

Implementation of Insurance Fraud in Healthcare Industry using Blockchain Technology with AI

Swapnali Tanaji Nagtilak¹, Ram Joshi², Nisha W Kimmatkar³ and Seema Kedar⁴

¹⁻⁴Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune, 411033, India
Email: swapnali6777@gmail.com, rbjoshi_it}@jspmrscoe.edu.in

Abstract— The escalating cost of healthcare has made cost control a critical concern for health insurance agencies. Traditional service-based reimbursement models often result in excessive and unjustifiable claims. Although the single-disease payment model has gained traction to mitigate this, it still presents a considerable risk of fraudulent activities. Insurance operations involve multiple stakeholders and are frequently hindered by outdated, opaque claim processing mechanisms. Blockchain technology, with its immutable ledger and decentralized validation, offers a promising solution to address these inefficiencies and fraud vulnerabilities. This paper proposes a consortium blockchain-based framework enhanced with machine learning to detect fraudulent insurance claims automatically. The system includes an explainable AI component that assesses the legitimacy of medical codes based on patient symptoms. Moreover, the use of blockchain ensures secure, transparent, and auditable storage of healthcare records, thereby assisting auditors and reinforcing trust in the claim settlement process.

Index Terms— Blockchain artificial intelligence healthcare management EHR.

I. INTRODUCTION

Health insurance (HI) serves as a financial agreement wherein insurers compensate subscribers for healthcare expenditures arising from medical conditions, injuries, or accidents. As healthcare costs surge globally, HI has become essential for individuals, not only for medical coverage but also for tax-saving incentives. Historically, claim processing in health insurance involved manual, paper-based procedures that required in-person visits to insurance offices, making the process time-consuming and prone to errors or manipulation.

In the realm of healthcare, Blockchain finds diverse applications and functionalities. Its ledger technology aids healthcare researchers in deciphering genetic codes by ensuring the secure exchange of patient medical records. Moreover, it facilitates the management of the drug supply chain and ensures the safe transfer of patient medical data. In research, AI has revolutionized data analysis by scrutinizing vast datasets and uncovering intricate patterns beyond human detection capabilities. This has paved the way for breakthroughs in fields like genomics and drug discovery. Healthcare settings leverage AI to devise diagnostic tools and tailor personalized treatment plans. This paper presents a consortium blockchain-based architecture enhanced with explainable AI for automatically detecting and preventing healthcare insurance fraud. The proposed system integrates a three-tier framework encompassing data prediction, user interaction, and blockchain validation. By combining technological strengths, the system not only enhances fraud detection but also improves efficiency, auditability, and trust in insurance operations.

II. RELATED WORK

This research focuses on building a blockchain-driven model to enhance security and minimize fraud in the insurance industry. Traditional claim processing mechanisms are identified as major risk factors due to their centralized and manual nature. Kapadiya et al. [4] proposed a blockchain-AI hybrid for secure claim handling, emphasizing smart contracts for automated validation. Triyono and Ginting [3] assessed machine learning models like Naïve Bayes and K-NN for fraud detection, establishing baseline performance metrics. Ali et al. [1] introduced blockchain with IPFS for secure EHR storage, demonstrating enhanced access control. Unlike prior works that treat storage and detection as separate entities, our approach integrates supervised learning (SVM) with SHA-256-backed blockchain validation, enabling transparent, automated, and tamper-resistant fraud mitigation within a unified architecture.

III. PROPOSED METHODOLOGY

A. System Architecture

The proposed system is structured into three key layers: the Prediction Layer, User Layer, and Blockchain Layer. In the Prediction Layer, insurance claim data is collected and preprocessed before being split into training and testing sets. Machine learning algorithms are then applied to classify claims as either genuine or fraudulent. The User Layer involves interactions between the insurance subscriber (patient), insurance provider, and service providers such as hospitals or labs. Here, insurance claims are generated and submitted based on services received. Once a claim is verified as genuine, it proceeds to the Blockchain Layer. In this layer, the claim is validated and added to a block, which is then securely added to the blockchain. This process ensures transparency, data integrity, and protection against fraudulent activities.

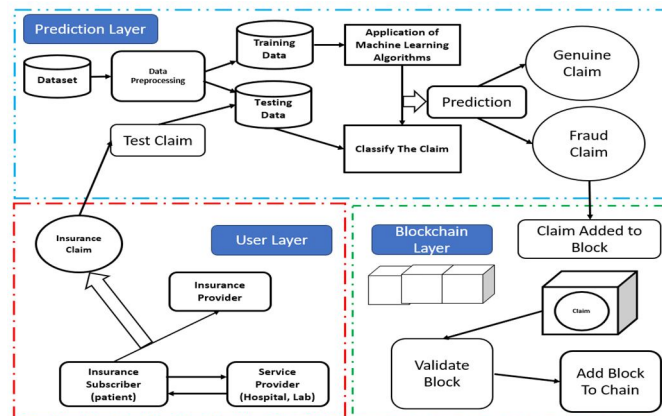


Figure 1: System Architecture

B. Algorithm

Support Vector Machine (SVM)

SVM (Support vector machine) is a popular algorithm used for many classification problems.

- It is one of the supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.
- An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

Input: D dataset, on-demand features, aggregation-based features

Output: Classification of Application

SHA256

SHA-256 (256-bit) is part of SHA-2 set of cryptographic hash functions, designed by the U.S. National Security Agency (NSA) and published in 2001 by the NIST as a U.S. Federal Information Processing Standard (FIPS). A hash function is an algorithm that transforms (hashes) an arbitrary set of data elements, such as a text file, into a single fixed length value (the hash). The computed hash value may then be used to verify the integrity of

copies of the original data without providing any means to derive said original data. Irreversible, a hash value may be freely distributed, stored and used for comparative purposes. SHA stands for Secure Hash Algorithm. SHA-2 includes a significant number of changes from its predecessor.

IV. RESULTS AND DISCUSSION

Insurance Fraud Prediction

User Provided Data:

Gender	0
Location	0
Employer	0
Relationship	0
Cause	0
Fee Charged	3798
Membership Period (years)	4
Number of Claims	2
Number of Dependents	4
Patient Age	27

Provided details are: LEGITIMATE

Figure 2: Based on the provided details of insured person, fraud is predicted whether the claim made by the insured person is legitimate or fraudulent

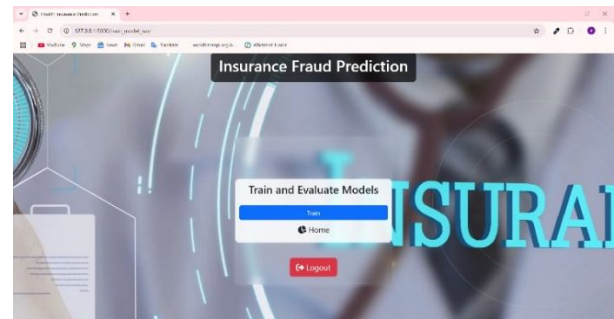


Figure 3: After prediction, the model is trained

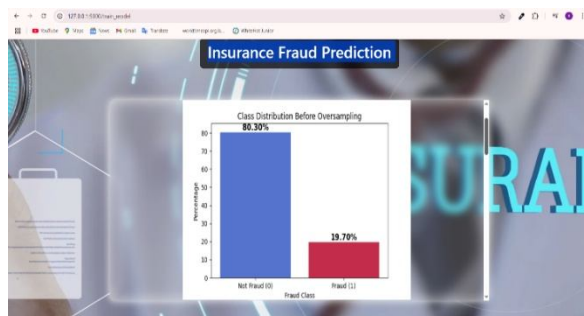


Figure 4: Class Distribution before oversampling, where not fraud cases are more compared to fraud cases

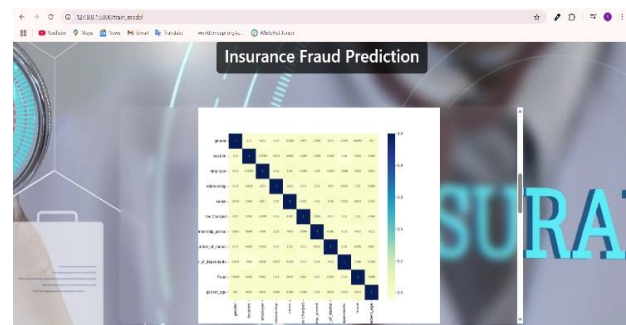


Figure 5: The image shows a correlation heatmap between various features like gender, location, employer, relationship, cause, Fee Charged, and the target variable fraud

- Diagonal values are 1 because each feature is perfectly correlated with itself.
- Off-diagonal values are close to zero, which means the features are mostly uncorrelated with each other.

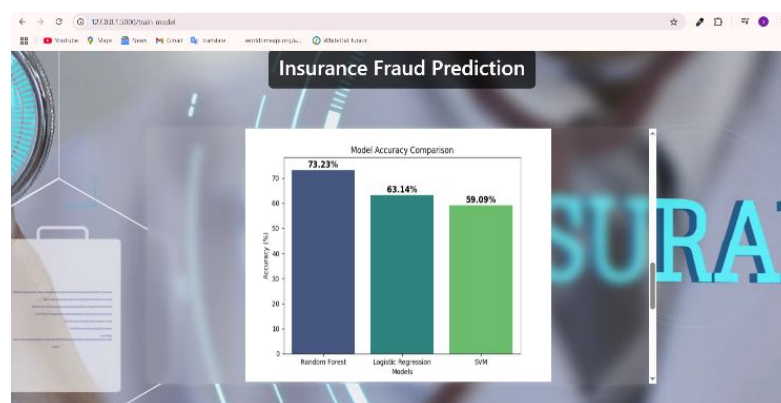


Figure 6: This chart showcases the Model Accuracy Comparison for insurance fraud prediction models.

Random Forest likely performs better because it handles both numerical and categorical data well, and it's less prone to overfitting due to ensemble learning. Logistic Regression is interpretable and works best when there's a linear decision boundary. SVM might not be ideal here if your data isn't well-separated or isn't scaled properly.

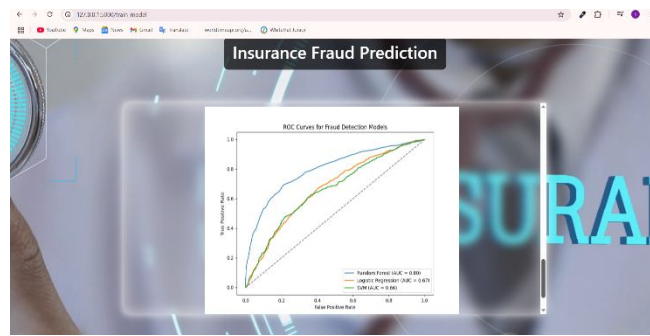


Figure 7: Random Forest model is outperforming the others for fraud detection.

ROC Curve Analysis

- Random Forest:
 - AUC = 0.80 — strong performance, solid at distinguishing fraud from non-fraud.
- Logistic Regression
 - AUC = 0.67 — moderate discrimination capability.
- SVM:
 - AUC = 0.66 — slightly below logistic regression.

V. CONCLUSIONS

The findings and results will demonstrate that the utilization of blockchain for storing health information can achieve effective security by distributing data across multiple machines, overseen and authorized by a decentralized community, as opposed to a centralized approach. This approach allows all parties involved to access and verify data additions and modifications. In this project we predict fraud by entering the details of the insured person. The insured person details are entered, and the fraud is predicted. Based on the provided details, fraud is predicted whether the claim made by the insured person is legitimate or fraudulent.

REFERENCES

- [1] M. Ali, S. Ahmed, M. I. Hossain, A. B. M. Alim Al Islam, and J. Noor, "Electronic health record's security and access control using blockchain and IPFS," in **Proc. 7th Int. Congr. Inf. Commun. Technol.**, Singapore: Springer, 2023, pp. 493–505.
- [2] S. Gupta, H. K. Sharma, and M. Kapoor, "Introduction to blockchain and its application in smart healthcare system," in **Blockchain for Secure Healthcare Using Internet of Medical Things (IoMT)**, Cham: Springer, 2023, pp. 55–65.
- [3] G. Triyono and D. Ginting, "Comparative analysis performance of Naïve Bayes and K-NN using confusion matrix and AUC to predict insurance fraud," **J. Media Informatika Budidarma**, vol. 6, no. 4, pp. 2293–2300, 2022.
- [4] K. Kapadiya et al., "Blockchain and AI-empowered healthcare insurance fraud detection: An analysis, architecture, and future prospects," **IEEE Access**, vol. 10, pp. 79606–79627, 2022.
- [5] S. Vyas and S. Serasiya, "Fraud detection in insurance claim system: A review," in **Proc. 2nd Int. Conf. Artif. Intell. Smart Energy (ICAIS)**, 2022, pp. 922–927, doi: 10.1109/ICAIS53314.2022.9742984.
- [6] S. E. Hiererra et al., "Blockchain technology for fraud detection and risk prevention in insurance industry," **Unpublished manuscript**.
- [7] A. Gupta, J. Patel, M. Gupta, and H. Gupta, "Issues and effectiveness of blockchain technology on digital voting," **Int. J. Eng. Manuf. Sci.**, vol. 7, no. 1, 2017.
- [8] A. Navya, R. Roopini, and A. S. SaiNiranjana, "Electronic voting machine based on blockchain technology and Aadhar verification," **Int. J. Adv. Res. Ideas Innov. Technol.**, vol. 4, no. 2, 2018.
- [9] F. S. Hardwick, R. N. Akram, and K. Markantonakis, "E-voting with blockchain: An e-voting protocol with decentralisation and voter privacy," **arXiv preprint arXiv:1805.10258**, 2018.
- [10] C. Meter, "Design of distributed voting systems," **arXiv preprint arXiv:1702.02566**, 2017.
- [11] S. Panja and B. K. Roy, "A secure end-to-end verifiable e-voting system using zero knowledge-based blockchain," **Unpublished**.
- [12] M. A. Makary and M. Daniel, "Medical error—the third leading cause of death in the US," **BMJ: British Med. J. (Online)**, vol. 353, 2016.
- [13] P. T. S. Liu, "Medical record system using blockchain, big data and tokenization," in **Int. Conf. Inf. Commun. Security**, Springer, 2016, pp. 254–261.

- [14] C. Decker and R. Wattenhofer, "Information propagation in the Bitcoin network," in **Proc. 13th IEEE Int. Conf. Peer-to-Peer Comput.**, 2013, pp. 1–10.
- [15] T. Neudecker, P. Andelfinger, and H. Hartenstein, "Timing analysis for inferring the topology of the Bitcoin peer-to-peer network," in **IEEE Int. Conf. UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld**, 2016, pp. 358–367.
- [16] D. Zhang, "Resilience enhancement of container-based cloud load balancing service," **PeerJ Preprints**, 2018.
- [17] Y. Kou, C. T. Lu, and S. Sirwongwattana, "Survey of fraud detection techniques," in **Proc. IEEE Int. Conf. Networking, Sensing and Control**, 2004.
- [18] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," **J. Network Comput. Appl.**, Elsevier, 2016.
- [19] V. Almendra, "Finding the needle: A risk-based ranking of product listings at online auction sites for non- delivery fraud prediction," **Expert Syst. Appl.**, 2013. [Online]. Available: <https://colab.ws/articles/10.1016%2Fj.eswa.2013.02.027>
- [20] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in **Proc. Int. Conf. Comput., Netw. Informatics (ICCNI)**, IEEE, 2017.
- [21] J. Perols, "Financial statement fraud detection: An analysis of statistical and machine learning algorithms," **Auditing: J. Pract. Theory**, AAA, 2011.
- [22] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," **Procedia Comput. Sci.**, Elsevier, 2019.
- [23] P. Raghavan and N. El Gayar, "Fraud detection using machine learning and deep learning," in **Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)**, IEEE, 2019.
- [24] A. Thennakoon, C. Bhagyan, and I. Perera, "Real-time credit card fraud detection using machine learning," in **Int. Conf. Cloud Comput. Intell. Syst.**, IEEE, 2019.
- [25] R. Sailusha, V. Gnaneswar, and R. Ramesh, "Credit card fraud detection using machine learning," in **Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)**, IEEE, 2020.
- [26] Anonymous, "Sequence classification for credit-card fraud detection," **Expert Syst. Appl.**, 2018.
- [27] Kimmattkar, Nisha Vishnupant, and B. Vijaya Babu. "Novel approach for emotion detection and stabilizing mental state by using machine learning techniques." *Computers* 10.3 (2021): 37.
- [28] Kimmattkar, Nisha Vishnupant, and B. Vijaya Babu. "Human emotion detection with electroencephalography signals and accuracy analysis using feature fusion techniques and a multimodal approach for multiclass classification." *Engineering, Technology & Applied Science Research* 12.4 (2022): 9012-9017.
- [29] P. Sahane, A. Obla and P. Bhangale, "Student Achievements Portfolio Management using Blockchain," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 1787-1792, doi: 10.1109/ICICT57646.2023.10134076.
- [30] R. Deshmukh, T. Lahamge, I. Phadatare, D. Shinde and R. Manhas, "Real-Time Marathi Sign Language Recognition using Deep Learning Techniques," 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2024, pp. 1326-1331, doi: 10.1109/ICACRS62842.2024.10841608.
- [31] Kedar, Seema, Saurabh Sutar, and Himanshu Prasad. "Smart Analyzer: Assisting College Management through Machine Learning and Data Analysis." *Turkish Journal of Computer and Mathematics Education* 12.1S (2021): 137-145.]
- [32] Manjrekar, A.A., Bhosale, A.H., Khandagale, H.P., Nikam, S.S. (2025). Data Storage with Attribute-Based Encryption and De-duplication Filter on Hybrid Cloud. In: Tuba, M., Akashe, S., Joshi, A. (eds) *ICT Systems and Sustainability. ICT4SD 2024. Lecture Notes in Networks and Systems*, vol 1159. Springer, Singapore. https://doi.org/10.1007/978-981-97-8526-1_42