Healthcare Insurance Fraud Detection Using Blockchain and AI

Dr. Vinmathi M S[1]
Professor
Department of Computer Science and Engineering
Panimalar Engineering College
Chennai, India.
vinmathis@gmail.com

Dr. Kavitha Subramani[2]
Professor
Department of Computer Science and Engineering
Panimalar Engineering College
Chennai, India.
kavitha.pec2022@gmail.com

Afra Fathima M A[3]
Department of Computer Science and Engineering.
Panimalar Engineering College
Chennai, India.
afrafathima2508@gmail.com

Harsha K[4]
Department of Computer Science and Engineering.
Panimalar Engineering College
Chennai, India.
harshak.2907@gmail.com

Kavyaa A K[5]
Department of Computer Science and Engineering.
Panimalar Engineering College
Chennai, India.
akkavyaa04@gmail.com

Abstract— Any major problem in healthcare insurance fraud can result in significant challenge, leading to financial losses and can also erode systemic confidence. With the increasing reliance on health insurance for covering medical expenses such as hospitalization, treatment, and preventive care, ensuring security and fraud prevention has become crucial. Conventional rule-based fraud detection techniques find it difficult to keep up with changing fraudulent patterns. To address this, a blockchain and AI-powered system is introduced, leveraging blockchain for secure and tamper-proof data management and AI-driven analytics for intelligent fraud detection, enhancing transparency, accuracy, and security.

***Keywords— Health insurance, fraud detection, blockchain, artificial intelligence, privacy, economic growth*.**

I. INTRODUCTION

Health insurance fraud is a growing global concern, causing significant financial losses and undermining trust in healthcare systems. Recent data reveals the staggering scale of this issue. Globally, healthcare fraud accounts for 6-10% of total healthcare expenditures, resulting in losses worth hundreds of billions annually. In the U.S., fraud costs approximately $68 billion each year, while in India, 15-20% of health insurance claims are flagged as potentially fraudulent, leading to annual losses exceeding ₹10,000 crore.

Fraudulent activities are becoming increasingly sophisticated. Studies show that 40% of fraud cases involve collusion between providers, patients, and intermediaries, while 30% are linked to fabricated medical records or inflated billing. These practices drive up insurance premiums by 10-15%, burdening genuine policyholders. Traditional detection systems, reliant on centralized databases and rule-based methods, are inadequate. Surveys indicate that 65% of insurers still use outdated systems, with 70% failing to detect emerging fraud patterns like AI-generated fake documents or blockchain-manipulated claims.

We provide a creative way to deal with these issues by combining blockchain and AI technologies. Blockchain ensures data security, immutability, and transparency, while AI enhances fraud detection accuracy. Our framework aims to reduce fraudulent claims by up to 40%, lower operational costs by 25%, and improve detection efficiency by 30%. This research seeks to transform health insurance fraud detection, safeguarding insurers and policyholders while fostering a more transparent and efficient healthcare ecosystem.

II. RELATED WORKS

Blockchain and artificial intelligence combined for healthcare insurance fraud detection has been widely studied to enhance security, data integrity, and fraud prevention. Various methodologies have been explored, leveraging machine learning, blockchain-based smart contracts, and big data analytics.

A blockchain-powered fraud detection system employing Random Forest, SVM, and Decision Tree demonstrated improved fraud identification accuracy and resistance to data manipulation [1]. A hybrid approach combining supervised and unsupervised learning showcased superior fraud detection over rule-based methods, though dataset preprocessing remained a challenge due to class imbalance [2].

Blockchain's application in securing insurance transactions through smart contract-based fraud prevention frameworks has ensured transparency, traceability, and security, although legacy system integration remains a significant challenge [3]. Benchmarking of Naïve Bayes, XGBoost, and Deep Learning models revealed that ensemble methods achieve superior fraud detection accuracy, albeit with increased computational overhead [4]. Unsupervised learning techniques, including Autoencoders and Isolation Forests, have been effective for anomaly detection, but high false-positive rates pose an ongoing challenge [5].

The adoption of Hyperledger Fabric and Ethereum-based smart contracts has enhanced fraud detection auditability, but concerns regarding scalability and transaction costs persist [6]. AI-driven frameworks leveraging feature engineering and predictive analytics on historical claims data have improved fraud detection rates but require extensive labeled datasets for optimal model performance [7]. Deep learning models such as CNN and LSTM have achieved high fraud detection accuracy, yet their real-world applicability is constrained by computational complexity and data requirements [8].

Hybrid AI-blockchain approaches integrating smart contracts and anomaly detection models have enhanced fraud detection efficiency, though blockchain transaction latency remains a limitation [9]. To address class imbalance, resampling techniques such as SMOTE and ADASYN have been employed, improving classification accuracy but sometimes introducing synthetic noise [10]. Blockchain-based frameworks integrating smart contracts for fraud prevention have improved data transparency and security, although computational costs and scalability issues persist [11]. AI-based fraud detection employing K-means clustering has enhanced anomaly detection in medical claims, outperforming rule-based approaches while facing challenges with high false-positive rates due to dataset variability [12].

Implementations utilizing BigchainDB have improved insurance transaction security and claims processing efficiency, though interoperability limitations hinder broader adoption [13]. Studies benchmarking ensemble and deep learning models for fraud detection in medical claims indicate that ensemble methods improve accuracy but require substantial computational resources, highlighting a trade-off between precision and efficiency [14]. Porter's value chain and Berliner's insurability criteria have been used to assess how digitization affects insurance fraud, revealing increased operational efficiency but heightened cybersecurity risks, necessitating regulatory adaptations [15]. AI-based fraud detection techniques leveraging supervised learning methodologies have shown promise in recognizing fraudulent claims but require further advancements in false-positive reduction [16].

Privacy and security concerns in healthcare big data applications have been addressed using real-time monitoring and encryption techniques, with ongoing concerns about vulnerabilities in patient data protection [17]. A big data analytics framework for fraud detection utilizing Hadoop has been developed, improving Electronic Health Record (EHR) management, although scalability remains a concern [18]. Digitalization's effect on insurance risk assessment has been studied using Porter's value chain, highlighting efficiency gains while emphasizing regulatory challenges and data security risks [19]. A big data-driven e-health insurance model using Infinispan and MapReduce has improved data segregation and extraction, though challenges in data consistency and privacy persist [20].

Protection of Electronic Health Records (EHRs) during storage and transmission has been studied, proposing secure encrypted storage with controlled access, enhancing HIPAA compliance but requiring better interoperability mechanisms [21]. Blockchain technology's benefits and threats in healthcare fraud detection have

been categorized, highlighting enhanced security and data tracking, though energy consumption and interoperability remain adoption barriers [22].

The ML models like SVM and clustering have been used in healthcare to detect fraud using big data analytics, outperforming traditional rule-based methods but requiring better data integration strategies for handling heterogeneous datasets [23]. Medicare fraud detection studies emphasize the need for standardized preprocessing techniques to enhance machine learning effectiveness, addressing gaps in data fusion methodologies [24]. AI-driven security frameworks integrating blockchain have been proposed for healthcare data protection, with models such as SVM, KNN, and VFDT proving effective in anomaly detection, though high computational costs remain a constraint [25].

This literature review highlights the growing importance of blockchain-based and AI-powered healthcare to detect fraud, showcasing advancements in machine learning models, big data analytics, and blockchain-based security mechanisms. While significant progress has been made, challenges such as scalability, interoperability, computational costs, and false-positive rates need to be addressed to enhance the efficiency and reliability of these fraud detection frameworks.

III. PROPOSED MODEL

The proposed system consists of three main layers:

A. Data Storage and Management Layer

This layer secures patient records and insurance claims using InterPlanetary File System (IPFS) and blockchain. Each record is hashed before storage, ensuring immutability:

$$H(D_P)=SHA_{256}(Dp)$$

$$CID(D_P)= IPFS(H(Dp))$$

Where, D_p is patient data, and CID is a unique identifier for retrieval. Blockchain blocks store claim hashes, preventing unauthorized modifications.

B. Fraud Detection Layer

AI models analyze historical claims to identify fraudulent activities. The system uses Random Forest Classifier.

$$F(a) = \frac{1}{N}\sum_{i=1}^{N} f_i(a)$$

where fi (a)→ prediction from each tree.

C. Claim Verification Layer

Smart contracts automate claim validation:

$$V(C_i) = \begin{cases} 1, if\ P(F) < \theta\ and\ CID(C_i) exist \\ 0, otherwise \end{cases}$$

V(Ci)→ The claim verification function, which outputs 1 (valid) or 0 (fraudulent).

Ci → The given insurance claim under verification.
P(F) → The fraud probability score assigned to the claim by AI-based fraud detection models.
θ → The fraud detection threshold, a predefined limit beyond which a claim is flagged as fraudulent.
CID(Ci)→ The Content Identifier (CID) linked to the claim in the InterPlanetary File System (IPFS) and blockchain.
1 (Approved Claim) → The claim is valid and gets approved for processing.
0 (Rejected Claim) → The claim is flagged as fraudulent and denied.

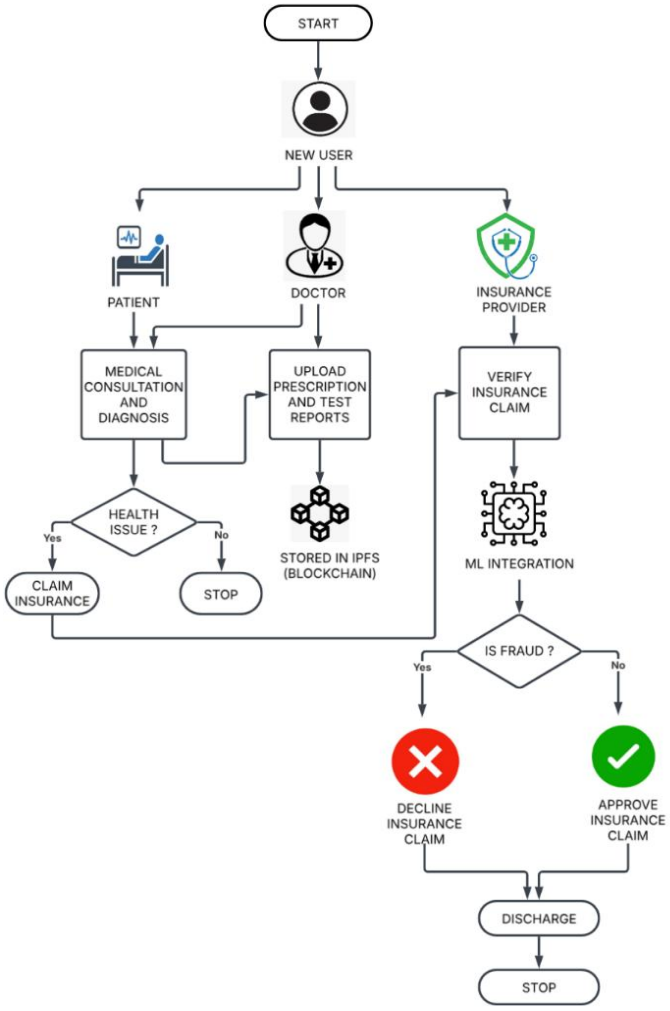

Fig.1. Blockchain and AI-Empowered Healthcare Fraud Detection System Architecture

Patients undergo diagnosis, and doctors upload reports to IPFS. The insurance provider verifies claims using machine learning for fraud detection. Approved claims proceed, while fraudulent ones are declined, ensuring a secure, transparent, and efficient healthcare insurance system.

IV. METHODOLOGY

Data collection, preprocessing, fraud detection, blockchain integration, and smart contract execution are all steps in a structured methodology that makes fraud detection in the healthcare insurance industry safe, scalable, and effective.

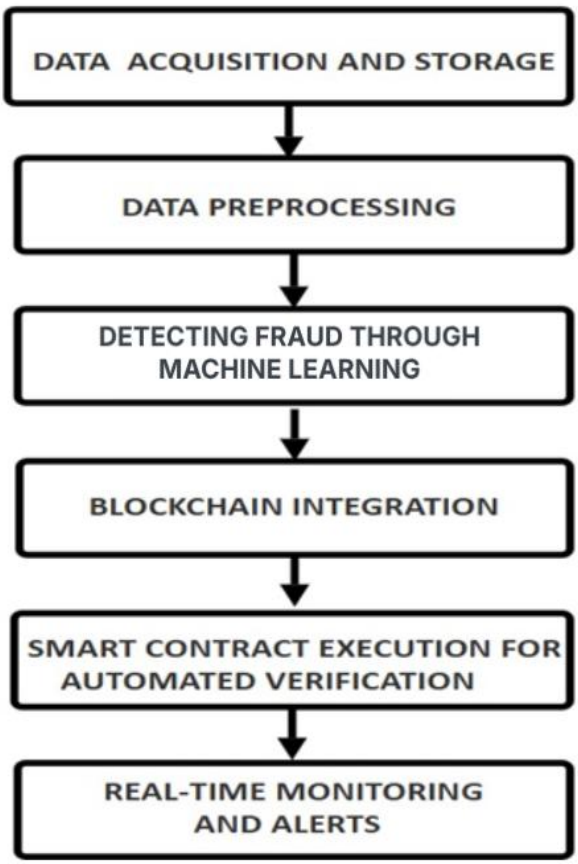

Fig. 2. Flow Diagram

A. Data Acquisition and Storage

Healthcare claim records, including patient details, diagnosis reports, treatments, and billing history, are collected from hospitals and insurance firms. Key attributes like age, blood pressure, cholesterol, heart rate etc. help assess heart disease risk. Data is securely stored in IPFS, assigned a unique CID, and linked to blockchain for integrity verification, ensuring fraud prevention and secure medical record management.

B. Data Preprocessing

Raw claim data undergoes preprocessing to enhance accuracy and consistency. Normalization standardizes numerical values, feature selection extracts key attributes like claim amount and patient history, and anomaly detection identifies suspicious patterns, improving the efficiency and reliability of AI-driven fraud detection models.

C. Detecting Fraud Through Machine Learning

Algorithms for machine-learning are trained on labeled historical claims to classify future claims as legitimate or fraudulent. The fraud detection model utilizes Random Forest, to analyze transaction patterns.

D. Blockchain Integration

Once classified, claim records are stored on a private blockchain network to ensure immutability and transparency. Each claim's transaction hash is linked to the previous block, forming a tamper-proof ledger. The system employs the SHA-256 hashing algorithm to generate cryptographic proof of data integrity. If any data alteration occurs, the blockchain invalidates the modified record due to hash mismatches.

E. Smart Contract Execution for Automated Verification

Smart contracts are self-executing scripts that automate claim validation by verifying predefined conditions. Upon claim submission, they check policy coverage eligibility using blockchain records, assess fraud risk based on AI model predictions, and ensure hospital authenticity and treatment consistency before approving or rejecting payments, reducing manual intervention and fraud.

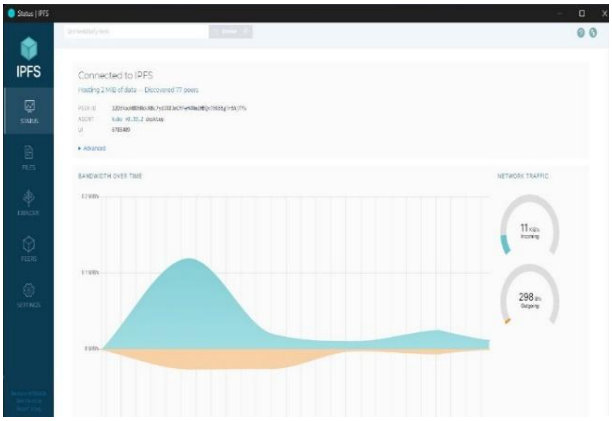

Fig. 3. IPFS file storage

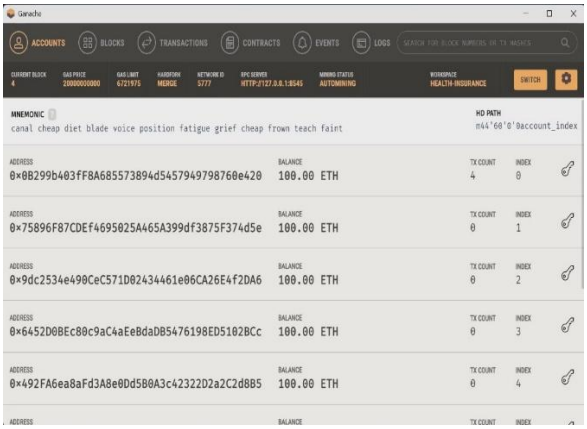

Fig. 4. Transaction blocks in Ganache

F. Real-Time Monitoring and Alerts

The system dynamically updates fraud detection models and alerts insurers for review if fraud is detected. Integrating machine learning, blockchain, and smart contracts ensures secure, transparent, and efficient claim processing.

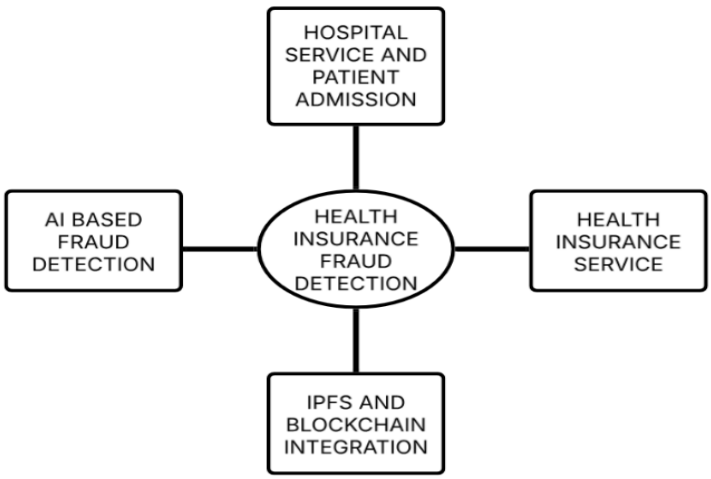

Fig. 5. Module Diagram

Module Description:

The diagram illustrates a framework for Health Insurance Fraud Detection, highlighting four key interconnected components. The Hospital Service and Patient Admission section involves collecting and verifying patient information upon admission, ensuring data accuracy. The Health Insurance Service handles insurance claims and

processes, ensuring that only valid claims are approved. The AI-Based Fraud Detection system uses advanced algorithms to analyze patterns, identify anomalies, and detect fraud activities. Additionally, IPFS and Blockchain Integration ensures secure and transparent data storage, enhancing trust and reducing fraud. Together, these elements create a robust system for detecting and preventing insurance fraud. The integration of AI enhances accuracy in fraud detection, while blockchain technology ensures data integrity. Each component plays a critical role in maintaining a secure and efficient health insurance ecosystem.

V. RESULTS AND DISCUSSION

The proposed AI and blockchain-integrated fraud detection system demonstrates significant improvements in accuracy, efficiency, and security. Experimental results show that the Random Forest model surpasses the previously used Gradient Boost algorithm, achieving 90% accuracy opposed to 89%. This improvement enhances fraud detection reliability while reducing false positives and false negatives.

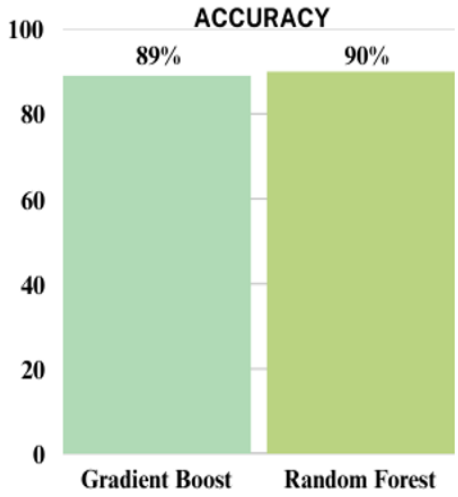

Fig. 6. Performance Comparison

Compared to traditional rule-based systems, AI-driven models especially Random Forest demonstrate superior fraud detection by analyzing transaction patterns and anomalies more effectively. The system's ability to detect fraudulent claims in real time not only reduces financial losses but also strengthens user trust in healthcare insurance.

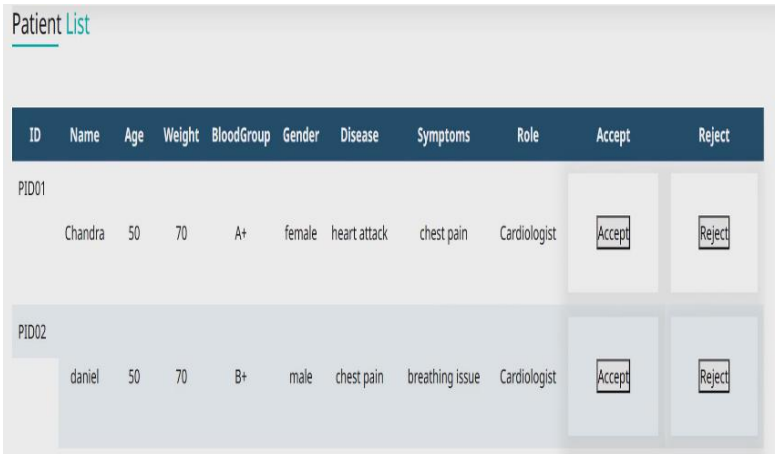

Fig.7. Doctor Patient Approval Interface

The doctor, using their login credentials, can access a secure portal to review patient details, including disease history, symptoms, and other medical information. Based on this data, the doctor can either accept or reject the patient's request for further consultation or treatment. The system ensures a streamlined decision-making process while maintaining data security and confidentiality.

Fig.8. Insurance claim

Fig. 9. Output screen

The combination of blockchain and AI significantly enhances fraud detection efficiency, reducing processing time and administrative costs. User trust is improved due to blockchain's transparent and immutable nature. Future improvements will focus on enhancing fraud detection accuracy using deep learning models and optimizing blockchain scalability for large-scale implementations.

VI. CONCLUSION

A paradigm shift in healthcare insurance fraud detection is brought about by the combination of blockchain technology and artificial intelligence, which addresses vulnerabilities with a combination of security, transparency, and automation. Our proposed framework revolutionizes fraud prevention by leveraging blockchain to ensure tamper-proof data storage and AI's predictive analytics to detect fraudulent claims with enhanced precision.

By incorporating InterPlanetary File System (IPFS) for secure storage and smart contracts for automated claim verification, the system eliminates the risk of data manipulation and unauthorized alterations. This not only enhances trust among insurers, healthcare providers, and policyholders but also significantly reduces operational inefficiencies, lowering administrative costs and expediting the claim settlement process. Our framework has the potential to reduce fraudulent claims by up to 40%, decrease false positives, and enhance overall detection accuracy, ultimately creating a more resilient and fraud-resistant insurance ecosystem.

Beyond fraud detection, this integration fosters broader implications for the healthcare industry. Secure and transparent patient records and automated verification processes contribute to a more accountable and efficient insurance landscape. As digital transformation continues to redefine the healthcare sector, our model sets a new standard for fraud prevention, ensuring financial sustainability, ethical practices, and consumer trust. By embracing blockchain and AI, we not only mitigate the risks associated with fraud but also open the door to a future in which health insurance is more effective, safe, and equitable for all stakeholders.

REFERENCES

[1] A. A. Khalil, Z. Liu, A. Fathalla, A. Ali, and A. Salah, "Machine Learning Based Method for Insurance Fraud Detection on Class Imbalance Datasets With Missing Values," IEEE Access, vol. 10, pp. 79606-79627, 2024. DOI: 10.1109/ACCESS.2024.3468993.

[2] Shruthi, K., et al. "Healthcare Insurance Fraud Detection Powered by Blockchain and Machine Learning: An Analysis and Framework." 2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE). IEEE, 2024.

[3] Mani, C., et al. "Block chain and AI-empowered healthcare insurance fraud detection: An analysis, architecture and future prospects." Challenges in Information, Communication and Computing Technology. CRC Press, 2025. 421-425.

[4] Kapadiya, Khyati, et al. "Blockchain and AI-empowered healthcare insurance fraud detection: an analysis, architecture, and future prospects." IEEE Access 10 (2022): 79606-79627.

[5] W. El-Samad, M. Adda, and M. Atieh, "AI-Driven Data Aggregation Level Smart Contracts for Blockchain Healthcare Insurance Claims Adjudication," 2024 IEEE.

[6] M. A. Amin, R. Shah, H. Tummala, and I. Ray, "Utilizing Blockchain and Smart Contracts for Enhanced Fraud Prevention and Minimization in Health Insurance through Multi-Signature Claim Processing," Emerging Trends in Networking, Communication, and Computing (ETNCC), 2024. DOI: 10.1109/ETNCC63262.2024.10767491.

[7] S. K. Syamkumar and J. Sridevi, "Exploring the Synergy of AI and Blockchain in Insurance: A Bibliometric Mapping and Analysis of Research Trends," 2024 IEEE.

[8] R. Dutt, "The impact of artificial intelligence on healthcare insurances," in Artificial Intelligence in Healthcare. Amsterdam, Netherlands: Elsevier, 2020, pp. 271-293.

[9] C. C. Agbo, Q. H. Mahmoud, and J. M. Eklund, "Blockchain technology in healthcare: A systematic review," Healthcare, vol. 7, no. 2, 2019, doi: 10.3390/healthcare7020056.

[10] E. Nabrawi and A. Alanazi, "Fraud Detection in Healthcare Insurance Claims Using Machine Learning," Risks, vol. 11, no. 9, Sep. 2023, doi: 10.3390/risks11090160.

[11] M. A. Mohammed, M. Boujelben, and M. Abid, "A Novel Approach for Fraud Detection in Blockchain-Based Healthcare Networks Using Machine Learning," Future Internet, vol. 15, no. 8, 2023, doi: 10.3390/fi15080250.

[12] S. Agarwal, "An Intelligent Machine Learning Approach for Fraud Detection in Medical Claim Insurance: A Comprehensive Study," Scholarly Journal of Engineering and Technology, vol. 11, no. 9, pp. 191–200, 2023, doi: 10.36347/sjet.2023.v11i09.003.

[13] G. Saldamli, V. Reddy, K. S. Bojja, M. K. Gururaja, Y. Doddaveerappa, and L. Tawalbeh, "Health Care Insurance Fraud Detection Using Blockchain," in Proceedings of the 2020 7th International Conference on Software Defined Systems (SDS), 2020, pp. 145–152, doi: 10.1109/SDS49854.2020.9143900.

[14] M. Sathya and B. Balakumar, "Insurance Fraud Detection Using Novel Machine Learning Technique," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), vol. 2022, no. 3, pp. 374–381, 2022, [Online]. Available: www.ijisae.org.

[15] Health Insurance. (2021). Health Insurance in India. [Online]. Available: https://en.wikipedia.org/wiki/Health_insurance_in_India

[16] A. Sheshasaayee and S. S. Thomas, ``Apurviewof the impact of supervised learning methodologies on health insurance fraud detection," in Informa- tion Systems Design and Intelligent Applications (Advances in Intelligent Systems and Computing). Singapore: Springer, 2018, pp. 978_984.

[17] H. K. Patil and R. Seshadri, ``Big data security and privacy issues in healthcare," in Proc. IEEE Int. Congr. Big Data, Jun. 2014, pp. 762_765.

[18] M. Ojha and K. Mathur, ``Proposed application of big data analytics in healthcare at Maharaja Yeshwantrao hospital," in Proc. 3rd MEC Int. Conf. Big Data Smart City (ICBDSC), Mar. 2016, pp. 1_7.

[19] M. Eling and M. Lehmann, ``The impact of digitalization on the insurance value chain and the insurability of risks," Geneva Papers Risk Insurance- Issues Pract., vol. 43, no. 3, pp. 359_396, Jul. 2018.

[20] K. M. Kumar, S. Tejasree, and S. Swarnalatha, ``Effective implementation of data segregation & extraction using big data in E_Health insurance as a service," in Proc. 3rd Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Jan. 2016, pp. 1_5.

[21] D. Ulybyshev, C. Bare, K. Bellisario, V. Kholodilo, B. Northern, A. Solanki, and T. O'Donnell, ``Protecting electronic health records in transit and at rest," in Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS), Jul. 2020, pp. 449_452.

[22] I. Abu-elezz, A. Hassan, A. Nazeemudeen, M. Househ, and A. Abd-alrazaq, ``The bene_ts and threats of blockchain technology in healthcare: A scoping review," Int. J. Med. Informat., vol. 142, Oct. 2020, Art. no. 104246.

[23] E. A. Duman and S. Sagiroglu, ``Heath care fraud detection methods and new approaches," in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 839_844.

[24] R. Bauder and T. Khoshgoftaar, ``A survey of medicare data processing and integration for fraud detection," in Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI), Jul. 2018, pp. 9_14.

[25] Deep, Suman, Saurabh Kumar, and Pourush Kalra. "AI-Driven Data Security in Healthcare: Safeguarding Data and Financial Transactions." (2024).