Improving Security in Healthcare Applications using Federated Learning System with Blockchain Technology

Aofan Liu  
School of Computing & Data Science  
Xiamen University MalaysiaSelangor, Malaysia  
SWE2009510@xmu.edu.my

Qianqian Tan  
School of Computing & Data Science  
Xiamen University MalaysiaSelangor, Malaysia  
SWE2009510@xmu.edu.my

Dr.Burra Venkata Durga Kumar\*  
School of Computing & Data Science  
Xiamen University MalaysiaSelangor, Malaysia  
SWE2009510@xmu.edu.my

*Abstract*— Data security is of the utmost importance in the healthcare area, as sensitive patient information is constantly sent around and analyzed by many different parties. The use of federated learning, which enables data to be evaluated locally on devices rather than being transferred to a central server, has emerged as a potential solution for protecting the privacy of user information. To protect against data breaches and unauthorized access, federated learning alone might not be adequate. In this context, the application of blockchain technology could provide the system extra protection. This study proposes a distributed federated learning system that is built on blockchain technology in order to enhance security in healthcare. This makes it possible for a wide variety of healthcare providers to work together on data analysis without raising concerns about the confidentiality of the data. The technical aspects of the system, including as the design and implementation of distributed learning algorithms, consensus mechanisms, and smart contracts, are also investigated as part of this process. The technique that was offered is a workable alternative that addresses concerns about the safety of healthcare while also fostering collaborative research and the interchange of data.

Keywords—Data privacy, Distributed system, Federated learning, Machine learning

# Introduction

In today's healthcare sector, safeguarding sensitive patient information is a critical issue as it circulates among various entities for analysis and utilization. Undoubtedly, data security takes precedence in ensuring the privacy and confidentiality of user information. Federated learning (FL) is a novel approach and it has emerged as a potential solution for safeguarding data privacy in the healthcare domain [1]. As a distributed machine learning technique, federated learning allows indirect data sharing through the sharing of model parameters, but it relies on third-party servers.

Technologically, federated learning disrupts data isolation, allowing for the efficient sharing and utilization of extensive user data. This has gained considerable interest from both the academic and industrial sectors. Nevertheless, depending solely on federated learning may not provide adequate protection against data breaches and unauthorized access.

The conventional Federated Learning framework encounters several problems, with the most notable being the communication model based on a master-slave architecture. This communication model is vulnerable to attacks from malicious nodes in distributed environments, significantly impacting the resilience of FL models. Nguyen et al. addressed the issue of single point failures in traditional Federated Learning, where a bottleneck caused by a single point failure can disrupt the entire Federated Learning system [3]. Similarly, Zhilin et al. provided a detailed explanation of the critical influence of any issues occurring in the central server on the entire Federated Learning system in traditional Federated Learning systems. Moreover, dishonest participants can also have an impact on related issues [4].

Therefore, we urgently need to develop a more decentralized Federated Learning method without using a central server, in order to solve the security and scalability issues, and thus realize the next generation of intelligent edge networks.

In order to enhance the security infrastructure, the utilization of blockchain technology presents a promising pathway. Integrating blockchain into the system provides an additional layer of protection, reducing the risks associated with data breaches and unauthorized access. This approach enables collaboration among various healthcare providers for data analysis while preserving the confidentiality of sensitive information. The overall project architecture is roughly illustrated in the following diagram.

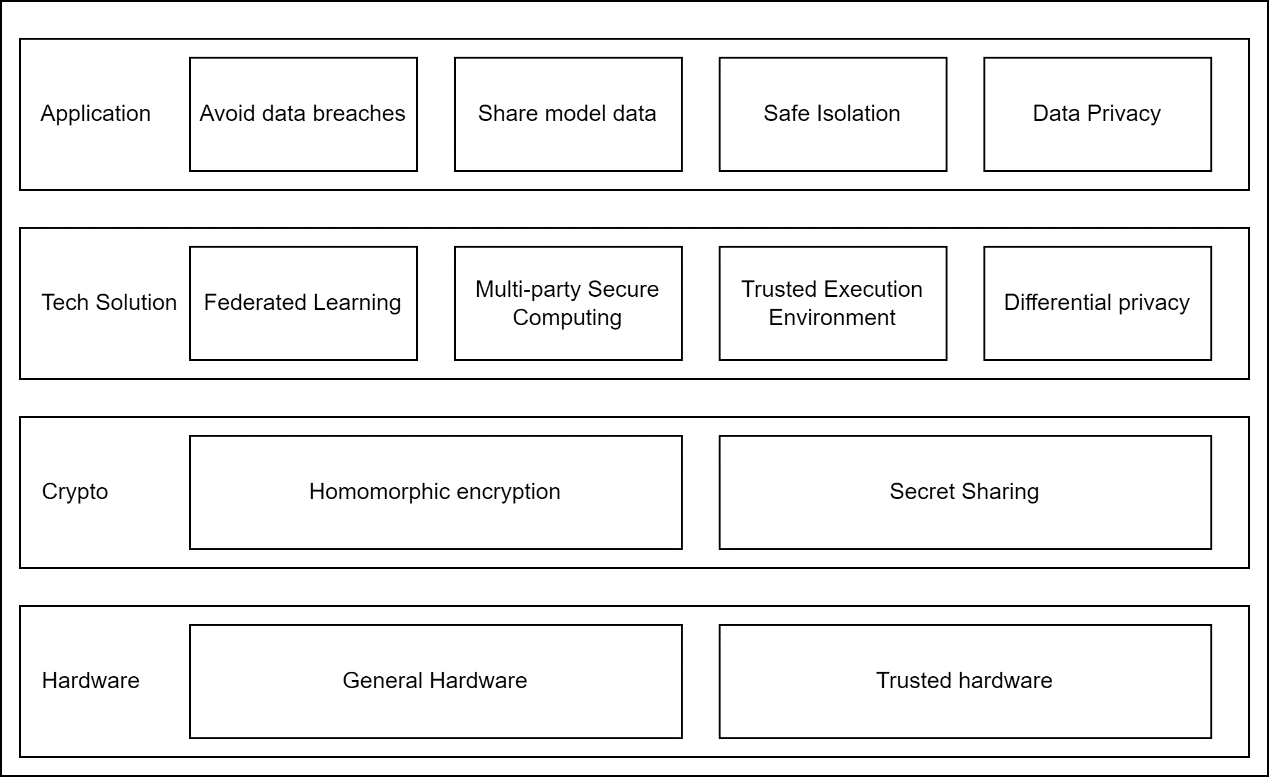


Fig. 1 Overall Structure \*Adapted From [5]

This viable alternative addresses issues surrounding healthcare security while facilitating collaborative research and facilitating data exchange between healthcare providers.

In this approach, we have developed a training member selection algorithm based on similarity, incorporating the use of Euclidean distance to enhance training efficiency. The algorithm aims to identify data owners with similar characteristics to the requirements of the data requester, thereby forming a federated community responsible for training the federated model [5]. By selecting a few highly matched training nodes, our goal is to improve training efficiency and effectiveness.

In summary, our study has made remarkable progress in ensuring data security within the healthcare sector. By integrating federated learning and blockchain technology, the technical exploration and practical implementation of this system provide a promising solution to enhance security, foster collaboration, and facilitate data exchange among healthcare professionals.

# Problems

Prentosito et al.[6] cited the healthcare data breach that happened in March 2021 and created user angst, fear, and concern. As the number of healthcare data collecting sites and the organisations that use the data continues to rise, the amount of data that must be transmitted to and from the point also continues to grow, which eventually raises the risk of data leakage and cybercrimes. In this scenario and considering issues such as intellectual property and privacy protection, models involving healthcare data tend to choose the federal learning system. Although the localized training model of the federal learning system can reduce the leakage of sensitive data, it cannot effectively reduce the impact of cybercrimes [7]. In this regard, blockchain can be introduced to solve this problem. Blockchain can provide immutable tamper-proof records, and has the characteristics of distribution and decentralization, which can reduce the risk of healthcare data forgery and tampering. Its paired keys can ensure normal data access control and reduce unauthorized access.

# Objectives

The primary aim of this essay is to offer a solution to the issue of healthcare data security, namely how to guarantee the privacy, availability, and integrity of healthcare data shared among healthcare organisations. The following three goals are suggested in order to accomplish and guarantee the accomplishment of the above objectives:

1. To prevent the widespread distribution of raw data by employing a self-training technique for every healthcare data point and a federated learning system to assure data confidentiality.
2. To use blockchain technology to assure the availability and integrity of transmitted data due to its distributed storage and tamper-evident design.
3. Find a method to combine the blockchain technology and federated learning system in order to maximize the data security

# RESEARCH QUESTION

RQ1: How to combine federated learning system with blockchain technology?

RQ2: How to maximize the improvement on security with the above methodology?

# FRAMEWORK & METHOD

Based on current research, we believe that the literature study can be effectively divided into three key sections and organized using the following models:

## Overlapped model

Three overlapped model will be discussed here. Lu et al. [8] proposed a digital twin wireless networks (DTWN) model. It is implemented by assigning training data and computing tasks for training to each BS based on the association between digital twins and BSs. Rahman et al. [9] suggested a simple framework , which architecture employs additive and multiplicative encryption. Client devices and compound nodes apply DP and introduce noise to balance the trade-offs between accuracy and privacy budget in order to protect data privacy. Chai et al. [10] proposed a model whose collection of local models is performed by a consensus process and the learned knowledge is packed into blocks and hierarchically agreed by FR in GC and BS in TC.

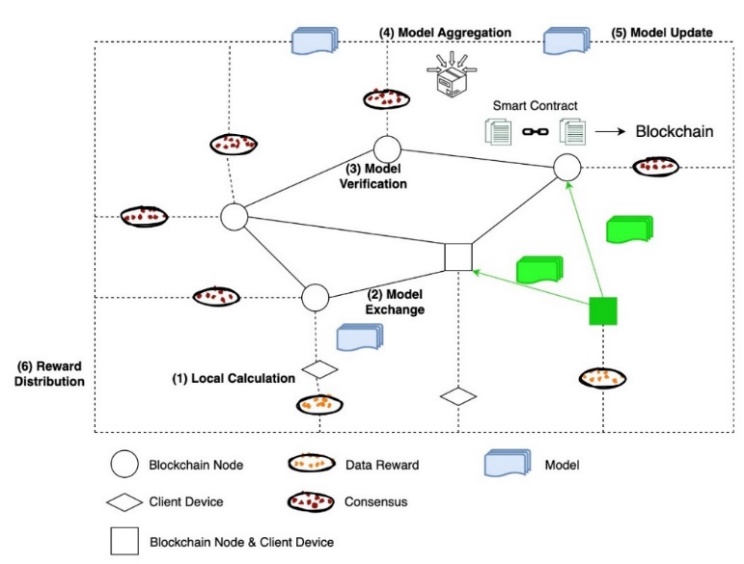


Figure 2: Workflow of overlapped model [5]

## Coupled Model

Two frameworks of coupled model will be discussed here. Y. Qu et al. [1] proposed a framework, selecting miners and related devices to come together to form an ephemeral aggregator that broadcasts and updates global parameters across the blockchain while using global model parameters for training. Wang et al. [11] proposed SFAC, a local differential privacy (LDP)-based framework for drone-assisted MCS to achieve ideal aggregation accuracy and strict privacy preservation via on-device perturbations.

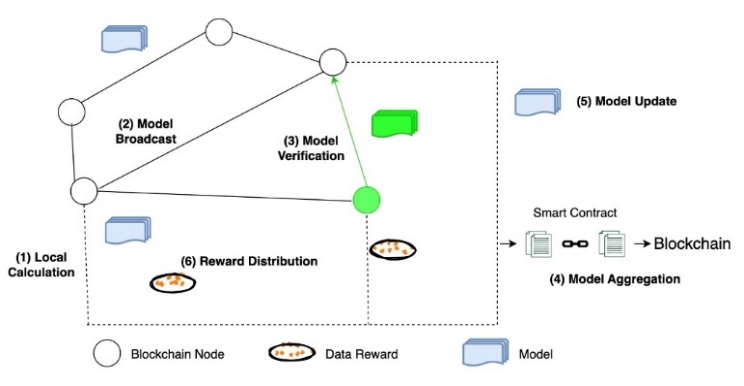


Figure 3: Workflow of coupled model [7]

## Decoupled Model

Two frameworks of decoupled model will be discussed here. Kumar et al. [12] proposed a framework, which is a multi-organizational blockchain design that divides participants into regions, and distance calculation and a node weight matrix allow data retrieval. Lu et al. [8] proposed a framework, which Node Selection formulates and solves optimization problems by using a DRL algorithm to select participating vehicles. The selected data is then trained locally and updates its trained local model for global aggregation.

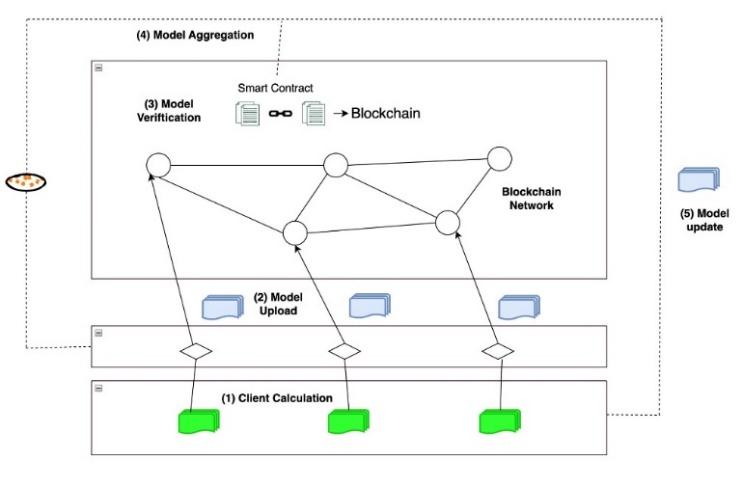


Figure 4: Workflow of decoupled model

# Findings& Discussion

The coupled model of Federated Learning System with Blockchain has a low communication overhead since the topology of the entire system is the same as that of each subsystem. The coupled model's development costs are further decreased by the seamless integration of blockchain and federated learning. However, each node in the coupled model must carry out labour-intensive tasks including consensus and model training. This action raises the hardware cost, which is detrimental to the advertising of the system for the healthcare industry, which requires a significant amount of calculation [13]. Importantly, the coupled model goes against the intention of maintaining information security because each node has more obligations than nodes in a single subsystem [14].

The overlapped model of Federated Learning System with Blockchain can optimise the blockchain and federated learning roles given to nodes based on their resource availability, security level, etc. Although overlapping models have distinct advantages, they also have drawbacks, much as coupled models [15]. Because different nodes in the linked model play very distinct functions in the total system, it is inevitably more complex to design and configure, which is detrimental to system advancement.

In contrast, each node only takes part in one subsystem, such as federated learning or blockchain, in decoupled model of Federated Learning System with Blockchain. The decoupled paradigm does not require nodes to deliver sufficient resources, as opposed to participating in two subsystems [16]. Furthermore, demonstrating the simplicity of design and configuration, the design and configuration of a single subsystem is adequate to meet the requirements of the complete blockchain-enabled federal learning system [17]. However, there are still drawbacks, such as a significant communication overhead brought on by the substantial node connectivity.

In contrast, each node only takes part in one subsystem, such as federated learning or blockchain, in decoupled model of Federated Learning System with Blockchain. The decoupled paradigm does not require nodes to deliver sufficient resources, as opposed to participating in two subsystems [16]. Furthermore, demonstrating the simplicity of design and configuration, the design and configuration of a single subsystem is adequate to meet the requirements of the complete blockchain-enabled federal learning system [17]. However, there are still drawbacks, such as a significant communication overhead brought on by the substantial node connectivity.

# RECOMMENDATIONS

In this paper, we propose a data sharing model that combines blockchain and federated learning. This model integrates the underlying architecture of blockchain technology as suggested by Y. Qu et al., the parameter aggregation among federated nodes proposed by Rahman et al. [9], and the advantages of the multi-organizational structure highlighted by Kumar et al.

At the top level of the architecture, we recommend using the consensus algorithm-based federated learning approach advocated by [1]., as it performs exceptionally well in the proposed training phase. They have demonstrated the existence of a unique Nash equilibrium in the dynamic game of federated learning node selection. Additionally, the blockchain-based consensus mechanism provides continuous incentives for well-performing miners, which is highly beneficial for federated learning.

We suggest incorporating the concept of federated communities introduced by Zhang et al. [5], ensuring training performance by selecting a small number of highly matched training nodes. The data standardization proposed by Kumar et al. should also be included in our model to address the problem of diverse data sources. Rahman et al.'s approach of utilizing blockchain smart contracts to manage edge training plans and federated node authentication should also be kept. In this manner, each node anonymously performs additive encryption and allows the blockchain to execute multiplicative encryption for updating the aggregated model. This effectively resolves the single-point failure problem highlighted by Wang et al. [11] in traditional federated learning and addresses the issue of potentially dishonest clients.

The blockchain-based smart contract utilizes a similarity-based training member selection algorithm to choose qualified data owners from a decoupled model-based pool of data owners, creating a federated learning community. Member nodes within the community employ local data to train their individual models, incorporating threshold homomorphic encryption and differential privacy techniques.

These trained local models, encrypted with homomorphic encryption, are transmitted to data requesters for model aggregation. Subsequently, cooperative decryption operations are performed to generate a differentially private global model. The overall process is shown in the figure below:

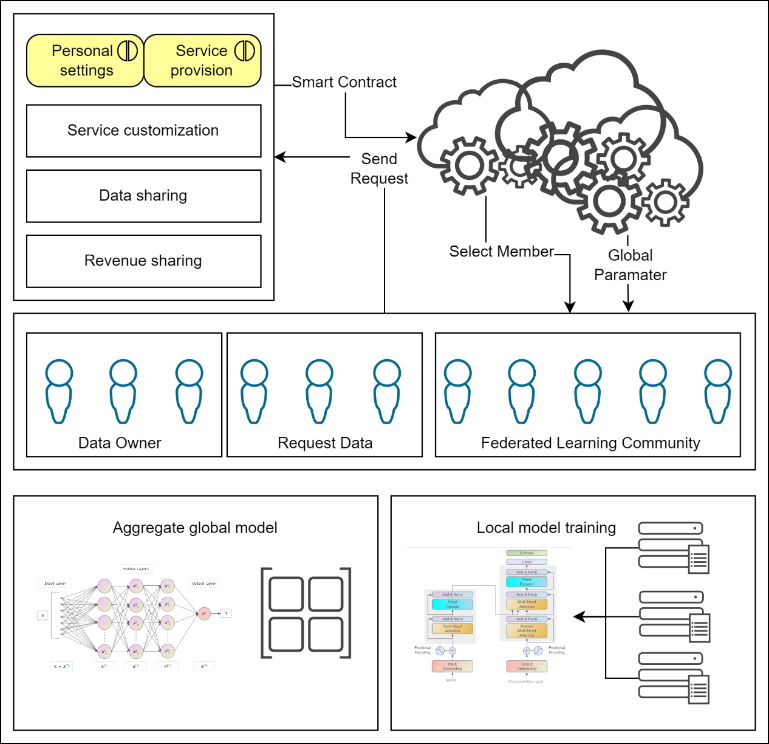


Figure 6: Overall Structure

# RECOMMENDATIONS

With continued technological improvement and the demand for security and privacy protection in the healthcare industry continues to grow, federated learning system and blockchain technology will play an increasingly important role in the healthcare area. In this research, a data sharing model combining a federated learning system with blockchain technology at Euclidean distance is proposed to address the issues of how to combine a federated learning system with blockchain technology and how to maximise security.

The paper synthesises several previously proposed models and classifies them into three categories, including decoupled model, coupled model and overlapped model, discusses their strengths and weaknesses, and partially adopts and improves them. The blockchain as the underlying architecture, the federated node parameter aggregation and the multi-organisational architecture can satisfy Objective 1 and 2 well.

This method achieves privacy while sharing data without affecting the model training's accuracy. The model includes a data normalisation approach to address the issue of data source diversity. Blockchain-enabled smart contract-based management of edge training schedules and node authentication is retained in the model, and the problem of single points of failure and dishonest customers in traditional federation learning is addressed by performing multiplicative cryptographic execution and updating aggregated models via additive encryption and blockchain to secure the data. In addition, future tests will be conducted to evaluate the effectiveness of the model for healthcare data security improvement.

In summary, the objectives of this study have been met and, with some adaptation, can be widely applied to various medical research platforms and healthcare organisations in the future. This will help the medical research field to collect and analyse data in a more secure manner and actively explore sustainable research models and collaboration mechanisms.

##### References

1. [1] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A Blockchained Federated Learning Framework for Cognitive Computing in Industry 4.0 Networks," IEEE Transactions on Industrial Informatics, vol. 17, no. 4, pp. 2964-2973, 2021, doi: 10.1109/TII.2020.3007817.
2. [2] W. Xing and F. Yushun, "Decentralized federated LDA algorithm for user demand mining," Computer Integrated Manufacturing System, vol. 29, no. 4, p. 1055, 2023.
3. [3] D. C. Nguyen et al., "Federated learning meets blockchain in edge computing: Opportunities and challenges," IEEE Internet of Things Journal, vol. 8, no. 16, pp. 12806-12825, 2021.
4. [4] Z. Wang and Q. Hu, "Blockchain-based federated learning: A comprehensive survey," arXiv preprint arXiv:2110.02182, 2021.
5. [5] C. ZHANG, H. YANG, H. WANG, J. WANG, C. ZENG, and R. LI, "Efficient Model Collaborative Training and Sharing Scheme of Blockchain Based on Hybrid Privacy," 电子与信息学报, vol. 44, pp. 1-9, 2022.
6. [6] A. Prentosito, M. Skoczen, L. Kahrs, and S. Bhunia, "Case Study on a Session Hijacking Attack: The 2021 CVS Health Data Breach," in Mobile Web and Intelligent Information Systems, Cham, I. Awan, M. Younas, and A. Poniszewska-Marańda, Eds., 2022// 2022: Springer International Publishing, pp. 93-105.
7. [7] D. Hou, J. Zhang, K. L. Man, J. Ma, and Z. Peng, "A Systematic Literature Review of Blockchain-based Federated Learning: Architectures, Applications and Issues," in 2021 2nd Information Communication Technologies Conference (ICTC), 7-9 May 2021 2021, pp. 302-307, doi: 10.1109/ICTC51749.2021.9441499.
8. [8] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Low-Latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G Networks," IEEE Transactions on Industrial Informatics, vol. 17, no. 7, pp. 5098-5107, 2021, doi: 10.1109/TII.2020.3017668.
9. [9] M. A. Rahman, M. S. Hossain, M. S. Islam, N. A. Alrajeh, and G. Muhammad, "Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach," IEEE Access, vol. 8, pp. 205071-205087, 2020, doi: 10.1109/ACCESS.2020.3037474.
10. [10] H. Chai, S. Leng, Y. Chen, and K. Zhang, "A Hierarchical Blockchain-Enabled Federated Learning Algorithm for Knowledge Sharing in Internet of Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 3975-3986, 2021, doi: 10.1109/TITS.2020.3002712.
11. [11] Y. Wang, Z. Su, N. Zhang, and A. Benslimane, "Learning in the Air: Secure Federated Learning for UAV-Assisted Crowdsensing," IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1055-1069, 2021, doi: 10.1109/TNSE.2020.3014385.
12. [12] R. Kumar et al., "Blockchain-Federated-Learning and Deep Learning Models for COVID-19 Detection Using CT Imaging," IEEE Sensors Journal, vol. 21, no. 14, pp. 16301-16314, 2021, doi: 10.1109/JSEN.2021.3076767.
13. [13] R. Lei et al., "Federated Learning Over Coupled Graphs," IEEE Transactions on Parallel and Distributed Systems, vol. 34, no. 4, pp. 1159-1172, 2023.
14. [14] M. Alkahtani, Q. S. Khalid, M. Jalees, M. Omair, G. Hussain, and C. I. Pruncu, "E-Agricultural Supply Chain Management Coupled with Blockchain Effect and Cooperative Strategies," Sustainability, vol. 13, no. 2, p. 816, 2021. [Online]. Available: <https://www.mdpi.com/2071-1050/13/2/816>.
15. [15] M. Baza, A. Rasheed, A. Alourani, G. Srivastava, H. Alshahrani, and A. Alshehri, "Privacy-preserving Blockchain-assisted private-parking scheme with efficient matching," Computers and Electrical Engineering, vol. 103, p. 108340, 2022/10/01/ 2022, doi: <https://doi.org/10.1016/j.compeleceng.2022.108340>.
16. [16] J. Zhu, J. Cao, D. Saxena, S. Jiang, and H. Ferradi, "Blockchain-empowered Federated Learning: Challenges, Solutions, and Future Directions," ACM Comput. Surv., vol. 55, no. 11, p. Article 240, 2023, doi: 10.1145/3570953.
17. [17] G. S. Aujla and A. Jindal, "A Decoupled Blockchain Approach for Edge-Envisioned IoT-Based Healthcare Monitoring," IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 491-499, 2021, doi: 10.1109/JSAC.2020.3020655.