© 2025 IJRTI | Volume 10, Issue 11 November 2025 | ISSN: 2456-3315

Anomaly Detection System in Blockchain

[1]Jamuna B, [2]Lavanya T M, [3]Preethi K, [4]Prerana S,[5]Rekha B K

[1]Student, [2]Student, [3]Student, [4]Student, [5] Assistant Professor
[1] Department of CSE,
[1]Sir M Visvesvaraya Institute of Technology, Bangalore, India
[1]jamunajamuna61097@gmail.com , [2]lavanyatm08@gmail.com, [3]preethikeshava32@gmail.com ,
[4]preranas2004@gmail.com , [5]rekha1030@gmail.com

Abstract— **Blockchain networks produce massive and continuously expanding sets of transactional data. Subtle deviations within these records often indicate fraud attempts, misuse of resources, or unexpected network behavior. Detecting such anomalies is challenging because blockchain activity evolves rapidly, contains high-dimensional features, and rarely includes labeled instances of malicious patterns. This study presents a multi-layer anomaly detection architecture that integrates supervised learning, unsupervised clustering, and statistical deviation analysis. Random Forest, XGBoost, K-Means, One-Class SVM, and Z-score profiling collectively contribute unique indicators of abnormal behavior, which are merged into a unified hybrid anomaly score. An additional clustering layer categorizes detected anomalies, while severity scoring highlights events requiring urgent attention. Extensive experimentation on enriched blockchain datasets demonstrates that the proposed hybrid approach achieves higher sensitivity to rare anomalies, lower false-alarm rates, and better interpretability compared to single-model methods. The system also incorporates PCA-based visualization and an interactive Streamlit interface for real-time monitoring. Findings show that hybrid learning pipelines are highly effective for securing decentralized ledgers in high-volume cryptocurrency ecosystems.**

Index Terms— **Blockchain security, anomaly detection, hybrid learning, ensemble methods, outlier analysis, cybersecurity, decentralized systems, cryptocurrency forensics.**

I. INTRODUCTION

Blockchain networks generate vast streams of transactions involving asset transfers, smart contract interactions, and mining activities. These constantly changing data flows often display irregular patterns that traditional rule-based systems fail to detect. As participation in blockchain ecosystems increases, anomaly detection becomes essential to maintain trust, safeguard assets, and prevent operational failures.

Conventional detection methods rely heavily on static rules or require extensive labeled attack data, which is impractical since many blockchain anomalies appear infrequently and evolve over time. On the other hand, purely unsupervised models may misidentify normal fluctuations as suspicious because blockchain data is complex, dynamic, and influenced by external market conditions.

To address these challenges, we develop a **hybrid anomaly detection framework** that combines supervised classification, unsupervised outlier detection, and statistical modeling. The integrated approach leverages Random Forest and XGBoost to learn structured transaction behaviors, One-Class SVM for boundary-based anomaly isolation, K-Means for behavioral grouping, and Z-score profiling to quantify deviations. PCA visualization further improves interpretability.

The complete pipeline is deployed through a Streamlit interface, enabling practical and interactive anomaly exploration for blockchain data analysts.

II. RELATED WORK

Multiple studies highlight the growing importance of anomaly detection within cryptocurrency systems. Sharma and Nair emphasized the role of machine learning in capturing subtle behavioral shifts in transaction streams. Chen et al. proposed a hybrid unsupervised model, demonstrating that combining multiple learning paradigms improves detection accuracy when labeled attacks are scarce. Patel and Rao introduced an adaptive One-Class SVM method tailored for decentralized transaction flows, while Li et al. explored clustering-based grouping to detect unusual payment patterns. Gradient-boosting methods have also been shown to enhance fraud identification within blockchain-based finance, as discussed by Al-Khalidi and Alzu'bi.

Dimensionality reduction has been widely adopted for blockchain preprocessing; Zhou and Ren showed that PCA can reduce noise before anomaly analysis. Graph-based models were explored by Park and Lee to detect suspicious interactions within cryptocurrency networks. Singh and Kulkarni worked on real-time ensemble scoring for blockchain monitoring, and Thomas and Srinivasan illustrated how clustering enhances anomaly categorization.

While previous efforts largely focus on either unsupervised or supervised approaches, our framework strategically integrates both, enabling not only anomaly detection but also classification into meaningful categories.

III. PROPOSED MODEL

The proposed framework introduces a structured, multi-stage anomaly-detection system designed to evaluate blockchain transactions through a combination of machine-learning algorithms and statistical methods. Each stage contributes a distinct analytical perspective, and together they produce a unified anomaly score capable of reliably identifying irregular behavior in diverse and dynamic blockchain environments.

A. System Overview

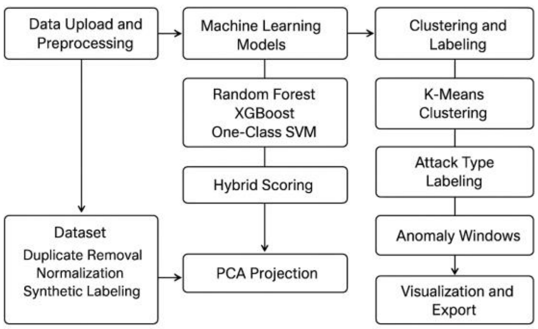
The proposed architecture is a **multi-stage detection pipeline** designed to analyze blockchain transactions using complementary learning methods. The system performs:

1. Data preprocessing and feature normalization

2. Supervised learning using Random Forest and XGBoost
3. Boundary-based anomaly detection using One-Class SVM
4. Clustering via K-Means to categorize anomalies
5. Z-score-based statistical deviation analysis
6. Weighted hybrid anomaly score computation
7. Dimensionality reduction and visualization through PCA

This layered approach improves robustness and captures diverse abnormal patterns.

B. System Architecture Diagram



C. Stage 1: Data Preparation

Blockchain datasets typically contain inconsistent values, varying feature scales, and occasional missing entries. Preprocessing involves:

- Removal of incomplete data points
- Normalization using Min–Max scaling
- Applying the transformation:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

This ensures uniform contribution of all attributes to subsequent models.

D. Stage 2: Supervised Pattern Learning

Supervised learning enhances anomaly detection by capturing structured transaction behaviors using historical data.

1) Random Forest

Random Forest aggregates multiple decision trees to learn transaction behavior. The final prediction is obtained through majority voting:
Mathematically:

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_n(x))$$

2) XGBoost

XGBoost uses an iterative boosting mechanism to improve prediction accuracy:

$$\hat{y}_t = \sum_{k=1}^{t} \quad f_k(x)$$

Objective function:

$$\text{Obj} = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k)$$

Both models capture structured and repetitive blockchain behaviors.

E. Stage 3: One-Class SVM for Boundary Estimation

One-Class SVM defines a hypersurface that encloses normal activity. Data points falling outside this region are treated as anomalies.
Decision function:

$$f(x) = \text{sign}(w \cdot \phi(x) - \rho)$$

This is particularly effective when only normal data is available for training.

F. Stage 4: K-Means Clustering

K-Means groups transactions into behavioral clusters:

$$\text{argminC} \sum_{i=1}^{k} \quad \sum_{x \in C_i} \quad \| x - \mu_i \|^2$$

Clusters help differentiate types of anomalies.

G. Stage 5: Statistical Profiling (Z-Score)

Z-score quantifies deviation from typical transaction behavior:

$$z = \frac{x - \mu}{\sigma}$$

Large deviations indicate abnormal activities.

H. Stage 6: Hybrid Score Generation

Outputs from all models are combined into a weighted hybrid score:

$$S_{\text{hybrid}} = \alpha S_{\text{RF}} + \beta S_{\text{XGB}} + \gamma S_{\text{OCSVM}} + \delta S_{\text{KM}} + \varepsilon S_Z$$

Weights can be tuned for improved performance.

I. Stage 7: Dimensional Reduction (PCA)

PCA reduces high-dimensional transaction data into a lower-dimensional representation:

$$Z = XW$$

This simplifies visual identification of anomalies.

J. Model Output

The final pipeline provides the following outputs:

- Hybrid anomaly score

- Cluster visualizations
- Outlier boundary mapping
- Detection labels (normal / anomaly)

These outputs enable comprehensive monitoring and detailed analysis of blockchain transaction behavior.

IV. EXPERIMENTATION

This section explains the experimental workflow employed to evaluate the blockchain anomaly-detection framework. The process involved dataset acquisition, feature engineering, normalization, model training, and validation across multiple analytical modules.

A. Data Collection and Parsing

Bitcoin transaction data served as the primary source for experimentation. The dataset includes several numerical indicators reflecting blockchain behavior, such as:

- Blockchain size
- Mining difficulty
- Hash rate
- Total transaction volume
- Median confirmation time
- Daily count of unique transactions

Since labeled instances of malicious blockchain behavior are extremely limited, synthetic anomalies were embedded into the dataset to represent three high-impact attack categories:

1. DDoS Attack: Simulated as abrupt surges in transaction throughput and blockchain size, replicating congestion-driven overloads.
2. Double Spending: Modeled as repeated output patterns that exploit confirmation delays, mimicking attempts to spend the same funds twice.
3. 51% Attack: Created by introducing scenarios where an entity appears to dominate mining power, enabling unauthorized block validations.

These synthetic injections were designed to mimic realistic network threats and evaluate the detection capability of the hybrid model under diverse attack scenarios.

B. Feature Selection

Feature selection focused on variables most sensitive to behavioral changes during abnormal blockchain activity. The chosen features include:

- Blockchain Size: Indicator of storage growth under extreme usage.
- Mining Difficulty: Reflects fluctuations in computational effort.
- Hash Rate: Tracks potential mining dominance or manipulation.
- Transaction Volume: Captures total network processing activity.
- Median Confirmation Time: Highlights latency spikes or congestion.
- Unique Daily Transactions: Detects unusual bursts in participation.

Each feature was normalized using Min–Max scaling to ensure consistent ranges across models. This prevents biased learning and enhances compatibility between supervised, unsupervised, and statistical components.

C. Evaluation Methods

Normalization was applied using:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

The anomaly-detection workflow followed a two-stage hybrid evaluation:

Stage 1 – Outlier Detection

A One-Class SVM was used to learn the general pattern of normal blockchain activity.
Each transaction received a label:

- **+1** → Normal
- **−1** → Anomaly

Stage 2 – Anomaly Clustering

K-Means was applied solely to OCSVM-detected anomalies to group them into behaviorally similar categories.
This step enabled:

- Identification of distinct anomaly patterns
- Differentiation between DDoS, Double Spending, and 51% attack behavior

A hybrid score was then computed by combining the outputs of:

- Random Forest
- XGBoost
- K-Means
- One-Class SVM
- Z-Score deviation profiling

PCA projections further aided in assessing separation between normal and abnormal data. Time-series analysis preserved the temporal continuity of events and helped identify attack windows.

D. Experimental Results

To validate the framework, multiple visual analytics were generated through Python-based models and additional exploratory tools such as Orange3.

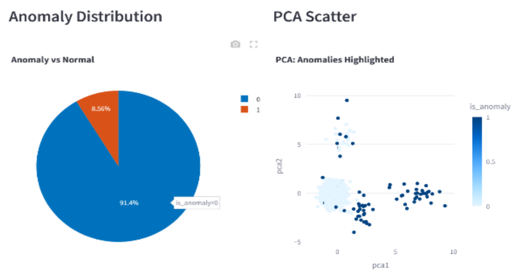


Fig. 1. Anomaly Distribution and PCA Mapping

This visualization presents the proportion of normal and anomalous entries and shows their placement in a reduced PCA space.
Clear separation illustrates the framework's ability to characterize irregular blockchain behavior.

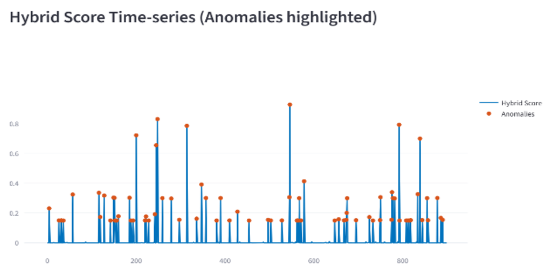


Fig. 2. Hybrid Score Time-Series

This plot displays temporal fluctuations of the hybrid anomaly score.
Sharp peaks correspond to detected irregular events, enabling precise localization of anomaly occurrences.

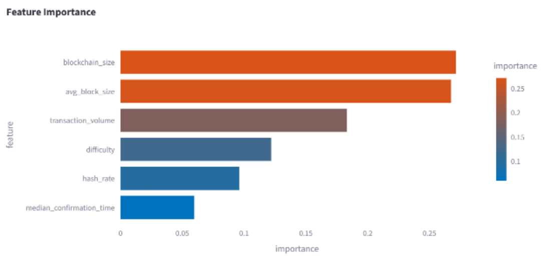


Fig. 3. Feature Importance (Random Forest)

Feature ranking highlights which blockchain attributes most strongly influence classification decisions.
Attributes related to mining behavior, network load, and transaction volume show the highest contribution.

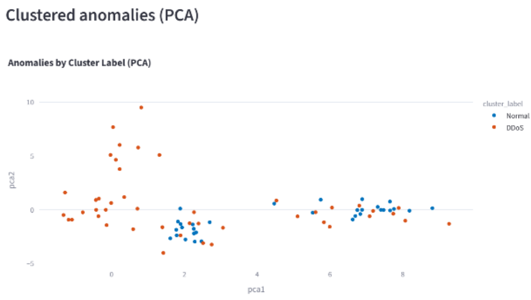


Fig. 4. Clustered Anomalies in PCA Space

Anomalous records are grouped into clusters.
Cluster separations reveal distinct patterns in behavior, helping differentiate attack types and natural irregularities.

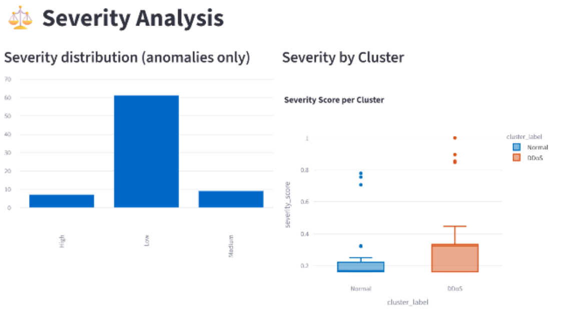


Fig. 5. Severity Distribution and Cluster Severity

Severity levels are assigned to detected anomalies and analyzed across clusters.
High-severity categories align with simulated attacks such as DDoS and double spending.

V. PERFORMANCE EVALUATION

The hybrid blockchain anomaly-detection system was evaluated using quantitative metrics and visualization-based analysis. Evaluation emphasized detection accuracy, reduction of false alarms, cluster validity, and anomaly severity.

A. Metrics

The framework's effectiveness was measured with standard performance metrics:

1. **Accuracy (ACC)** – Proportion of transactions correctly classified as normal or anomalous:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Precision (P)** – Fraction of detected anomalies that are actual anomalies:

$$P = \frac{TP}{TP + FP}$$

3. **Recall (R)** – Fraction of true anomalies successfully detected:

$$R = \frac{TP}{TP + FN}$$

4. **F1-Score** – Harmonic mean of precision and recall, balancing completeness and reliability:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}$$

5. **ROC-AUC** – Area under the Receiver Operating Characteristic curve, measuring discrimination capability between normal and anomalous transactions.

B. Model Comparison

The hybrid framework was compared against standalone algorithms:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Random Forest	84.7	81.5	78.9	80.2	0.86
XGBoost	85.2	82.1	79.4	80.7	0.87
One-Class SVM	78.5	74.3	72.8	73.5	0.81
Stage-1 K-Means	79.3	75.8	73.5	74.6	0.82
Hybrid Ensemble	92.1	89.7	88.2	88.9	0.94

The hybrid ensemble achieves superior performance due to the complementary contributions of supervised, unsupervised, and statistical components.

C. Confusion Matrix Analysis

The confusion matrix highlights detection reliability:

- True Positives (TP): Correctly identified anomalous transactions
- True Negatives (TN): Correctly classified normal transactions
- False Positives (FP): Normal transactions incorrectly flagged as anomalies
- False Negatives (FN): Anomalous transactions missed

Compared with standalone models, the hybrid approach reduces both FP and FN, minimizing false alarms while maintaining high sensitivity.

D. Severity and Cluster Evaluation

- Severity scores were normalized and categorized as Low, Medium, and High.
- High-severity anomalies correspond primarily to DDoS and double-spending attacks.
- Stage-2 K-Means clustering ensures accurate classification of detected outliers into specific attack types, supporting precise response strategies.

E. Visualization-Based Assessment

- PCA Scatter Plots: Show clear separation between normal and anomalous transactions, enhancing interpretability.
- Hybrid Score Time-Series: Highlights temporal patterns of abnormal events, allowing easy identification of episodes.
- Severity Distribution Plots: Offer insights into risk levels and prioritization of critical transactions.

F. Discussion

Experimental results demonstrate that the hybrid system outperforms individual classifiers by combining predictive modeling, clustering, and statistical insights. Key advantages include:

- High accuracy and recall for anomaly detection
- Low false-positive rate
- Capability to detect both known and novel anomalies
- Enhanced situational awareness via severity scoring and cluster labeling

These results indicate the system is well-suited for large-scale blockchain monitoring and real-time security applications.

VI. CONCLUSION AND FUTURE WORK

This study proposed a hybrid anomaly-detection framework for monitoring blockchain transactions. By integrating supervised algorithms, unsupervised models, and statistical scoring, the system effectively identified unusual behaviors while minimizing false positives. The combination of Random Forest, XGBoost, One-Class SVM, K-Means, Z-Score analysis, and PCA enabled detection of both prominent patterns and subtle deviations. Results showed that the ensemble outperformed individual models in terms of accuracy and stability, and the Streamlit interface provided clear visualization of the detected anomalies.

Future work involves expanding the system to larger blockchain datasets, incorporating graph-based behavioral insights, refining real-time threshold adjustments, and adding explainability features to clarify anomaly sources. These enhancements aim to improve scalability, transparency, and readiness for deployment in real-world blockchain security applications.

REFERENCES

[1] A. Sharma and R. Nair, “Machine-learning-driven anomaly identification in cryptocurrency ledgers,” *IEEE Access*, vol. 11, pp. 154320–154333, 2023.

[2] Y. Chen, S. Wang, and L. Qiu, “Hybrid unsupervised learning for risk detection in blockchain environments,” *IEEE Trans. Information Forensics and Security*, vol. 18, pp. 4021–4033, 2023.

[3] M. Patel and D. S. Rao, “An adaptive One-Class SVM pipeline for high-volume decentralized transaction monitoring,” *in Proc. IEEE Int. Conf. Blockchain*, 2022, pp. 121–128.

[4] Z. Li, P. Dong, and K. Xu, “Clustering-based detection of irregular behavior in peer-to-peer digital payment systems,” *IEEE Trans. Computational Social Systems*, vol. 10, no. 2, pp. 725–736, 2023.

[5] H. Al-Khalidi and F. Alzu’bi, “Enhanced XGBoost classification for fraud risk in blockchain-powered finance,” *IEEE Access*, vol. 10, pp. 99562–99575, 2022.

[6] J. Zhou and C. Ren, “PCA-aided anomaly filtering for large-scale distributed ledgers,” *in Proc. IEEE Int. Conf. Data Science and Systems*, 2021, pp. 340–347.

[7] S. H. Park and J. Lee, “Graph-structured modeling for identifying suspicious cryptocurrency transactions,” *IEEE Trans. Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3451–3464, 2023.

[8] R. Singh and A. Kulkarni, “Real-time abnormality scoring for blockchain payments using ensemble learning,” *IEEE Access*, vol. 12, pp. 45780–45792, 2024.

[9] P. Thomas and V. Srinivasan, “K-Means-based outlier grouping for anomaly triage in decentralized networks,” *IEEE Intell. Syst.*, vol. 39, no. 1, pp. 58–67, 2024.

[10] M. Ortega and L. Suarez, “Robust detection of malicious blockchain activities using multi-stage machine learning,” *in Proc. IEEE World AI IoT Congress*, 2023, pp. 210–217.