Empowering Health Insurance with Personalization: A Comprehensive Architecture Leveraging Federated Learning and BlockChain

Anant Jain[1] [0009-0000-2859-8911] anant.07019011921@ipu.ac.in, Gauranshi Gupta[1] [0009-0001-9918-3120] gauranshi.00719011921@ipu.ac.in, Rahul Johari*[1] [0000-0002-7675-8550]

rahul@ipu.ac.in, and Deo Prakash Vidyarthi[3] [0000-0003-4151-0552] dpv@mail.jnu.ac.in

[1] SWINGER : **S**ecurity, **W**ireless, **I**oT **N**etwork **G**roup of **E**ngineering and **R**esearch, University School of Automation and Robotics(USAR), Guru Gobind Singh Indraprastha University, Delhi, India

[2] School of Computer and Systems Sciences, Parallel and Distributed System Lab JNU, Delhi, India

Abstract. Historically, the health insurance domain has adhered to uniform premium pricing structures, often sidelining the distinct health nuances of each policyholder. Such broad-brush approaches can result in misaligned premiums, not truly reflective of individual health dynamics. Venturing into uncharted territories, this study proposes an innovative fusion of Federated Machine Learning (FML) and BlockChain methodologies to recalibrate health insurance pricing. FML, a cutting-edge evolution in the machine learning sphere, emphasizes decentralized data processing. By harnessing FML, we can timely insights that mirror an individual's health trajectory. This continuous insight generation facilitates agile adjustments to insurance premiums, ensuring they resonate with real-time health conditions. BlockChain's decentralized ledger system acts as the bedrock of this model. Each premium modification, steered by FML-derived insights, is indelibly etched onto the BlockChain. This architecture not only masters transparency but also instills a heightened sense of trust, given the unalterable nature of BlockChain entries.

Keywords: Machine Learning · Federated Learning · Data Science · BlockChain · HealthCare · Health Insurance

1 Introduction

Machine Learning (ML), an integral branch of artificial intelligence, has rapidly evolved over the past few decades. Unlike traditional computational methods

* corresponding author

that rely on explicit programming, ML models improve autonomously by continuously processing new data which makes it particularly potent in environments with vast and dynamic datasets, such as healthcare. In the realm of health insurance, ML can be a game-changer. Its capability can lead to more accurate premium pricing, tailored health interventions, and proactive policy adjustments based on predicted health trajectories.

However, the power of ML comes with challenges. Centralized ML models often require the aggregation of data in a single location, raising significant concerns about data privacy and security. The sensitivity of health data further amplifies these concerns, necessitating innovative solutions to harness ML's potential without compromising individual privacy.

BlockChain is a decentralized and distributed ledger, where data is stored in cryptographically linked blocks. Each block, once added, is virtually immutable, ensuring data integrity and transparency. In sectors like health insurance, where trust is paramount, BlockChain can provide a transparent and verifiable record of transactions, be it policy adjustments, claim submissions, or premium changes.

1.1 Towards a Synergized Future: Federated ML on BlockChain

Merging the predictive prowess of ML with the transparency and security of BlockChain leads us to an innovative proposition: a dynamic health insurance system where premiums are not static but evolve based on real-time health data. By leveraging FML, a decentralized approach to ML, we can train models on localized data without central aggregation. When combined with BlockChain's transparent ledger system, we get a framework where premium adjustments are not only personalized but also verifiable and tamper-proof.

This research delves into the intricacies of such integration, exploring its potential to reshape the health insurance landscape, the technical challenges it presents, and the broader societal implications of a truly personalized and transparent health insurance system.

2 Problem Statement

The current health insurance landscape often overlooks individual health profiles, leading to generalized premium rates. This approach can unfairly burden healthier policyholders by making them subsidize higher-risk individuals. To address this, our study proposes a personalized insurance model, leveraging Machine Learning to analyze individual health data and offer more accurate premium rates. Complementing this, BlockChain technology introduces a layer of transparency and security. Its smart contracts, informed by data-driven insights, can autonomously adjust premiums, ensuring a more dynamic and trustworthy insurance system. This research evaluates the feasibility and benefits of integrating these technologies to enhance fairness and efficiency in health insurance.

3 Literature Survey

In [1], a BlockChain-driven encryption method combined with federated learning is presented to enhance electronic health records (EHR) security. The system focuses on protecting remote patient records with end-to-end encryption. Dynamic smart contracts and an efficient re-encryption mechanism are used. Tests on Ethereum showed promising results.

In [2], traditional machine learning centralizes data, causing privacy and communication issues. BlockChain-based federated learning protects AIoT devices without central server vulnerabilities. The research proposes a cross-chain framework for scalable training and a model update compression scheme, showing a decentralized system for AIoT.

In [3], the integration of machine learning with the Internet of Medical Things (IoMT) is discussed. Traditional models face data fraud challenges in IoMT. The paper introduces FL-BETS, a federated learning-based framework, focusing on energy consumption and deadlines.

In [4], Federated learning (FL) allows users to collaboratively solve machine learning problems while keeping data private. A systematic review on FL, Machine Learning, and BlockChain Technology is presented. Out of many papers, 84 met specific criteria. The review highlights growing interest in ML using FL, but challenges remain.

In [5], solutions to healthcare challenges posed by Big Data are discussed. Distributed learning, which analyzes databases without centralizing patient data, is highlighted as a solution. This method shares research insights without transferring actual data, ensuring patient privacy.

In[6], Artificial intelligence (AI) has significantly impacted the healthcare industry, aiding in diagnosis, personalized care plans, and managing the COVID-19 pandemic through symptom prediction and accelerating research. However, developing robust AI models for real-time application is challenging due to privacy concerns and reluctance to share data, along with difficulties in creating generalized models from fragmented patient data. This paper proposes integrating BlockChain and AI, utilizing BlockChain for secure data access and AI-based federated learning to develop reliable models suitable for global and real-time application.

In[7], paper introduces a BlockChain-based federated learning approach for smart healthcare, addressing data privacy, security, and service quality in the Medical Internet of Things (MIoT). It utilizes edge nodes and MIoT devices for distributed clinical data processing, incorporating an adaptive differential privacy algorithm for data protection and a gradient verification consensus protocol to counteract poisoning attacks. Tested on a real-world diabetes dataset, this method demonstrates high accuracy, efficient run-time, and effective privacy budget management, while also being resilient to poisoning attacks.

In[8], article proposes a secure architecture for smart healthcare in smart cities, integrating BlockChain and Federated Learning technologies. As IoT deployment escalates, protecting user data becomes crucial. The proposed BlockChain-based IoT cloud platforms ensure data security and privacy, while Federated

Learning enables scalable, privacy-preserving machine learning applications in healthcare. This approach allows users to benefit from advanced machine learning models without compromising personal data privacy. Additionally, the article explores the use of federated learning for creating a distributed, secure environment within a smart city.

In[9], paper introduces a lightweight hybrid federated learning (FL) framework, enhancing the security and privacy of Internet of Health Things (IoHT) data. Addressing the need for data provenance, accuracy, and integrity in IoHT applications, the framework combines FL, differential privacy (DP), and BlockChain technology. It allows private IoHT data training on-premises or on edge devices, managed by BlockChain smart contracts for edge training plans, trust, authentication, and reputation. The framework employs additive encryption at federated edge nodes and multiplicative encryption via BlockChain, ensuring full dataset and model encryption, including during inference. Lightweight DP is integrated for complete IoHT data anonymization. Tested in deep learning applications for COVID-19 clinical trials, the framework shows promise for secure IoHT-based health management adoption.

In[10], This work explores how BlockChain technology, particularly through Ethereum-based smart contracts, enhances healthcare data management by ensuring security, privacy, and efficient access. The focus is on using BlockChain to securely track medical devices, drugs, claims, payments, and data exchange in healthcare. The Ethereum framework facilitates the creation of healthcare smart contracts, aiding in managing patient records, claim payments, and maintaining data security and privacy. It highlights a BlockChain-based intelligent contract, designed and developed using Solidity language on REMIX IDE, emphasizing its potential to secure electronic medical records and streamline healthcare processes.

In[11], This paper discusses the integration of BlockChain technology in healthcare for securing patient data collected through various smart devices and systems. As healthcare increasingly relies on big data stored in the cloud for smart diagnosis and health management, powered by AI and ML, data security and privacy emerge as major concerns. The paper presents a plugin developed for health management systems, ensuring data storage on the BlockChain using the Secure Hash Algorithm (SHA-256) for enhanced security. The application allows secure data storage and sharing between providers. Its effectiveness is demonstrated by storing and creating a BlockChain network with liver and chronic kidney disease datasets from Kaggle. This represents a significant step toward practical enterprise-level applications of BlockChain in healthcare.

[12] This research introduces "Medichain," a BlockChain framework developed to enhance the security of Electronic Medical Records (EMR) in hospital management systems. Recognizing the increasing threats from hackers to electronic records, this work proposes using BlockChain containers on multiple ports to store patient medical records securely. Medichain contains essential BlockChain functionalities to protect patient data. It stores each patient's records in individual blocks of the BlockChain, allowing users to upload records

as JSON files within a secure, distributed, and decentralized network. The framework's BlockChain algorithm includes cryptographic hash, proof of work, and a Merkle tree formulation, all simulated in Python. The results of this implementation have been positive, suggesting a promising approach to safeguarding patient data in healthcare systems.

In [13] the author told that Federated Learning (FL) is a collaboratively decentralized privacy-preserving technology to overcome challenges of data silos and data sensibility. This study aims to review prevailing application in industrial engineering to guide for the future landing application.

In [14], the author elucidates the transformative potential of federated learning in the realm of healthcare data integration. With burgeoning access to disparate healthcare data, the paper underscores the challenge of fragmentation and privacy constraints. It highlights federated learning as a promising solution, allowing shared model training while preserving sensitive data locally. The survey delves into statistical and system challenges, addressing privacy issues in federated learning. The findings emphasize the profound implications and potentials of this approach for advancing data-driven insights and enhancing healthcare quality.

In [15], the author scrutinizes the application of BlockChain technology primarily within the insurance domain, emphasizing its efficacy in fostering efficient transactions, real-time transparency, and the execution of smart contracts. The literature notes a scarcity of academic exploration into leveraging BlockChain for mitigating information asymmetry and combating insurance fraud. The study posits a theoretical framework wherein BlockChain adoption addresses information asymmetry in health insurance, facilitating secure and effective sharing of medical information. It further contends that BlockChain technology plays a pivotal role in curbing soft insurance fraud through real-time medical information sharing and automated claims processing.

In [16], the author underscores the critical role of health insurance in accessing quality healthcare amidst rising costs. However, due to fragmented and unsynchronized data across providers, healthcare fraud has surged. The National Health Care Anti-Fraud Association estimates annual losses in the billions. Proposing a solution, the paper advocates for a BlockChain-based system to securely integrate and monitor insurance activities, leveraging BlockChain's immutable data capabilities for effective health insurance fraud detection.

In [17], the author addresses the escalating concern over rising healthcare insurance fraud incidents. With fraud taking diverse forms, the existing system's resource-intensive nature exacerbates the issue. The paper proposes a BlockChain-based healthcare insurance anti-fraud system, leveraging cloud computing architecture. The system incorporates key entities like command centers, traffic control, judicial bodies, and insurance agencies, forming a comprehensive BlockChain network. Utilizing medical data, including expenses, prescriptions, and treatment records, the system facilitates services such as medical process inspection, analysis of excessive medical behavior, third-party liability inspection, and medical invoice data review.

In [18], the author explores the evolution of BlockChain applications, termed BlockChain 3.0, following Bitcoin's success. Recognizing BlockChain's attributes like enhanced security and privacy, the paper addresses the pervasive issue of fraud in the National Health Insurance Scheme (NHIS) in Ghana. With fraud and data-related problems threatening financial stability, the study proposes a BlockChain-based solution to safeguard the NHIS. The paper presents a comprehensive system design, including conceptual views, sequence and use case diagrams, a data management framework, a smart notification system, and a smart claim processing system. Evaluation using the DeLone McLean Information Systems Success Model highlights the significant influence of information quality and user satisfaction on system success.

In [19], the authors examine the role of Federated Learning (FL) in revolutionizing smart healthcare amid advances in communication technologies and the Internet of Medical Things. Addressing challenges of centralized AI techniques in healthcare, the survey explores FL's recent advances, motivations, and requirements. It covers diverse FL designs and reviews applications in health data management, remote monitoring, medical imaging, and COVID-19 detection. The paper analyzes FL-based smart healthcare projects, distills key lessons, and outlines research challenges, offering insights into future directions for FL in healthcare research.

4 Methodology Adopted

1. User Onboarding:

Upon deciding to use our health insurance system, the user begins by registering on our platform. This involves providing basic personal details and setting up secure authentication methods. In the backend, the user's registration data is encrypted and stored securely.

2. Data Submission:

The user is prompted to upload their Electronic Health Records (EHR) or any relevant health data onto the platform. Then uploaded data undergoes encryption using AES to ensure its security. Once encrypted, the data is stored in a decentralized cloud system, ensuring that there's no single point of failure.

3. Data Analysis and Model Training:

With the user's consent, the system uses their health data to train personalized machine learning models using Federated Learning. The Federated Averaging Algorithm is employed to train models across decentralized datasets which ensures user privacy.

4. Premium Calculation:

Based on the insights derived from the user's data and the trained model, a personalized insurance premium is calculated. The machine learning model assesses the user's health risks and predicts future health expenses. This prediction is then used to determine a fair premium.

5. Transparent Logging with BlockChain:

All activities, including premium calculations, model updates, and data access, are logged on the BlockChain. In the backend, smart contracts are triggered to log these activities.

6. Policy Issuance:

Once the user agrees to the calculated premium, they can proceed to purchase the health insurance policy. The policy details and the agreed premium and terms are stored on the BlockChain using the smart contract. This ensures that the policy terms are tamper-proof.

7. Continuous Updates:

As users continue to update their health data or undergo medical check-ups, they can upload new data to the platform. The system continuously retrains the model with the new data, ensuring that the insurance premium remains reflective of the user's current health status. All updates are, again, logged transparently on the BlockChain.

9. Claims and Verification:

In the event of a health issue, the user can file a claim on the platform. The claim is verified against the user's health data and the terms of the policy stored on the BlockChain. Once verified, the claim is processed, ensuring a transparent and tamper-proof claims process.

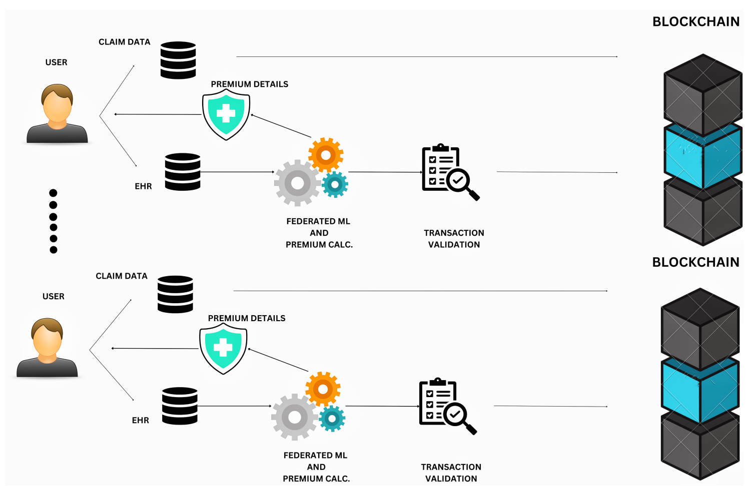

Fig. 1: System architecture of Personalized Health Insurance System.

4.1 Algorithm Utilized

In the health insurance system, a combined approach of Federated Learning and BlockChain is employed. **The federated averaging algorithm,** as detailed in [6], is a key component of this methodology which can be interpreted as:

1. Initialization: A global model M is initialized on the server. The model parameters are denoted as w.

2. Client Selection: At each round t, a subset of m (m $\in$ K) clients are selected to participate, out of a total of K clients. This subset is chosen randomly.

3. Client Training: Each selected client ki computes an update based on its local dataset ki[D]. The client trains the model on its local data to produce a set of updated weights w_k^t.

$$w_k^{(t)} = w - \eta \nabla L_k(w)$$

Where:

η is the learning factor.

L_k^t is the loss of client's data.

4. Model Aggregation: The server aggregates the updated weights from the selected clients to produce a global model update.

$$w^{(t+1)} = \frac{1}{m} \sum_{k=1}^{m} w_k^{(t)}$$

The global model weights are updated with the averaged weights.

5. Repeat: Steps 2-4 are repeated for a number of communication rounds until convergence.

The BlockChain model can be showcased as:

```

records = ARRAY of {timestamp, recordType, data}
owner = SET AS CurrentUserAddress()

FUNCTION AddRecord(type: ENUM, data: STRING)
    IF CurrentUserAddress() == owner THEN
        newRecord = {timestamp: CurrentTimestamp(), recordType: type,
            data: data}
        APPEND newRecord TO records
    END IF
END FUNCTION

FUNCTION GetRecord(index: INTEGER) RETURNS Record
    RETURN records[index]
END FUNCTION

FUNCTION GetTotalRecords() RETURNS INTEGER
    RETURN Length(records)
END FUNCTION
```

5 Flowchart

Here is a flowchart illustrating the key steps and processes of our algorithm. Refer to Figure 2.

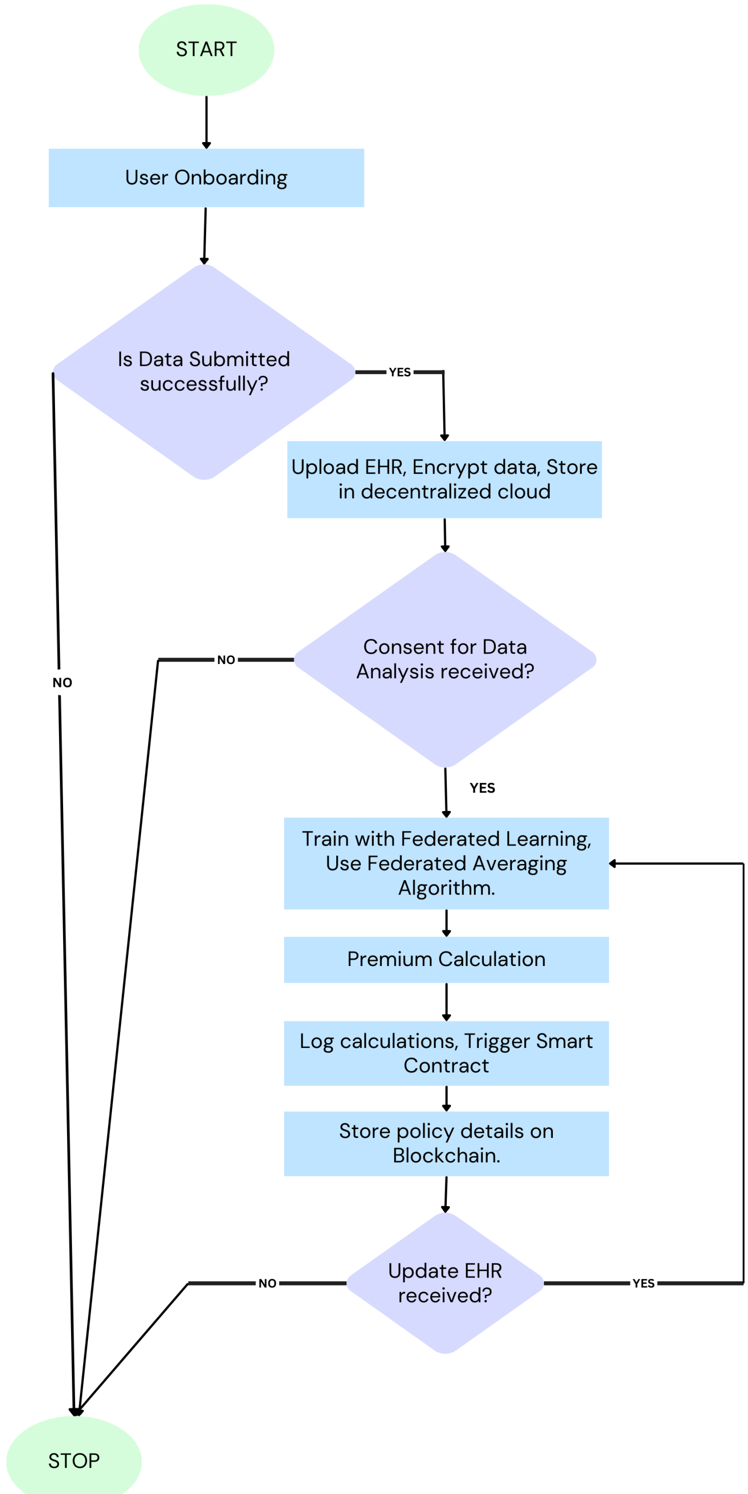

Fig. 2: Flowchart illustrating the Personalized Health Insurance System

6 Results

6.1 Data Encryption:

Encrypting user data ensures confidentiality. The encryption ensures that unauthorized entities cannot decipher the original data without the decryption key.

6.2 Premium Calculation:

By leveraging the Federated Averaging algorithm, we can achieve a more granular and personalized risk assessment and policyholders are likely to receive premium values that are more attuned to their actual health conditions and lifestyles. Consequently, many users might find themselves paying a more reasonable amount, potentially lower than what they would under a traditional architecture that doesn't account for individual nuances.

6.3 Data Analysis and Federated Learning:

Federated Learning operates on the principle of decentralized machine learning, where data remains localized which can reduce data transfer and enhance privacy, and can be contrasted against traditional centralized learning.

6.4 Continuous Updates:

In real-life scenarios, health conditions, lifestyle choices, and medical advancements are constantly evolving. The ability to continuously update the model ensures that the insurance system remains relevant and accurate in its risk assessments. By regularly integrating new data, the model can converge faster to an optimal solution, ensuring that predictions remain consistent and reliable. Furthermore, this distributed approach not only speeds up the learning process but also ensures that the model reflects the most recent and diverse data from various sources, leading to a more comprehensive and up-to-date understanding of policyholders' health risks.

7 Conclusion and Future Work

7.1 Conclusion: In this research, we delved into the intricate architecture of a personalized health insurance system, leveraging the strengths of Federated Learning and BlockChain. The proposed system promises to revolutionize the way health insurance operates by ensuring data privacy, enhancing transparency, and offering personalized premium calculations. While the theoretical framework has been laid out meticulously, the real test of its efficacy will be in its practical implementation, which we anticipate will address the existing challenges in the health insurance domain.

7.2 Future Work: Moving forward, our primary focus will be on the practical implementation and real-world testing of the proposed system. We aim to collaborate with healthcare institutions and insurance providers to gather real-world data and feedback. Additionally, we will explore the integration of more advanced machine learning models and BlockChain optimizations to further enhance system efficiency and scalability. As technology and healthcare landscapes evolve, we also plan to continuously update our system to cater to emerging needs and challenges.

8 References

1. Guduri, Manisha, Chinmay Chakraborty, and Martin Margala. "Blockchain-based Federated Learning Technique for Privacy Preservation and Security of Smart Electronic Health Records." IEEE Transactions on Consumer Electronics (2023).

2. Kang, Jiawen, Xuandi Li, Jiangtian Nie, Yi Liu, Minrui Xu, Zehui Xiong, Dusit Niyato, and Qiang Yan. "Communication-efficient and cross-chain empowered federated learning for artificial intelligence of things." IEEE Transactions on Network Science and Engineering 9, no. 5 (2022): 2966-2977.

3. Lakhan, Abdullah, Mazin Abed Mohammed, Jan Nedoma, Radek Martinek, Prayag Tiwari, Ankit Vidyarthi, Ahmed Alkhayyat, and Weiyu Wang. "Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare." IEEE journal of biomedical and health informatics 27, no. 2 (2022): 664-672.

4. Ogundokun, Roseline Oluwaseun, Sanjay Misra, Rytis Maskeliunas, and Robertas Damasevicius. "A review on federated learning and machine learning approaches: Categorization, application areas, and blockchain technology." Information 13, no. 5 (2022): 263.

5. Zerka, Fadila, Samir Barakat, Sean Walsh, Marta Bogowicz, Ralph TH Leijenaar, Arthur Jochems, Benjamin Miraglio, David Townend, and Philippe Lambin. "Systematic review of privacy-preserving distributed machine learning from federated databases in health care." JCO clinical cancer informatics 4 (2020): 184-200.

6. Aich, Satyabrata, Nday Kabulo Sinai, Saurabh Kumar, Mohammed Ali, Yu Ran Choi, Moon-IL Joo, and Hee-Cheol Kim. "Protecting personal healthcare record using blockchain federated learning technologies." In 2022 24th International Conference on Advanced Communication Technology (ICACT), pp. 109-112. IEEE, 2022.

7. Chang, Yuxia, Chen Fang, and Wenzhuo Sun. "A blockchain-based federated learning method for smart healthcare." Computational Intelligence and Neuroscience 2021 (2021).

8. Singh, Saurabh, Shailendra Rathore, Osama Alfarraj, Amr Tolba, and Byungun Yoon. "A framework for privacy-preservation of IoT healthcare data using Federated Learning and blockchain technology." Future Generation Computer Systems 129 (2022): 380-388.

9. Rahman, Mohamed Abdur, M. Shamim Hossain, Mohammad Saiful Islam, Nabil A. Alrajeh, and Ghulam Muhammad. "Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach." Ieee Access 8 (2020): 205071-205087.

10. Aggarwal, Harsha, Rahul Johari, Deo Prakash Vidyarthi, Kalpana Gupta, and Arun Arora. "BEST: Blockchain-Enabled Secure Technology in a healthcare management system." In Blockchain Applications for Healthcare Informatics, pp. 267-282. Academic Press, 2022.

11. Khandelwal, Parth, Rahul Johari, Medha Chugh, and Anmol Goel. "EMBRACE: Electronic Medical Record Safety, Blockchain to the Rescue." In Pro-

ceedings of Data Analytics and Management: ICDAM 2021, Volume 2, pp. 605-613. Singapore: Springer Singapore, 2021.

12. Johari, Rahul, Vivek Kumar, Kalpana Gupta, and Deo Prakash Vidyarthi. "BLOSOM: BLOckchain technology for Security of Medical records." ICT Express 8, no. 1 (2022): 56-60.

13. Li, Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin. "A review of applications in federated learning." Computers Industrial Engineering 149 (2020): 106854.

14. Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang. "Federated learning for healthcare informatics." Journal of Healthcare Informatics Research 5 (2021): 1-19.

15. Park, Daehyeon, and Doojin Ryu. "Blockchain in health insurance: sharing medical information and preventing insurance fraud." Korean Journal of Financial Studies 48, no. 4 (2019): 417-447.

16. Saldamli, Gokay, Vamshi Reddy, Krishna S. Bojja, Manjunatha K. Gururaja, Yashaswi Doddaveerappa, and Loai Tawalbeh. "Health care insurance fraud detection using blockchain." In 2020 Seventh International Conference on Software Defined Systems (SDS), pp. 145-152. IEEE, 2020.

17. Liu, Wei, Qinyong Yu, Zesong Li, Zeyuan Li, Yu Su, and Jian Zhou. "A blockchain-based system for anti-fraud of healthcare insurance." In 2019 IEEE 5th International Conference on Computer and Communications (ICCC), pp. 1264-1268. IEEE, 2019.

18. Amponsah, Anokye Acheampong, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. "Improving the financial security of national health insurance using cloud-based blockchain technology application." International Journal of Information Management Data Insights 2, no. 1 (2022): 100081.

19. Nguyen, Dinh C., Quoc-Viet Pham, Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, and Won-Joo Hwang. "Federated learning for smart healthcare: A survey." ACM Computing Surveys (CSUR) 55, no. 3 (2022): 1-37.

20. Prayitno, Chi-Ren Shyu, Karisma Trinanda Putra, Hsing-Chung Chen, Yuan-Yu Tsai, K. S. M. Tozammel Hossain, Wei Jiang, and Zon-Yin Shae. 2021. "A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications" Applied Sciences 11, no. 23: 11191.