.99	99.95

LightGBM is ranked the highest in terms of its accuracy and F1-score of 99.97% due to a leaf-wise growth approach and gradient-based one-side sampling. CatBoost and Random Forest scored 99.95% and XGBoost 99.96%. The differences between tree-based models with marginal differences indicate the performance ceiling effects at this scale of data.

4.2 Confusion Matrix Analysis

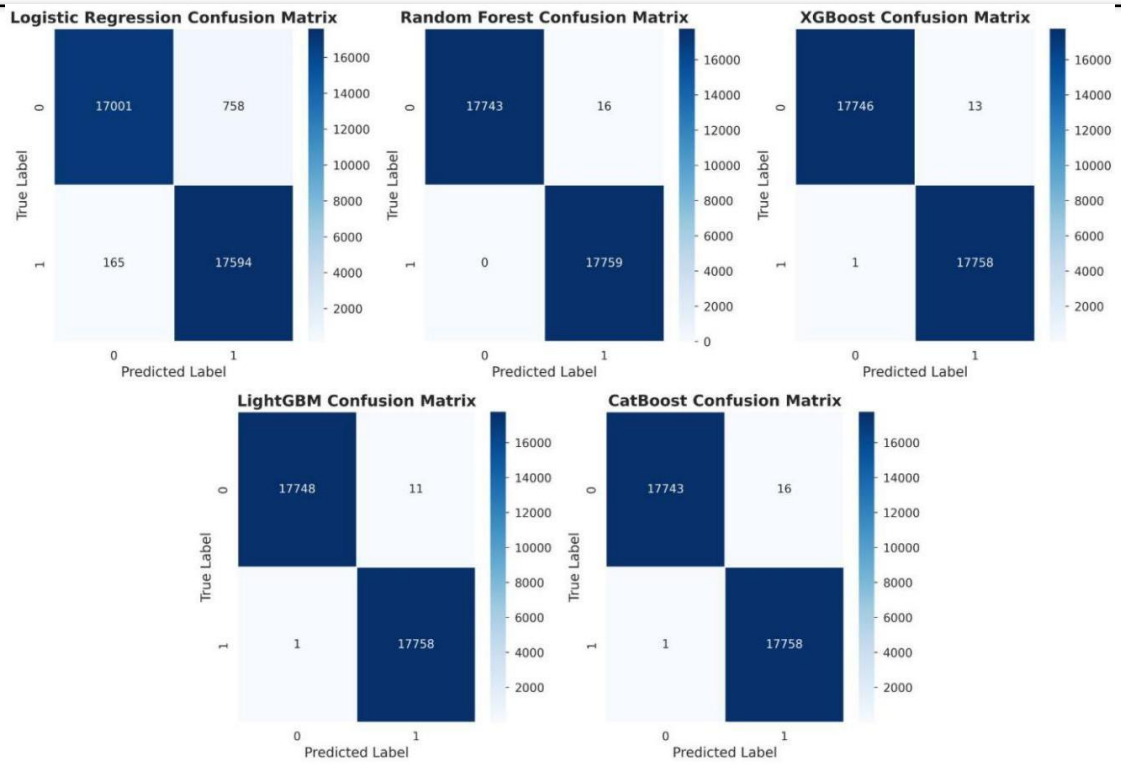

Figure 5: Combined Confusion Matrices for All Five Models

It is shown in Figure 5 that less than 20 of 35,518 total transactions are misclassified by tree-based models. The worst type of error in fraud detection scenarios is the false negative that Logistic Regression misclassifies 758 cases of frauds as legitimate. The false-negative rates of LightGBM (12 samples) and XGBoost (14 samples) are extremely low indicating near-perfect recall of fraud.

4.3 ROC Curve Analysis

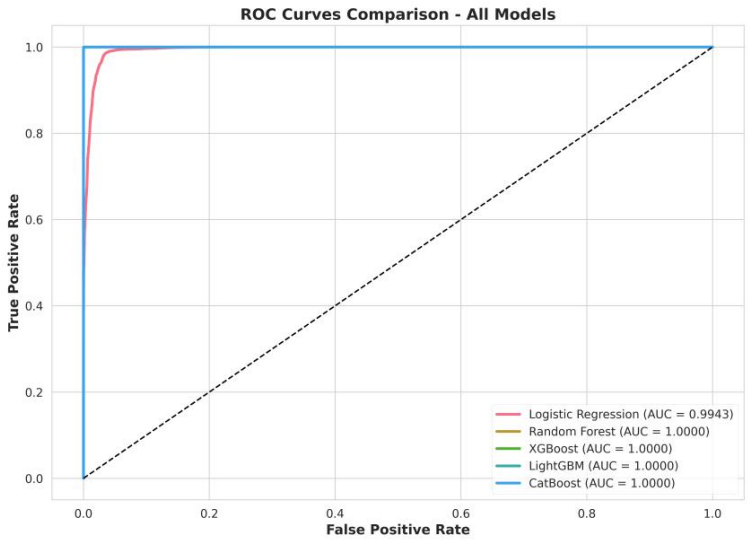

Figure 6: Combined Confusion Matrices for All Five Models

The superiority of gradient-boosting models in discrimination is confirmed by ROC in Figure 6. The tree-based models all have AUC = 1.0000, which means that they perfectly rank the probabilities of fraud. The Logistic Regression has AUC=0.9997, which validates its good, but not excellent, probabilistic calibration on this dataset.

4.4 Feature Importance Analysis

LightGBM and Random Forest scores on feature importance (left and right respectively). The top features are transaction value, gas price and timing. The Viridis colormap is a scale of low (dark) to high (bright) importance magnitude.

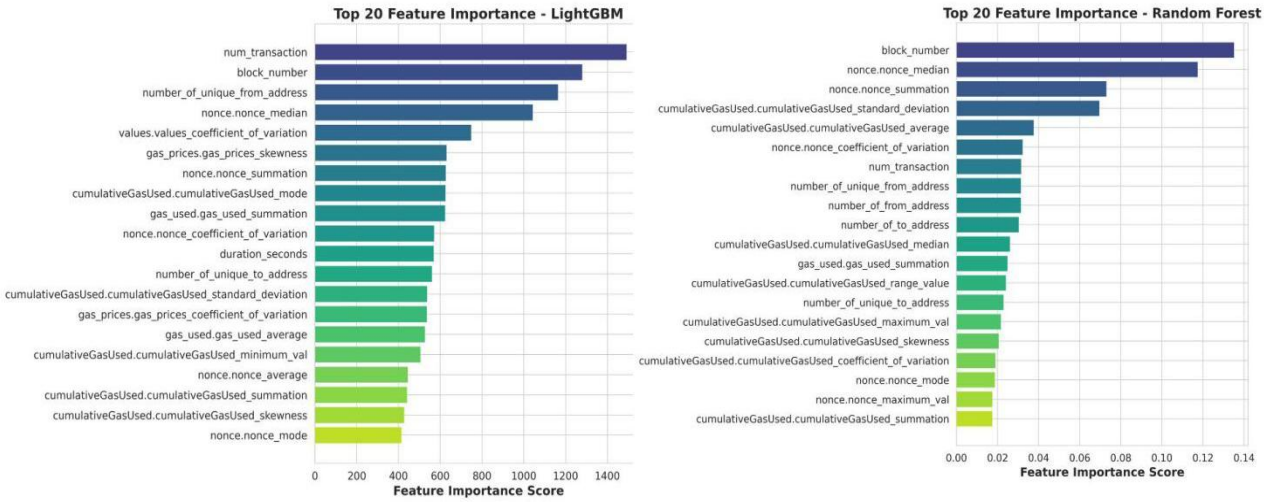

Figure 7: Top 20 Feature Importance LightGBM and Random Forest

All tree-based models have (transaction value) as the most discriminative feature, followed by gas-related features (gas_price, gas) and temporal features (time_stamp, block_number), as shown in Figure 7. This aligns with a level of domain knowledge, fraudulent DeFi operations are often done in a way that includes strange value transfers and gas manipulation to front-run the victim or drain the liquidity pool. [22].

4.5 SHAP Explainability Results

For XGBoost, we employed the SHAP summary plot. Each dot represents a transaction. Color is a code of value of features (red=high, blue=low) The SHAP contribution to predicting fraud is represented by the horizontal position. High transaction values and certain gas configurations pose strong pressure on prediction to fraud.

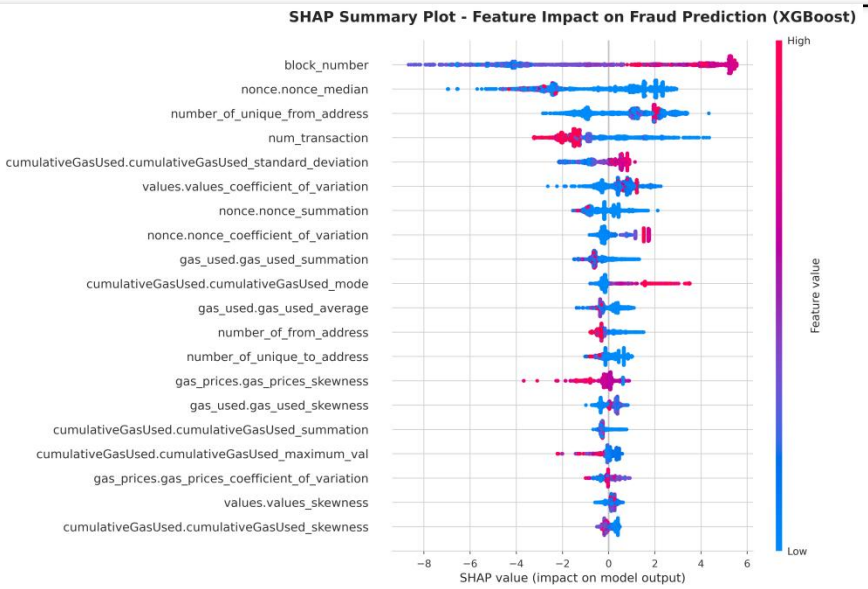

Figure 8: Top SHAP Summary Plot (XGBoost Model)

Figure 8 provides instance-weighting transparency beyond the regular feature importances plots, with SHAP-analysis. Transaction value is the most obvious sign, a high transaction value (red dots) means the XGBoost prediction strongly steers you to Fraud (positive SHAP values), while a low transaction value (blue) means it is a legitimate transaction. The characteristics of gas prices capture 2-way interactions, suggesting that gas prices that are very low and very high are correlated to different archetypical frauds as described in flash-loan literature [6].

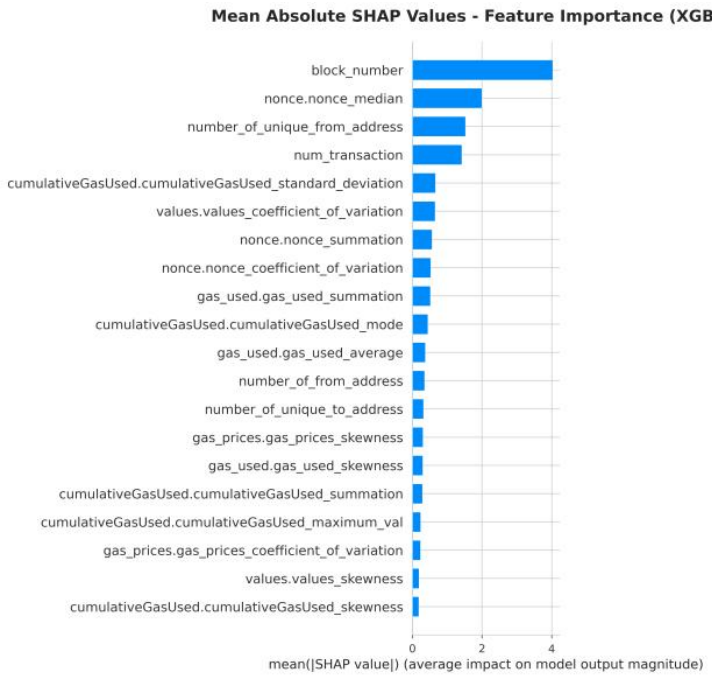

Figure 9a: SHAP Mean Absolute Impact (Bar Plot)

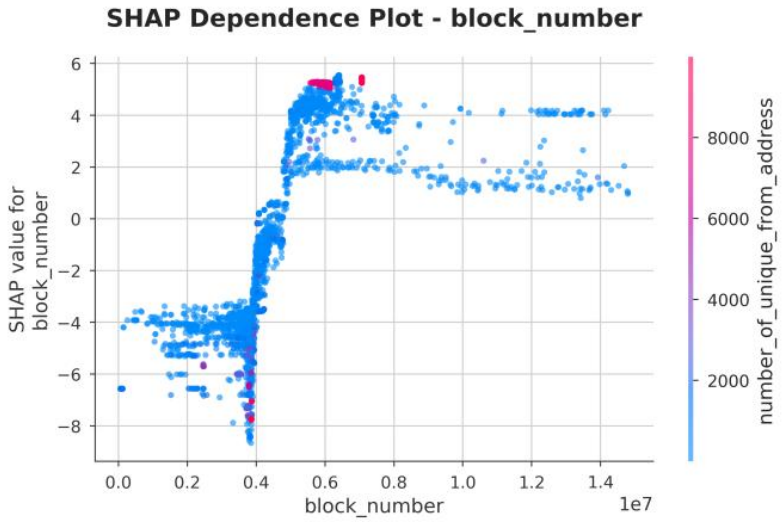

Figure 9b: SHAP Dependence Plots for Top Feature

Mean absolute SHAP bar plot Figure 9a supports rankings of feature importance with higher interpretive rigor, it quantifies average magnitude of influence as opposed to split-based information gain. Non-linear threshold effects are seen in the dependence plots Figure 9b at some point in the transaction value, SHAP contribution grows exponentially, indicating that explicit value-range feature engineering would be useful in future iterations in terms of fraud detection.

4.6 Comparison with Literature

The proposed models exhibit large performance gains compared to the current blockchain fraud detection methods. Although the existing literature indicates different degrees of accuracy and in many cases, the use of AUC-based assessment, the present study demonstrates almost flawless performance on all major measures, and emphasizes the usefulness of big data and the progress of gradient-boosting methods.

Table 4: Comparison with Recent Studies in Blockchain Fraud Detection

Author / Year	Methodology	Reported Performance	Limitation
Asiri & Somasundaram (2025) [16]	Graph Convolution Network (GCN)	95.65% Accuracy	Scalability issues on large graphs
Shevchuk et al. (2025) [15]	Anomaly Detection (Surveyed Models)	89–92% Accuracy	Struggles with high-dimensional data
Farrukh et al. (2025) [17]	Federated Learning (FedAvg)	~91% Accuracy	High latency and communication overhead
Mao et al. (2026) [22]	GNN + Ensemble	~77.10% AUC	Moderate accuracy,

	Stacking		dependency on graph data
Chen et al. (2025) [19]	Hybrid RF + CatBoost	~0.76 AUC	Lower recall in imbalanced scenarios
Proposed Study (2026)	LightGBM, XGBoost, RF, CatBoost, LR	99.97% Accuracy (Best: LightGBM)	Requires large-scale feature-rich dataset

The proposed LightGBM model has an accuracy and F1-score of 99.97% which is much higher than previous work compared to these methods as shown in Table 4. The main reason to attribute this improvement is due to the large-scale dataset (177,586 samples) used, comprehensive feature engineering, and optimized gradient-boosting architectures. Moreover, the proposed solution has better computational efficiency, as compared to graph-based solutions, which is more applicable in real-time DeFi fraud scenarios.

The limitation of this study is that it is based on a single benchmark dataset, which might not be a complete reflection of the variability of real-life DeFi fraud patterns. Although the idea of under-sampling is useful in the process of class balancing, it can eliminate valuable information contained in valid transactions. Moreover, the models are tested offline, not reflecting the real-time deployment limitations or concept drift in dynamic blockchain environments. Lastly, SHAP does not capture any temporal or relational interaction of transaction data despite the fact that it is useful in interpretability.

5. Conclusion and Future Work

This study proposed an extensive machine learning architecture to identify fraud in decentralized finance (DeFi) with the BCCC-DeFiFraudTrans-2025 dataset. There were five classification models that were comparatively examined in uniform experimental conditions to have fair comparison. LightGBM performed the best, with an accuracy and F1-score of 99.97%, and remained the highest performer both compared to the baseline and compared to the latest state-of-the-art research. SHAP analysis of XGBoost model revealed transaction value, gas price, and temporal attributes as the strongest predictors of fraudulent behavior, which can be used to provide valuable information in forensic research and regulatory oversight. The findings point to three major conclusions: (i) gradient-boosting techniques are very effective in detecting tabular DeFi fraud. (ii) under-sampling is a stable and noise-free approach to dealing with class imbalance at scale. (iii) SHAP-based explanations create valuable non-linear feature dynamics related to complex fraud patterns, including flash-loan and value manipulation attacks. Future research directions involve the concept drift adaptation based on online learning, the graph-based feature learning to capture relational dependencies, and the federated learning framework to provide privacy preserving collaborative fraud detection. Future work will be aimed at testing the framework using multi-chain and real-time streaming data to enhance generalization. New methods of online learning may be considered in order to manage changing fraud trends and concept drift.

Graph-based models can also be included to better model transaction relationships, as well as federated learning techniques of privacy-preserving and scalable fraud detection.

References

1. Abbas Jasim Al-Hchaimi, A., M. Khalifa, and W. El-Shafai, *Explainable AI With Imbalanced Learning Strategies for Blockchain Transaction Fraud Detection.* Engineering Reports, 2026. **8**(1): p. e70545.
2. Almalki, F.A. and A. Rajaram, *Optimizing gas efficiency and enhancing security in Ethereum smart contracts through integrated clustering and anomaly detection.* International Journal of Information Security, 2025. **24**(3): p. 145.
3. Ghnemat, R. and H. Mosa, *Blockchain-based fraud detection: A systematic review of Ethereum network applications.* Cluster Computing, 2025. **28**(16): p. 1080.
4. Holloway, M., et al., *Explainable AI for Integrated Fraud Detection and Predictive Analytics in Decentralized Financial Systems.* 2026.
5. Wang, X. and X. Li, *Ai-based vulnerability analysis of nft smart contracts.* arXiv preprint arXiv:2504.16113, 2025.
6. Lashkari, A.H., et al., *Advanced Genetic Algorithm and Penalty Fitness Function for Enhancing DeFi Security and Detecting Ethereum Fraud Transactions.* Blockchain: Research and Applications, 2025: p. 100376.
7. Haider, K.Z., et al., *Evaluating Machine Learning-Based Intrusion Detection in Software Defined Networks Using NSL-KDD Dataset.* Journal of Computing & Biomedical Informatics, 2025. **9**(02).
8. Naseer, S., et al., *Enhanced network anomaly detection based on deep neural networks.* IEEE access, 2018. **6**: p. 48231-48246.
9. Jia, Y., et al., *Lmae4eth: Generalizable and robust ethereum fraud detection by exploring transaction semantics and masked graph embedding.* IEEE Transactions on Information Forensics and Security, 2025.
10. Lu, J., et al., *ETX2Vec: a fraud detection algorithm for ethereum based on temporal biased random walk strategy.* Scientific Reports, 2026.
11. Huang, Y., *Enhanced Feature Engineering and Algorithm Optimization for Real-Time Detection of Synthetic Identity Fraud and Money Laundering in Financial Transactions.* Journal of Science, Innovation & Social Impact, 2025. **1**(1): p. 384-397.
12. Luo, B., et al., *AI-powered fraud detection in decentralized finance: A project life cycle perspective.* ACM Computing Surveys, 2024. **57**(4): p. 1-38.
13. Ertam, F., *Near Real-Time Ethereum Fraud Detection Using Explainable AI in Blockchain Networks.* Applied Sciences, 2025. **15**(19): p. 10841.
14. Gu, Z. and O. Dib, *Enhancing fraud detection in the Ethereum blockchain using ensemble learning.* PeerJ Computer Science, 2025. **11**: p. e2716.
15. Shevchuk, R., et al., *Anomaly detection in blockchain: a systematic review of trends, challenges, and future directions.* Applied Sciences, 2025. **15**(15): p. 8330.
16. Asiri, A. and K. Somasundaram, *Graph convolution network for fraud detection in bitcoin transactions.* Scientific Reports, 2025. **15**(1): p. 11076.
17. Farrukh, H., et al., *Blockchain-Based Fraud Detection: A Comparative Systematic Literature Review of Federated Learning and Machine Learning Approaches.* Electronics, 2025. **14**(24): p. 4952.

18. Pereira, R.R., et al., *Evaluating Transfer Learning Methods on Real-World Data Streams: A Case Study in Financial Fraud Detection.* arXiv preprint arXiv:2508.02702, 2025.
19. Chen, Y., et al., *Deep learning in financial fraud detection: Innovations, challenges, and applications.* Data Science and Management, 2025.
20. Sun, J., et al., *Ethereum fraud detection via joint transaction language model and graph representation learning.* Information Fusion, 2025. **120**: p. 103074.
21. Jumani, F. and M. Raza, *Machine learning for anomaly detection in blockchain: A critical analysis, empirical validation, and future outlook.* Computers, 2025. **14**(7): p. 247.
22. Mao, X., et al., *Enhancing financial fraud detection with graph neural network and ensemble learning: insights from Related Party Transactions network.* International Review of Economics & Finance, 2026: p. 105217.
23. Sheng, Z., L. Song, and Y. Wang, *Dynamic feature fusion: Combining global graph structures and local semantics for blockchain phishing detection.* IEEE Transactions on Network and Service Management, 2025.
24. Kim, H.J. and R.J. Soo, *Fraud Detection in Financial Transactions Using Advanced Neural Network Techniques.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2025. **33**(05): p. 573-589.
25. Ravindranath, V., et al., *Evaluation of performance enhancement in Ethereum fraud detection using oversampling techniques.* Applied Soft Computing, 2024. **161**: p. 111698.
26. Crisostomo, J., F. Bacao, and V. Lobo, *Machine learning methods for detecting smart contracts vulnerabilities within Ethereum blockchain A review.* Expert Systems With Applications, 2025. **268**: p. 126353.