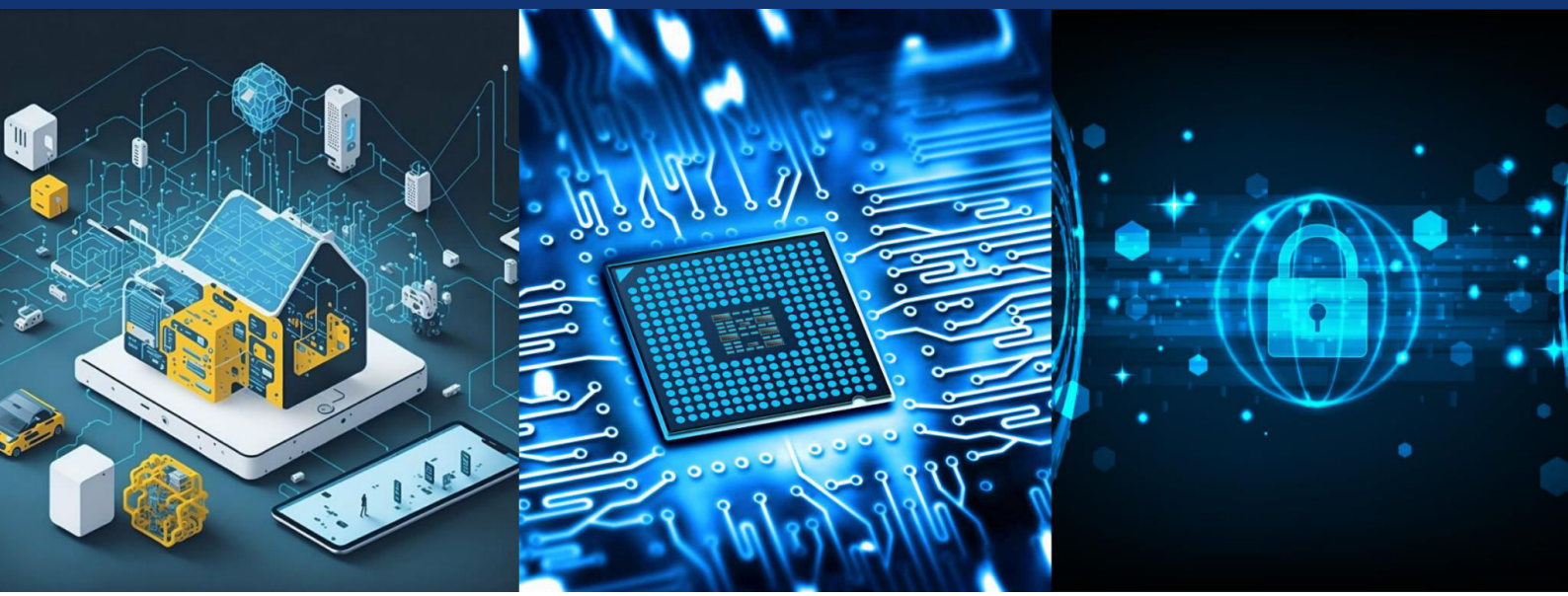
IJIRCCE

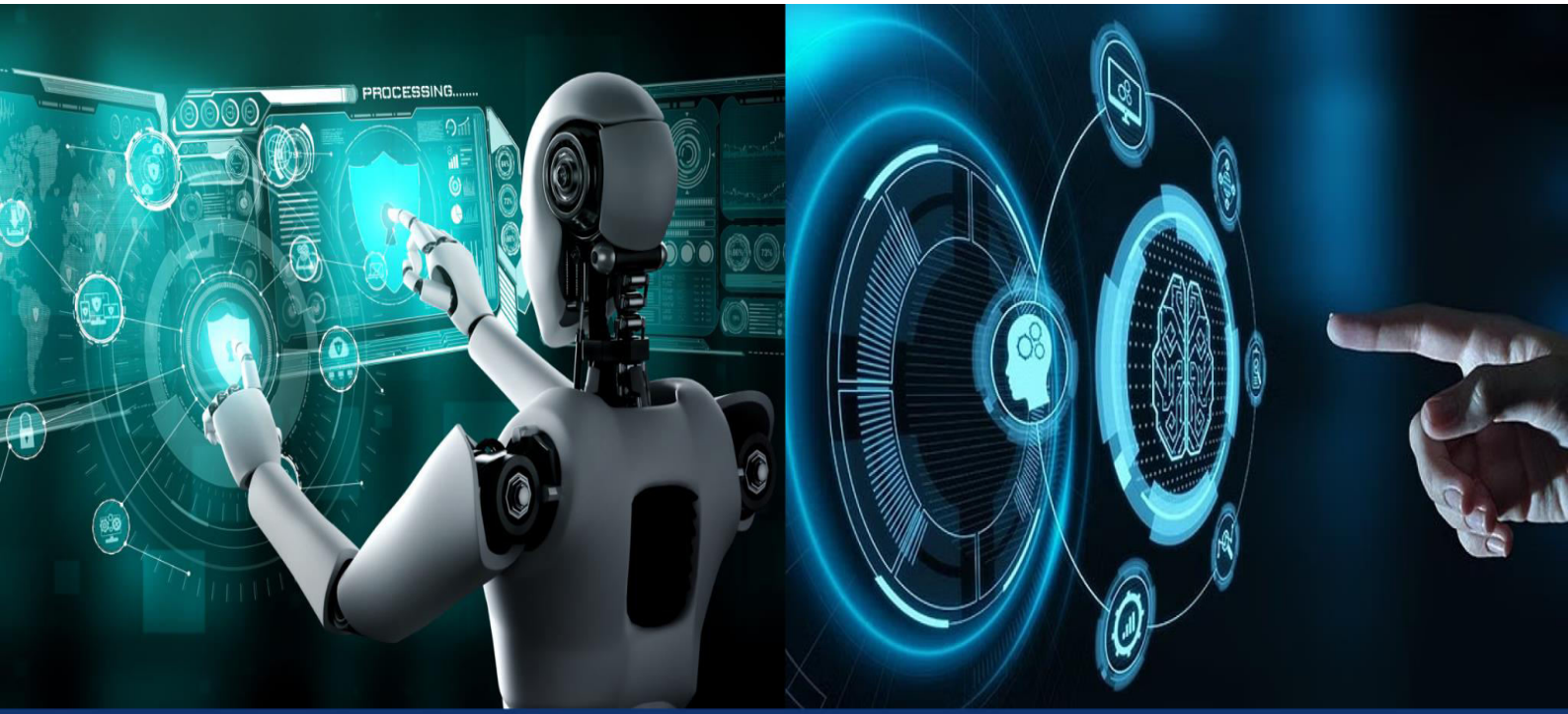
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304150

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning in Blockchain Technology for Effective Fraud Detection

Mohammad Ameeruddin[1], K. Roshini[2], K. Sai Lalithya[3], B. Poojitha [4], V. Jaya Sri[5]

U G Students, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[1,2,3,4]

Assistant Professor, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[5]

ABSTRACT: This project addresses the growing need for secure, intelligent blockchain systems by integrating Machine Learning (ML) for fraud detection. While blockchain ensures data immutability, it lacks built-in intelligence to identify fraudulent transactions. Existing research treats fraud detection and blockchain separately, creating a gap this work aims to fill. We embed an ML model into the blockchain's transaction layer to classify transactions before validation. A front-end interface facilitates user input and real-time prediction. Results show improved fraud detection accuracy and faster transaction validation. This approach enhances blockchain security and scalability, offering a novel direction for intelligent decentralized systems

KEYWORDS: Blockchain, Decentralized, Ethereum, Machine Learning, Artificial Intelligence, Fraud Detection, Anomaly Detection, Random Forest, Support Vector Machine.

I. INTRODUCTION

The development of decentralized systems like blockchain technology has transformed areas like finance, supply chains, healthcare, and even voting systems when paired with the internet. Blockchain technology's chief characteristics—transparency, immutability, and decentralization—provide immense security benefits, although they are not completely safe from fraudulent activities. As the number and complexity of transactions on blockchain networks increases, conventional transaction fraud detection techniques are becoming obsolete, making the case for advanced, flexible methods.

Artificial Intelligence (AI) is currently one of the most powerful domains that can be leveraged for analyzing and detecting anomalies through the use of Machine Learning (ML). With the ease of automation on the blockchain, ML can be trained on behavioural data trends, enabling identification of unusual activities, potential threats, and adaptation to evolving fraud strategies in real time.

This paper analyzes the integration of blockchain technology with machine learning in the context of improving fraud detection systems. It analyzes the implementation of machine learning algorithms on the blockchain framework for monitoring transaction data, considering anomaly detection and malicious activity prevention in a decentralized and secure fashion. The dissertation outlines the problems and possibilities in implementing ML-powered fraud detection systems on blockchain technology with the aim of enhancing the security and trust within decentralized systems.

The paper is structured as follows: Section II discusses related work, highlighting previous studies in ML-based fraud transaction. Section III provides a detailed background on the algorithms used in the project. Section IV introduces the proposed system, detailing the methodology and model architecture. Section V presents comparative results using graphical visualizations, and Section VI concludes the study with insights and future research directions.

II. RELATED WORKS

Several studies have explored fraud detection using Machine Learning (ML) and the inherent security of blockchain technology, but few have effectively integrated the two. Traditional fraud detection systems often rely on centralized databases, making them vulnerable to data tampering and single points of failure. On the other hand, blockchain offers data immutability and decentralized consensus, but lacks native intelligence for anomaly detection. Our work bridges this gap by embedding an ML classifier directly into the blockchain transaction layer, allowing proactive fraud

detection before transaction validation. This real-time, on-chain intelligence is a novel contribution toward more secure and scalable decentralized systems.

S.no	Author Details	Description	Limitations/Inference
1.	Siddamsetti, S., & Srivenkatesh, M. (2023).	This study addresses the detection of fraudulent transactions in the Ethereum blockchain using machine learning techniques. Utilizing a dataset of 9,841 Ethereum transactions, the authors employed algorithms such as decision trees, logistic regression, gradient boosting, XGBoost, and a hybrid model combining random forests with deep neural networks.	Further research with larger datasets is necessary to validate the model's effectiveness across diverse scenarios.
2.	Gupta, G., Mandal, B. K., Dwivedi, V., Sharma, V., S., R., Patil, A. K., & Kundu, D. (2024).	This research introduces an approach to detect fraud in health insurance claims by integrating blockchain technology with machine learning algorithms. claims data.	The paper focuses on the conceptual integration of blockchain and machine learning without providing detailed implementation results. Practical deployment challenges and real-world performance metrics are not extensively discussed..
3.	Gaikwad (Mohite), V., Meher, K., Dass, R., Sarah Jonista, A., D'Souza, J., & Victor, R. (2023).	This study develops a fraud detection system by combining machine learning algorithms with blockchain technology. Logistic Regression, Decision Tree, and Random Forest algorithms were implemented to improve the accuracy of detecting fraudulent activities.	The study provides a high-level overview of the integration without delving into specific performance metrics or comparative analysis with existing systems. Detailed evaluation of the system's effectiveness in real-world scenarios is lacking.
4.	Khan, M. A., Salah, K., & Crespi, N. (2023).	This survey explores the intersection of blockchain and machine learning, discussing how blockchain's decentralized nature can enhance the security and privacy of machine learning models.	While the survey provides a comprehensive overview, it primarily focuses on theoretical aspects and lacks detailed empirical studies or implementation examples.
5.	Wang, C., Dou, Y., Chen, M., Chen, J., Liu, Z., & Yu, P. S. (2021).	This paper addresses fraud detection in scenarios where node features are limited or unavailable. The authors propose a graph transformation method to capture structural information and introduce a graph pre-training strategy using contrastive learning.	The approach relies heavily on structural information, which may not be sufficient in cases where fraudulent patterns are not evident from graph structures alone. Integrating additional data sources could enhance detection capabilities.
6.	Sheng, Z., Song, L., & Wang, Y. (2025).	This study proposes a dynamic feature fusion model that combines graph-based representation learning and semantic feature extraction for blockchain fraud detection.	While the model shows improved performance, the complexity of integrating multiple data sources and the computational overhead may pose challenges for real-time fraud detection applications.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304150

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. BACKGROUND

1. Machine Learning Models

1.1 Random Forest

The image illustrates the Random Forest algorithm, an ensemble learning technique that combines multiple decision trees to improve prediction accuracy. Input data is passed through several trees, each trained on different subsets of the data. These trees make individual predictions, which are then aggregated—either by majority vote (for classification) or averaging (for regression)—to produce the final output. The diagram shows three decision trees leading to a combined prediction, symbolized by overlapping colored ovals. This approach reduces overfitting and enhances model generalization, making Random Forest highly effective for complex tasks involving classification, regression, and feature selection.

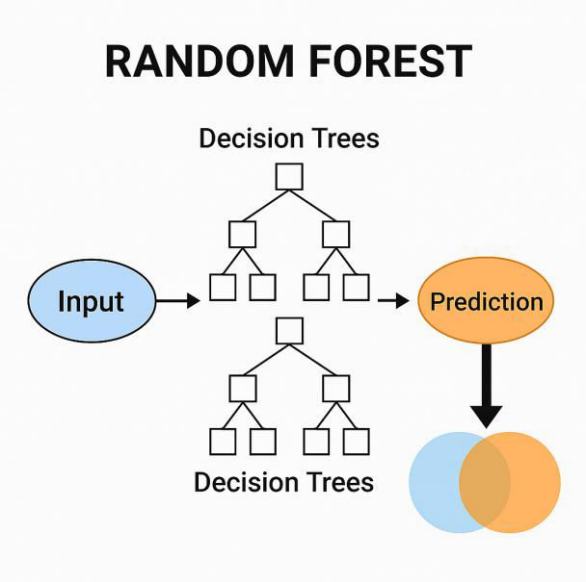


Figure 1: Random Forest

1.2 Support Vector Machine

Figure 2 demonstrates Support Vector Machine (SVM) is a powerful supervised machine learning algorithm widely used for classification and regression tasks. It is particularly effective in high-dimensional spaces and cases where the number of dimensions exceeds the number of samples. The fundamental idea of SVM is to find the optimal hyperplane that best separates data points of different classes by maximizing the margin between them. In cases where linear separation is not possible, SVM uses kernel functions (such as polynomial, radial basis function, or sigmoid) to transform the input space into a higher-dimensional space, enabling linear separation in the transformed space. SVM is known for its robustness against overfitting, especially in scenarios with limited data and high feature dimensionality. It performs well in applications such as text classification, bioinformatics, and image recognition.

The image illustrates the concept of a Support Vector Machine (SVM) for binary classification. It shows two classes of data points: blue circles and red crosses, plotted along Feature 1 and Feature 2 axes. A solid black line represents the hyperplane, which is the optimal decision boundary separating the two classes. Two dashed lines mark the margin, the maximum distance between the hyperplane and the closest data points from each class. These closest points are called support vectors and are crucial in defining the position and orientation of the hyperplane. SVM aims to maximize this margin for better generalization.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304150

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

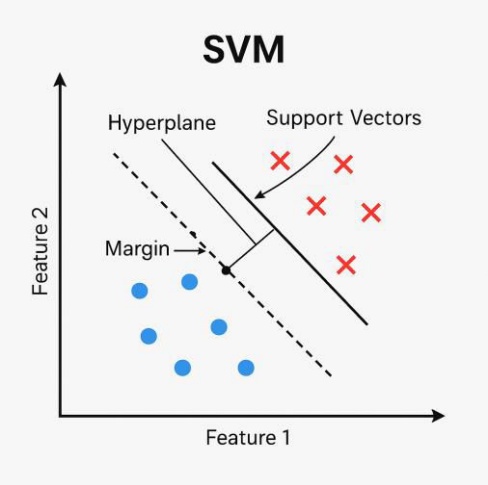


Figure 2: Support Vector Machine

1.3 Logistic Regression

Fig 3 As a baseline model, Logistic Regression was employed due to its interpretability and simplicity. It provides a probabilistic view of classification and serves to highlight the limitations of linear models on imbalanced, non-linear datasets like Ethereum transaction logs.

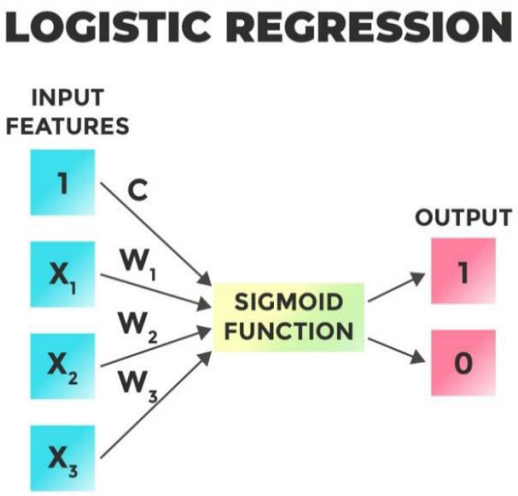


Figure 3: Logistic Regression

A Random Forest Classifier was selected as the primary model due to its superior performance in handling large datasets with complex feature interactions. It reduces variance through bagging and handles imbalanced class distributions more effectively than linear classifiers.

- Training/Test Split: The dataset was split into 80% training and 20% test sets.
- Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC Curve.
- Hyperparameter Tuning: Grid Search was used to optimize parameters such as the number of trees, maximum depth, and feature subsets.

The Random Forest model achieved the best balance between accuracy and false positive rate, and was therefore chosen for deployment with the smart contract layer

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304150

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. Dataset

Feature	Description
Index	The index number of a row.
Address	The address of the Ethereum account.
FLAG	Whether the transaction is fraud or not.
Avg min between sent tnx	Average time between sent transactions for the account in minutes.
Avg_min_between_received_tnx	Average time between received transactions for the account in minutes.
Time_Diff_between_first_and_last(Mins)	Time difference between the first and last transaction.
Sent_tnx	Total number of sent normal transactions.
Received_tnx	Total number of received normal transactions.
Number_of_Created_Contracts	Total number of created contract transactions.
Unique_Received_From_Addresses	Total unique addresses from which the account received transactions.
Unique_Sent_To_Addresses	Total unique addresses to which the account sent transactions.
Min_Value_Received	Minimum value in Ether ever received.
Max_Value_Received	Maximum value in Ether ever received.
Avg_Value_Received	Average value in Ether ever received.
Min_Val_Sent	Minimum value of Ether ever sent.
Max_Val_Sent	Maximum value of Ether ever sent.
Avg_Val_Sent	Average value of Ether ever sent.
Min_Value_Sent_To_Contract	Minimum value of Ether sent to a contract.
Max_Value_Sent_To_Contract	Maximum value of Ether sent to a contract.

Table 1: Dataset used in our model

Table 1: In this study, we utilized a comprehensive Ethereum blockchain dataset designed to facilitate effective fraud detection using machine learning techniques. The dataset includes transactional data from Ethereum addresses, featuring over 40 attributes that encompass behavioral patterns, financial metrics, and token interactions. Key features include the number of sent and received transactions, average time intervals between transactions, values of Ether and ERC20 tokens sent or received, and the number of unique interacting addresses. Each address is labeled with a binary classification flag indicating whether it is associated with fraudulent activity, thus enabling the use of supervised learning models. This rich feature set not only captures standard account activity but also delves into smart contract interactions and token-level behaviors, providing a multifaceted view of each address's operations. Such depth makes the dataset highly valuable for training machine learning models capable of distinguishing subtle fraudulent patterns from legitimate behavior. However, the dataset may suffer from class imbalance and potential labeling inaccuracies, which are common challenges in fraud detection applications.

IV. PROPOSED SYSTEM

1. Architecture

The diagram illustrates a four-layered architecture for integrating machine learning with blockchain technology to enhance functionalities such as fraud detection, transaction validation, and user interaction. Here's a detailed overview of each layer from top to bottom:

The system is structured into four primary layers:

1. Blockchain Layer: Hosts smart contracts and transaction records on Ethereum.
2. ML Layer: Contains the fraud detection model trained in Python using scikit-learn.
3. Backend Layer: Developed using Flask, serves the model via a REST API.
4. Frontend UI Layer: Built with HTML and CSS, allowing users to initiate transactions, view blockchain status, and receive fraud alerts.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304150

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

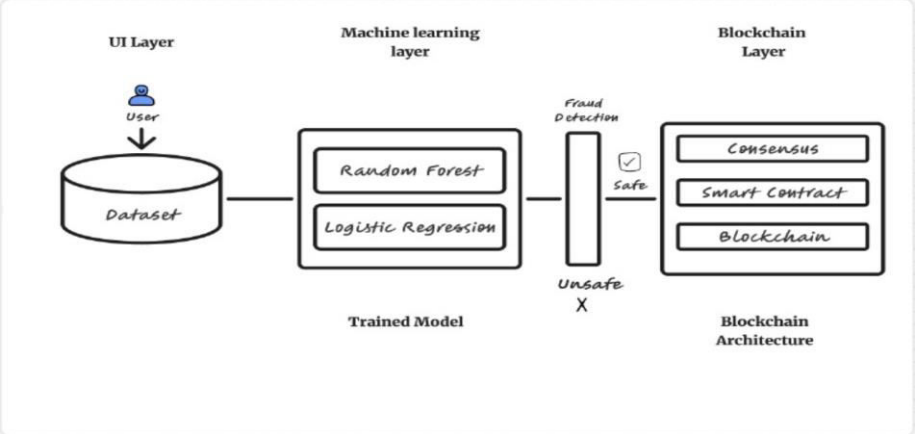


Figure 4: Architecture

2. Workflow

Fig.5 The flowchart illustrates a machine learning–integrated blockchain workflow designed for secure and fraud-resilient transaction processing. The process initiates when a user begins a transaction, which then moves into the Preprocessing stage. Here, the transaction data is cleaned, formatted, and prepared for analysis. The preprocessed data is passed through a Fraud Detection module powered by machine learning algorithms, which evaluate patterns and behavioural metrics to detect potentially malicious activity. If the transaction is flagged as invalid, it is rejected and returned for reassessment or flagged for investigation. If deemed valid, it proceeds to the Consensus stage, where network participants verify the authenticity and legitimacy of the transaction using a consensus mechanism such as Proof of Work or Proof of Stake. Upon successful consensus, the transaction activates a Smart Contract, which automatically executes predefined conditions associated with the transaction. Once the smart contract is executed, the changes are permanently written to the Blockchain Ledger, ensuring immutability and transparency. The process concludes with User Confirmation, notifying the user of the successful completion of the transaction. This workflow ensures a secure, automated, and intelligent transaction lifecycle, integrating fraud detection at an early stage to prevent malicious activities while maintaining the decentralized nature of blockchain systems.

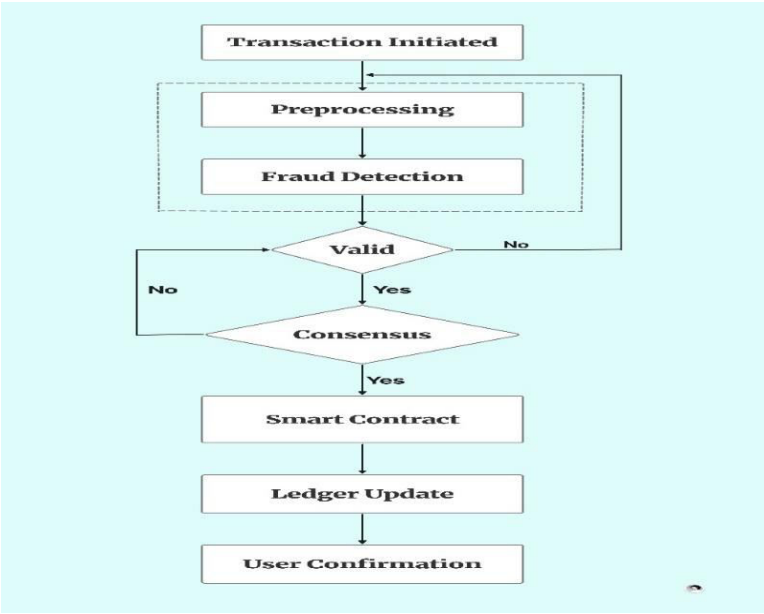


Figure 5: Workflow

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304150

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. RESULTS AND DISCUSSIONS

1. Comparative Analytic table

S.no	Model Name	Precision	Recall
1	Random Forest	0.93	0.92
2	Logistic Regression	0.85	0.80
3	Support Vector Machine	0.89	0.86

Table 2: Different model evaluations

Table 2 highlights the performance comparison of various machine learning models used for fraud detection in blockchain transactions. Among the evaluated models, Random Forest achieves the highest precision 0.93% and Recal 0.92% showcasing its robustness in handling complex patterns. Logistic Regression0.85%, with an Recall of 0.89%, offers decent performance but lacks in precision. Support Vector Machine (SVM) shows balanced metrics with a precision of 0.89% and an Recall of 0.86%, indicating its effectiveness in certain cases. Overall, the results suggest that ensemble models like Random Forest are more effective in identifying fraudulent activities compared to traditional classifiers.

2. Results

Fig6 The bar graph compares the performance of Random Forest, SVM, and Logistic Regression using Accuracy, Precision, and Recall. Random Forest generally outperforms the others, showing the highest scores in all three metrics (around 93% accuracy, 91% precision, and 92% recall). SVM performs reasonably well (around 89% accuracy, 87% precision, and 86% recall), while Logistic Regression has the lowest scores (around 85% accuracy, 82% precision, and 80% recall). This bar graph clearly indicates Random Forest as the strongest model among the three specific tasks

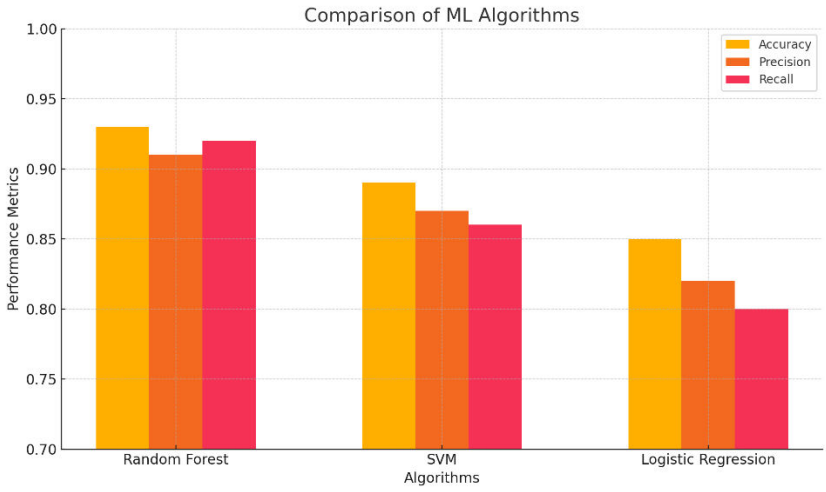


Figure 6: Performance of the algorithms

VI. CONCLUSION

The integration of Machine Learning and Blockchain represents a groundbreaking advancement in enhancing the security, transparency, and reliability of digital transactions. This study demonstrates that incorporating AI-driven models within the blockchain framework significantly improves fraud detection and threat mitigation, outperforming traditional rule-based mechanisms. The proposed system leverages intelligent algorithms to analyze transactional

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

behaviours in real-time, enabling the early identification of anomalies and malicious patterns. Despite these promising results, several challenges remain—particularly regarding computational overhead, privacy preservation, and the lack of standardized AI frameworks tailored for blockchain environments. Addressing these limitations will be essential for large-scale adoption. Looking ahead, the convergence of emerging technologies such as federated learning, quantum-resilient AI, and decentralized model training is poised to redefine blockchain security paradigms. These innovations will not only optimize system performance but also strengthen trust and resilience in decentralized networks. Therefore, the synergy between machine learning and blockchain is not merely a technological upgrade but a strategic step toward building smarter, safer, and more autonomous digital ecosystems.

REFERENCES

[1] “Fraud detection using machine learning and blockchain.” Agarwal, A., & Bansal, A. (2021). *Procedia Computer Science, 187*, 115–122.
[2] “A survey on the security of blockchain systems”. Li, X., Jiang, P., Chen, T., Luo, X., & Wen, Q. (2020). *Future Generation Computer Systems, 107*, 841–853.
[3] “Machine learning in cybersecurity: Applications and challenges”. Kaur, R., & Kumar, R. (2021). *IEEE Access, 9*, 80710–80735.
[4] Integrating AI and Blockchain Technology for Robust Fraud Detection Mechanisms," Prof. (Dr.) Khatib Noaman Umer, Rebecca John Kesavapattapa, *International Journal of Research in Commerce and Management Studies*, Vol. 6, No. 3, pp. 45–53, 2024.
[5] "Integrating Blockchain with Machine Learning for Fraud Detection in Health Insurance Claims Management," Ganesh Gupta, Bijay Kumar Mandal, Vinay Dwivedi, Vibhu Sharma, Ravindra S., Amit Kumar Patil, Debashis Kundu, *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 23s, pp. 128–139, 2024
[6] "Advanced Methodologies for Enhancing Credit Card Fraud Detection Utilizing Machine Learning, Blockchain Technologies, and Cryptographic Principles," Jitender Tanwar, Dipak Vijaykumar Bhosale, Vijay More, Vijit Srivastava, Tareek Pattewar, Kumar P., Pallavi Deshpande, *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 23s, pp. 140–150, 2024.
[7] "Credit Card Fraud Detection Utilizing Advanced ML and Blockchain Technologies," Adil Fahad, *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 23s, pp. 118–127, 2024.
[8] "Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain," Ranjitha H., Prof. M.N. Chandan, *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 11, No. 7, pp. 759–765, 2022
[9] "Efficient Fraud Detection in Ethereum Blockchain through Machine Learning and Deep Learning Approaches," Swapna Siddamsetti, Muktevi Srivenkatesh, *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 11, No. 11s, pp. 71–82, 2023.
[10] "Machine Learning Model for Detecting Fraudulent Transactions on the Ethereum Blockchain," Ayman Mohamed Mostafa, Ehab R. Mohamed, Reham Medhat, Asmaa Hanafy, *International Journal of Computers and Informatics (Zagazig University)*, Vol. 4, pp. 20–30, 2024.
[11] "Fraud Detection System for Identity Crime using Blockchain Technology and Data Mining Algorithms," Amol Jagdish Shakadwipi, Dinesh Chandra Jain, S. Nagini, *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 23s, pp. 151–160, 2024.
[12] "Insurance Fraud Detection Using Novel Machine Learning Technique," M. Sathya, B. Balakumar, *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 23s, pp. 161–170, 2024.
[13] "Fraud Detection Using Machine Learning and Blockchain," V. Gaikwad (Mohite), K. Meher, R. Dass, A. Sarah Jonista, J. D’Souza, R. Victor, *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 11, No. 6s, pp. 584–590, 2023.
[14] "Integration of Blockchain and Machine Learning for Secure Financial Transactions," by S. Kumar and L. Gupta, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 9, pp. 225-230, 2022.
[15] "Real-Time Fraud Detection Using Machine Learning in Blockchain Networks," by A. Verma and N. Singh, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 10, pp. 310-316, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details