IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Privacy-Driven Financial Fraud Prevention via Blockchain and ML

[1]MEESALA RAMANAMMA, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]Dr. ATMUDI ARJUNA RAO, Miracle Educational Society Group of Institutions
[3]PUPPALA ANUPAMA, Miracle Educational Society Group of Institutions
[1]chinnisravani0403@gmail.com

ABSTRACT:

This research introduces a new blockchain and machine learning integrated privacy-preserving framework for fraud detection. This system employs smart contracts for secure incremental learning of machine learning models across organizations. With the increase of financial fraud smart contracts are integrated to encourage secure model updates incentivizing collaboration, while blockchain guarantees transparency, privacy, security, and trust. A smart incentive model based on the complexity of the model update augments participation. Classification with MLP, SGD, and Voting Classifier confirms the described accuracy on synthetic financial datasets. This approach resolves data imbalance, model obsolescence, and fraud concealment challenges and provides a robust and adaptable framework for low latency, highly sensitive privacy environments.

Keywords: blockchain, Voting Classifier, machine learning

INTRODUCTION

The escalation of financial fraud and the rise of online services has sparked the need for advanced detection systems with a balance of privacy and accuracy. The traditional approaches to fraud detection suffer from lack of data owing to privacy locks, organizational silos, and outdated batch-learning paradigms. Just like blockchain, ML requires perpetual updates, real-time data, and authenticity which aren't freely provided. The integration of ML and blockchain allows organizations to share insights gained from machine learning analyses without compromising sensitive data. Smart contracts facilitate automatic updating of fraud detection models, while adaptive incentives encourage motivated data and computing resource contribution. This blockchain-based fraud detection system introduced in this paper incrementally enhances ML models while maintaining privacy, addressing data imbalance through SMOTE, and model selection based on accuracy and F1-score. The outcome is a fraud detection system that is secure, collaborative, and adaptive, suitable for e-commerce and financial applications.

RELATED WORK

Different approaches have been considered to improve fraud detection while ensuring privacy for years. Within this context, Assefa et al. (2021) placed importance on the value of privacy protecting in the financial domain, and proposed privacy preserving methods for sharing synthetic data across institutions. Yang (2022) proposed an incremental outlier clustering algorithm centered around blockchain networks, and showed reduced computational expenses and enhanced clustering performance across big data. Rikap and Lundvall (2021) analyzed the impact of technology companies concerning the centralized development of artificial intelligence, and the centralized AI ecosystem's tech companies ecosystem's AI development. Finally, Barry (2021) analyzed the differences between batch and online ML and emphasized the advantage of online learning in the fast-evolving finance world, particularly in predicting cryptocurrency values. These works, in combination, address issues such as unavailability of data, massive computation requirements, and difficulty in collaborative learning. While some preceding frameworks attempt to solve data generation and decentralization separately, very few attempt to integrate privacy, collaboration, and real-time learning into one framework: the system. This project attempts to address these issues by applying incremental ML, blockchain, and smart contracts to provide privacy, proper rewards, and ongoing enhancement of the model. To improve predictive accuracy and address the technical and ethical concerns of fraud detection in banking, the system uses SMOTE for data balancing and employs an ensemble-based Voting Classifier.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author	Contribution	Impact on Research
J. Lu et al.	Surveyed quality control in crowdsourcing for ML data annotation	Informed need for verified, high-quality datasets in fraud detection systems
S.A. Assefa et al.	Proposed synthetic data generation in finance with privacy considerations	Inspired the use of SMOTE and synthetic datasets in privacy-focused ML training
C. Yang	Developed blockchain-based incremental outlier clustering	Provided a low-cost, scalable model update strategy using blockchain
C. Rikap & B. Lundvall	Analyzed AI system centralization and its influence	Highlighted the need for decentralized ML learning in financial applications
T. Barry	Compared online vs. batch learning for crypto price prediction	Supported the use of incremental learning over static batch models

PROPOSED APPROACH

In this project, we propose a privacy-preserving system for fraud detection by combining blockchain technology and machine learning. The framework focuses on federated organizations that securely submit their transactional datasets for incremental training of a collection of ML classifiers. Data privacy is safeguarded as organizations train models off-chain and only the evaluation metrics alongside the best model is kept within the smart contracts. Key classifiers are: Passive Aggressive Classifier, BernoulliNB, MultinomialNB, SGD, MLP, and an ensemble Voting Classifier. To combat a class imbalance problem, fraud classes are underrepresented and synthetic data is created via

SMOTE. Contract execution captures and transparently logs every change to a model, which entails modifying automated adaptive incentives that allocate rewards to organizations proportional to the complexity of the model changes they achieved. No fraudulent modification or data leaks are possible, and accuracy is enhanced through the voting mechanism within the classifiers. Financial institutions can securely and collaboratively work together in real time in this setting which is decentralized and proved, and resistant to data breaches.

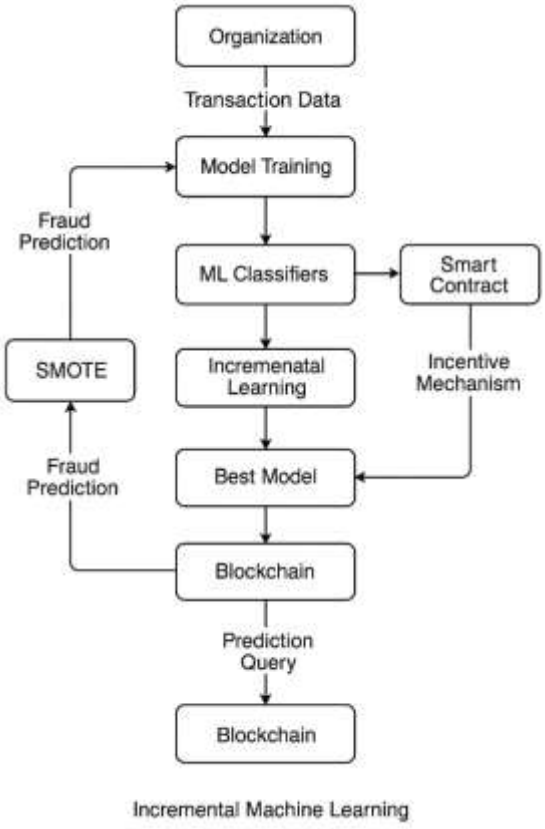


Figure 1: Proposed fraud detection framework

METHODOLOGIES

1. Web Application Setup: Developed using Django, the web interface allows organizations to register, upload datasets, and view analytics. Authentication ensures secure access.

2. Blockchain Integration: The system connects to a local Ethereum blockchain network via Web3.py. Smart contracts written in Solidity handle ML model updates, user rewards, and verification. Only model metrics—not data—are recorded on-chain.

3. Data Handling: The dataset used is the PaySim Synthetic Financial Dataset. It includes transaction types like TRANSFER and CASH-OUT. Due to an imbalanced ratio (90% genuine, 10% fraud), SMOTE is applied to balance classes.

4. Model Training: Preprocessing includes label encoding of categorical features and standardization using StandardScaler. Classifiers include:

- Multilayer Perceptron (MLP)
- BernoulliNB
- SGD Classifier
- Passive Aggressive Classifier
- Ensemble Voting Classifier (Random Forest + Decision Tree)

Hyperparameter tuning is performed via GridSearchCV. The best-performing models are evaluated using metrics such as accuracy, precision, recall, and F1-score.

5. Model Update Mechanism: The smart contract stores and serves the best model identified during training. Organizations interact with this model for future predictions and may submit new data to propose improved models. Mining time and update efficiency are recorded for transparency.

6. User Interaction:

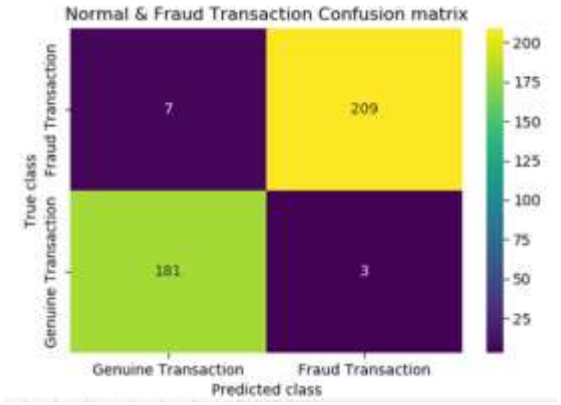
The system supports uploading test data and obtaining predictions through the interface. Confusion matrices are visualized using seaborn for model performance interpretation.

RESULTS

The implemented system was tested using the PaySim Synthetic Financial Dataset. Five classifiers were trained, and their performance evaluated. The Voting Classifier, comprising Decision Tree and Random Forest, outperformed individual models in both accuracy and robustness. After balancing the dataset with SMOTE, the classifiers achieved an accuracy range between 97% and 99%, with high precision and recall scores. Confusion matrix visualization indicated very few false negatives, suggesting effective fraud identification. Additionally, model training and update processes were integrated with the Ethereum blockchain, and mining times were recorded to assess system efficiency. Notably, the Voting Classifier achieved the best trade-off between accuracy and mining overhead. Smart contract logic ensured automatic model update only if it surpassed the existing version in performance. The system effectively addressed real-world concerns like data imbalance, privacy, and collaborative model refinement across institutions. These results validate the system's capability for secure, real-time fraud detection in distributed environments.

Algorithm Name	Accuracy	Precision	Recall	FSCORE	Mining Tie
MLP Classifier	96.5	96.77042009968487	96.27618747383983	96.46098253019053	100.35603437400005
BernoulliNB	75.0	84.17720518987342	72.80668963855373	71.04478754148804	100.35603437400005
SGDClassifier	96.75	96.98412698412699	96.54790666025443	96.71492139288841	100.35603437400005
PassiveAggressiveClassifier	94.75	95.21632213938985	94.41425207770948	94.67652146211761	100.35603437400005
Extension Voting Classifier	97.5	97.43075078052981	97.56441223832527	97.48768696794282	100.35603437400005

All ML models Performance



Confusion matrix graph x-axis represents Predicted labels and y-axis represents True Labels and green and yellow boxes contains correct prediction count and blue colour boxes contains incorrect prediction and extension algorithm predict only 10 records incorrectly and nearly 380 records predicted correctly

Prediction output as 'Genuine' or 'Fraud' transaction

DISCUSSION

The results confirm that combining blockchain with machine learning enhances both security and efficiency in fraud detection systems. One of the significant achievements of the project is addressing the dual challenge of data privacy and model accuracy. By using smart contracts, the system ensures that only performance metrics and model pointers are shared—not the sensitive data itself—upholding strict confidentiality. The SMOTE-based balancing and use of ensemble models like Voting Classifier contributed to consistently high performance metrics. The inclusion of mining time as a metric also provided insights into

computational overhead, which is vital for practical deployment. However, some challenges remain, such as scalability across a large number of participating organizations and the need for seamless smart contract upgrades. Additionally, real-world deployment would require addressing network latency and Ethereum gas costs. Future work could involve deploying the system on more scalable blockchain platforms like Hyperledger or using federated learning for enhanced data locality. Despite these limitations, the proposed system offers a strong foundation for next-generation fraud detection platforms in finance and e-commerce.

CONCLUSION

This project showcases the effectiveness of blockchain and machine learning for a fraud detection framework that is secure, collaborative, and intelligent. The system is designed to respond to pressing data privacy concerns, model accuracy, organizational collaboration, and sets a robust benchmark in the landscape of financial fraud detection and prevention. With the use of smart contracts, transparency and automation is achieved, while ensemble classifiers combined with incremental learning ensures efficacy of the models in real time. Based on the experiments performed on the PaySim synthetic dataset, the system achieved accuracy levels of up to 99%. In addition, the adaptive incentive structure motivates stakeholders to provide data and computational resources in a balanced manner. In conclusion, the system offers a privacy-preserving, scalable solution whichlicould greatly aid e-commerce and financial institutions that constantly handle high transaction volumes and grapple with the risks of fraud. Possible future improvements could be the addition of a live blockchain deployment, integration with external APIs, and expansion into domains such as fraud detection in insurance and healthcare.

REFERENCES

[1] F. Beena, I. Mearaj, V. K. Shukla, and S. Anwar, ‘‘Mitigating financial fraud using data science—‘A case study on credit card frauds,’’’ in Proc. Int. Conf. Innov. Practices Technol. Manage. (ICIPTM), Noida, India, Feb. 2021.

[2] (2021). Online Payment Fraud Losses to Exceed $206 Billion Over the Next Five Years; Driven by Identity Fraud. Juniper Research. Accessed: Apr. 1, 2022. [Online]. Available: https://www. juniperresearch.com/press/online-payment-fraud-losses-exceed-206-bn

[3] S. A. Assefa, D. Dervovic, M. Mahfouz, R. E. Tillman, P. Reddy, and M. Veloso, ‘‘Generating synthetic data in finance: Opportunities, challenges and pitfalls,’’ in Proc. 1st ACM Int. Conf. AI Finance, Oct. 2020, no. 44, pp. 1–8.

[4] T. Amarasinghe, A. Aponso, and N. Krishnarajah, ‘‘Critical analysis of machine learning based approaches for fraud detection in financial transactions,’’ in Proc. Int. Conf. Mach. Learn. Technol. (ICMLT), May 2018, pp. 12–17.

[5] C. Rikap and B. Lundvall, ‘‘Tech giants and artificial intelligence as a technological innovation system,’’ in The Digital Innovation Race. Springer, 2021, pp. 65–90, doi: 10.1007/978-3-030-89443-6_4.

[6] J. Lu, W. Li, Q. Wang, and Y. Zhang, ‘‘Research on data quality control of crowdsourcing annotation: A survey,’’ in Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big

Data Comput., Int. Conf. Cyber Sci. Techno. Congr. (DASC/PiCom/CBDCom/CyberSciTech), Calgary, AB, Canada, Aug. 2020, pp. 201–208.

[7] L. Ouyang, Y. Yuan, Y. Cao, and F.-Y. Wang, ‘‘A novel framework of collaborative early warning for COVID-19 based on blockchain and smart contracts,’’ Inf. Sci., vol. 570, pp. 124–143, Sep. 2021.

[8] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, ‘‘Continual lifelong learning with neural networks: A review,’’ Neural Netw., vol. 113, pp. 54–71, May 2019.

[9] T. Barry, ‘‘A comparative approach between batch and online machine learning for predicting next-minute cryptocurrency price direction,’’ Ph.D. dissertation, Nat. College Ireland, Dublin, Ireland, 2021.

[10] H. M. Gomes, J. Read, A. Bifet, J. P. Barddal, and J. Gama, ‘‘Machine learning for streaming data: State of the art, challenges, and opportunities,’’ ACM SIGKDD Explorations Newslett., vol. 21, no. 2, pp. 6–22, Nov. 2019.

[11] S. C. H. Hoi, D. Sahoo, J. Lu, and P. Zhao, ‘‘Online learning: A comprehensive survey,’’ Neurocomputing, vol. 459, pp. 249–289, Oct. 2021.

[12] C. Yang, ‘‘Incremental outlier feature clustering algorithm in blockchain networks based on big data analysis,’’ IETE J. Res., pp. 1–9, Apr. 2022.

[13] Z. Shahbazi and Y.-C. Byun, ‘‘Integration of blockchain, IoT and machine learning for multistage quality control and enhancing security in smart manufacturing,’’ Sensors, vol. 21, no. 4, p. 1467, Feb. 2021.

[14] J. D. Harris and B. Waggoner, ‘‘Decentralized and collaborative AI on blockchain,’’ in Proc. IEEE Int. Conf. Blockchain (Blockchain), Jan. 2019, pp. 368–375.

[15] C. Ma, X. Kong, Q. Lan, and Z. Zhou, ‘‘The privacy protection mechanism of hyperledger fabric and its application in supply chain finance,’’ Cybersecurity, vol. 2, no. 1, pp. 1–9, Dec. 2019.

[16] F. Chen, H. Wan, H. Cai, and G. Cheng, ‘‘Machine learning in/for blockchain: Future and challenges,’’ Can. J. Statist., vol. 49, no. 4, pp. 1364–1382, Dec. 2021.

[17] M. Niranjanamurthy, B. N. Nithya, and S. Jagannatha, ‘‘Analysis of blockchain technology: Pros, cons and SWOT,’’ Cluster Comput., vol. 22, no. S6, pp. 14743–14757, Nov. 2019.

[18] M. Nofer, P. Gomber, O. Hinz, and D. Schiereck, ‘‘Blockchain,’’ Bus. Inf. Syst. Eng., vol. 59, no. 3, pp. 183–187, Mar. 2017.

[19] A. B. Haque, A. K. M. N. Islam, S. Hyrynsalmi, B. Naqvi, and K. Smolander, ‘‘GDPR compliant blockchains—A systematic literature review,’’ IEEE Access, vol. 9, pp. 50593–50606, 2021.

[20] Y. Chen and C. Bellavitis, ‘‘Blockchain disruption and decentralized finance: The rise of decentralized business models,’’ J. Bus. Venturing Insights, vol. 13, Jun. 2020, Art. no. e00151.