Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.052570

ARTICLE

Privacy-Preserving Healthcare and Medical Data Collaboration Service System Based on Blockchain and Federated Learning

Fang Hu[1], Siyi Qiu[2], Xiaolian Yang[1], Chaolei Wu[1], Miguel Baptista Nunes[3] and Hui Chen[4,*]

[1]College of Information Engineering, Hubei University of Chinese Medicine, Wuhan, 430065, China

[2]College of Information Engineering, Zhongnan University of Economics and Law, Wuhan, 430073, China

[3]School of Advanced Technologies, Xi'an Jiaotong-Liverpool University, Suzhou, 215400, China

[4]School of Information Management, Central China Normal University, Wuhan, 430079, China

*Corresponding Author: Hui Chen. Email: h.chen@ccnu.edu.cn

Received: 07 April 2024 Accepted: 11 July 2024 Published: 15 August 2024

ABSTRACT

As the volume of healthcare and medical data increases from diverse sources, real-world scenarios involving data sharing and collaboration have certain challenges, including the risk of privacy leakage, difficulty in data fusion, low reliability of data storage, low effectiveness of data sharing, etc. To guarantee the service quality of data collaboration, this paper presents a privacy-preserving Healthcare and Medical Data Collaboration Service System combining Blockchain with Federated Learning, termed FL-HMChain. This system is composed of three layers: Data extraction and storage, data management, and data application. Focusing on healthcare and medical data, a healthcare and medical blockchain is constructed to realize data storage, transfer, processing, and access with security, real-time, reliability, and integrity. An improved master node selection consensus mechanism is presented to detect and prevent dishonest behavior, ensuring the overall reliability and trustworthiness of the collaborative model training process. Furthermore, healthcare and medical data collaboration services in real-world scenarios have been discussed and developed. To further validate the performance of FL-HMChain, a Convolutional Neural Network-based Federated Learning (FL-CNN-HMChain) model is investigated for medical image identification. This model achieves better performance compared to the baseline Convolutional Neural Network (CNN), having an average improvement of 4.7% on Area Under Curve (AUC) and 7% on Accuracy (ACC), respectively. Furthermore, the probability of privacy leakage can be effectively reduced by the blockchain-based parameter transfer mechanism in federated learning between local and global models.

KEYWORDS

Blockchain technique; federated learning; healthcare and medical data; collaboration service; privacy preservation

1 Introduction

In recent decades, the construction of healthcare and medical service systems has drawn a huge amount of attention, from which a large amount of healthcare and medical data has been generated and accumulated. The related medical institutions have proposed some innovative service systems that focus on academic research, clinical care, patient management, etc., to satisfy the requirements

of healthcare and medical data integration and application [1]. Many medical institutions have introduced some advanced medical systems, frameworks, and platforms to improve management efficiency and alleviate the shortage and uneven distribution of healthcare and medical resources [2]. Most of studies emphasized effective data management services, however, the security in data sharing and collaboration cannot be guaranteed. With medical development, a massive amount of heterogeneous data has been accumulated from mobile devices, clinical systems, healthcare monitoring platforms, etc., which are multi-source and cross-regional. Moreover, as an important part of healthcare and medical data, individual information requires high confidentiality because of its sensitivity. Additionally, different real-world scenarios require effective data services to support various applications, such as clinical research, biomedical development, and neuroscience studies. Therefore, how to construct a privacy-preserving healthcare and medical system to fuse heterogeneous data and provide effective data services in different real-world scenarios needs further study [3].

The combination of various advanced techniques, such as artificial intelligence (AI), big data, digital twin, blockchain, federated learning, etc., can promote the development of the medical and healthcare industries [4]. Blockchain has emerged as a promising technique with integrated, distributed, immutable, and reliable characteristics [5]. It can be used to create privacy-preserving, reliable, and effective data collaboration service systems for diverse research, such as bitcoin, healthcare monitoring, and clinical data sharing [6]. Recent studies have been developed for healthcare and medical data sharing and services using blockchain, which can provide guaranteed data sharing among untrusted participants [7,8]. Federated learning (FL) provides a privacy-preserving approach. Data owners train local models with their private data and just send the trained parameters to the server aggregation for global model training, achieving effective and secure data collaboration without sharing raw data [9]. Therefore, FL is applied in various scenarios, such as intelligent medical systems, the internet of vehicles, electronic health record sharing, etc. [10].

This study focuses on a healthcare and medical data collaboration service system encompassing clinical, research, and individual behavior data. The clinical data is primarily from the electronic medical record (EMR), consisting of text, images, or other multimedia data extracted during diagnosis, therapy, examinations, etc. It is usually utilized for intelligent diagnosis and treatment research [11]. The research data refers to clinical trial data, gene sequences, medical data, etc., which are mostly from various medical research and development institutions and can be used for disease-genome association studies, drug target predictions, etc. [12]. Individual behavior data, mainly collected by mobile smart devices, is transmitted to healthcare management systems, such as electronic health records (EHRs), referring to daily activities, heart rate, sleep patterns, etc. It may be applied for daily activity monitoring and analysis, chronic disease monitoring and prediction, etc. [13].

This aforementioned healthcare and medical data is stored across different areas or in different systems and platforms, which cannot be effectively integrated and shared during data storage, transferring, processing, and accessing [14]. Furthermore, these kinds of data cannot be effectively fused because of data duplicates, non-standard data, and data availability in heterogeneous scenarios [15]. Moreover, healthcare and medical data contain a lot of private and sensitive data for individuals. Without effective data security strategies or privacy-preserving mechanisms, private and sensitive data may be leaked and have serious consequences. In this regard, protecting data privacy is an important issue for data collaboration services. Overall, it is significant to construct a data sharing system to fuse healthcare and medical data from various sources, present some secure and privacy-preserving strategies to guarantee effective and reliable data storage, transfer, processing, and access, and then provide high-quality data services in various scenarios. There are some frameworks or platforms that have been explored using blockchain or federated learning to realize data sharing and collaboration

and enhance data security and privacy preservation. However, the existing systems or platforms cannot satisfy the overall requirements of healthcare and medical data collaboration services because they don't have sophisticated data fusion, storage, management, and application strategies [16].

The main challenges in healthcare and medical data storage, transfer, processing, and access are summarized as follows:

- Many mobile devices, platforms, and systems offer healthcare and medical data services. However, there are still obstacles in effectively merging multi-source, heterogeneous, and cross-regional data. Additionally, security mechanisms for data storage and sharing, as well as efficient collaboration services in real-world scenarios haven't been fully developed [17,18].
- Establishing trust in a decentralized system where multiple parties are involved is challenging. Blockchain helps by providing an immutable ledger, but ensuring the reliability and trustworthiness of all participants remains critical [19].
- For most studies, raw healthcare and medical data is shared and collaborated on models or systems to meet the requirements of different users, which may increase the risk of privacy leakage and restrict the development of healthcare and medical data services [10].

To address these challenges, we investigated a privacy-preserving Healthcare and Medical Data Collaboration Service System Based on Blockchain and Federated Learning (FL-HMChain), including the data extraction and storage layer, the data management layer, and the data application layer. Moreover, a CNN-based Federated Learning model (FL-CNN-HMChain) is proposed for medical image identification based on FL-HMChain. The following are the contributions to this study:

- A privacy-preserving Healthcare and Medical Data Collaboration System (FL-HMChain) is presented, which combines blockchain technology with federated learning. FL-HMChain is designed to effectively integrate diverse healthcare and medical data, ensure secure and reliable data sharing, and deliver a range of user-oriented services across different scenarios.
- An improved consensus mechanism is presented to elect master nodes with high scores and distribute control and verification to master nodes across the network. The validation process detects and filters out malicious updates and the use of rotating masters prevents any single node from consistently influencing the validation process.
- A CNN-based Federated Learning model with Blockchain, termed FL-CNN-HMChain, is investigated to identify medical images based on FL-HMChain. To guarantee privacy, it realizes the local/server model training and model parameter updating mechanisms in the blockchain without directly sharing raw data. Furthermore, this model provides more accurate predictions than the baseline Convolutional Neural Network (CNN) and VGG16.

The remainder of this paper is organized as follows: The existing studies are presented in Section 2. In Section 3, the investigated FL-HMChain system with three layers is depicted. In Section 4, a CNN-based federated learning model in the healthcare and medical chain is shown. In Section 5, the experiments are elaborately described. Finally, we give the conclusion and perspective in Section 6.

2 Related Work

2.1 Blockchain and Federated Learning in Healthcare and Medicine

We discuss some existing methods, systems, frameworks, and platforms for electronic healthcare records (EHRs), electronic medical records (EMRs), prescription studies, research on genomics, clinical trials, etc., which are based on blockchain, federated learning, and both of them.

Chamola et al. proposed a framework based on blockchain to store medical records that can generate a history medical report using artificial intelligence (AI) [20]. Haque et al. presented a blockchain-based model that employs the Secure Hash Algorithm, SHA256, to generate a 32-byte or 256-bit hash value that is both unique and identical for specific medical data [21]. Chen et al. presented a secure inter-hospital EMR sharing system based on blockchain, which employs a programmatic authorization mechanism using smart contracts to guarantee EMR security [22]. Zhang et al. presented a redactable blockchain approach with hierarchical access control for data sharing, which realizes effective owners' data control and weakens modifiers' privileges [7]. Chelladurai et al. built a system that provides a high level of security and integrity through hash functions and utilizes blockchain as a clinical data repository to provide patients with a complete distributed ledger record [23]. Lakhan et al. proposed a deep reinforcement learning and blockchain-based Task Scheduling (DRLBTS) algorithm, which provides secure and efficient scheduling for healthcare applications [24]. Jiang et al. proposed a blockchain-based platform for healthcare information exchange (BlocHIE), satisfying privacy and authenticability [25].

Lee et al. implemented a federated learning (FL) model on a clinical benchmark data set and provided reliability and privacy protection for data processing [26]. Huang et al. proposed a FL model combined with an adaptive boosting method, termed LoAdaBoost, to predict the mortality of patients according to their prescriptions [27]. Xue et al. investigated a novel fully decentralized federated framework by aggregating the Double Deep Q-Network models to extract information from EMRs for supporting clinical treatment [28]. Huang et al. presented a community-based FL algorithm by clustering the distributed clinical data to predict mortality and length of stay [29]. Ding et al. proposed a privacy-preserving federal learning framework for detecting seizures based on EEG signals in a fog computing-based medical Internet of Things [30]. Guo et al. investigated a method for processing real-time medical data based on federated learning, aiming to train disease diagnosis models from continuous medical real-time data streams to achieve assisted diagnosis [31]. Salim et al. presented an electronic health record sharing scheme based on federated learning for medical informatics to protect patient data privacy [32].

The combination of blockchain and federated learning has a significant improvement in security and privacy issues [19]. Połap et al. proposed a multi-agent system to process real-time medical data by combining the blockchain and threaded FL, which effectively guarantees the security of private data [33]. Inspired by a cross-chain method, Jin et al. proposed a cross-cluster FL system (CFL) to realize privacy protection and secure data sharing for the Internet of Medical Things [34]. El Rifai et al. combined blockchain and FL to realize a coordinating server, which can effectively improve prediction results while ensuring usage consent and data transparency [35]. Lian et al. proposed a blockchain-based personalized Federated Learning (FL) system based on the diversity of patient conditions, which can participate in personalized model training and has privacy, improving the security level of the system [36].

2.2 Problem Statement

Despite the aforementioned blockchain and federated learning-based methods, systems, platforms, and frameworks that can realize the services of data storage, transfer, processing, access, etc., most of them are incapable of balancing the effectiveness and security of data applications. Some studies focused on effective data services but ignored security as the process of data fusing, storing, and sharing. Some security strategies have been investigated; however, the effectiveness of data sharing and collaboration cannot be guaranteed. For instance, systems that prioritize effectiveness may compromise security measures, leaving data vulnerable to unauthorized access, tampering, or

leakage. In contrast, overly strict security measures may hinder data sharing and collaboration, reducing the overall effectiveness of the system. Moreover, blockchain, federated learning, or the combination of blockchain and federated learning are utilized for data services and data security. Blockchain-empowered federated learning has been explored in depth while federated learning-empowered blockchain is understudied, leading the unbalance of their mutual benefits. Therefore, we focus on the improvement of blockchain through federated learning model training, presenting a system that achieves privacy-preserving healthcare and medical data collaboration services. By employing the node consensus mechanism of blockchain, it combines verified data from multiple sources, securely realizing complex data storage and transmission and offering a variety of medical application services that orient users.

3 FL-HMChain System

Fig. 1 illustrates the system of privacy-preserving Healthcare and Medical Data Collaboration Services Based on Blockchain and Federated Learning (FL-HMChain), including the following three layers.

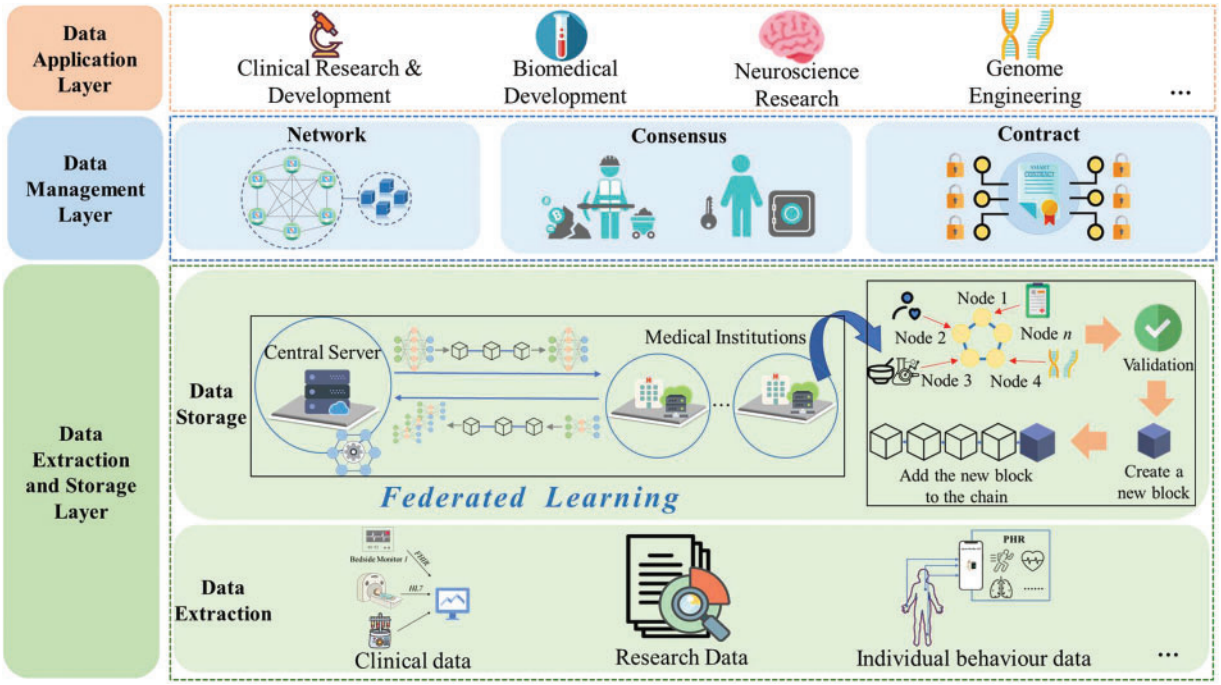

Figure 1: FL-HMChain system

3.1 Data Extraction and Storage Layer

The data extraction and storage layer provides effective and reliable data fusion and storage. The block generation and transaction mechanisms have been presented to store data securely and reliably. Federated learning (FL) can realize local and global model training and update the trained model's parameters on the blockchain without sharing raw data.

3.1.1 Data Extraction

Healthcare and medical data collected from different mobile equipment, platforms, and systems are with the features of multi-source, cross-regional, multi-relations, heterogeneity, etc. This system mainly extracts three classes of data, including clinical, research, and individual behavior data, from diverse healthcare and medical institutions as follows:

Clinical Data: This kind of data is mainly collected from the hospital information system (HIS), referring to different sub-systems, including doctor station, picture archiving and communication system (PACS), laboratory information system (LIS), microbial experiments, pharmaceutical administration system, pathological information system, etc. Clinical textual, image, and multimedia data referring to diagnosis, treatment, prescriptions, tests, examinations, etc., have been collected from EMR.

Research Data: This type of data refers to clinical trial data, gene sequence data, medical data, etc., mostly extracted from the systems or platforms of device research and development companies, outsourcing enterprises, genetic testing enterprises, or scientific research institutions.

Individual Behavior Data: This kind of data is real-time and collected from individual behaviors perceived by mobile equipment, such as mobile phones, smart watches, bracelets, etc. Mobile devices monitor an individual's real-time daily behaviors, such as sleep, step numbers, body temperature, heart rate, etc. Other related data is obtained from different online healthcare and medical systems or platforms, most of which are stored in the EHR.

3.1.2 Data Storage

In the data storage layer, two kinds of data are stored on the blockchain. One of them is clinical data, research data, and individual behavior data extracted from different institutions, mobile devices, systems, or platforms. Another data refers to the parameters generated from the global and local model training in a federated learning system for healthcare and medical data collaboration services. For the distribution, transparency, and tamper-resistance of diverse healthcare and medical data and model parameter storage, a blockchain has been created. Each block consists of an index, a time stamp, data, a current hash value, and a previous block hash value, and employs pointers to connect itself to other blocks. Such linked blocks form a blockchain, ensuring the tamper-resistance and integrity of healthcare and medical data and model parameters.

Blockchain uses a distributed ledger network. Firstly, a network node for healthcare and medical data requests a transaction to broadcast to peer nodes. After that, using a secure strategy like the Secure Hash Algorithm 256 (SHA-256) algorithm, this network generates a unique hash. Finally, all hashes are linked by the previous hash, which will create an unbreakable transaction network that can only be verified by a node or a smart contract. This immutable ledger can be appended to the transaction of blocks. With the generation of a decentralized network, some confirmed transactions, such as contracts, encryption, etc., will be verified. Algorithm 1 is block storage for extracted data and model parameters.

Algorithm 1: Storage of Blockchain

Input: Block()

1: index, previous_hash
2: data = healthcare and medical data, or parameters from model training
3: timestamp = time()

(Continued)

Algorithm 1 (continued)

```
4: hash = SHA-256() //hash value generation function
Input: AddBlock()
5: block = Block()
6: if block.previous_hash = previous_block.hash then
7:       chain.append(block)
8: end if
Input: Verification()
9: for i in range(1, len(chain)) do
10:    block = chain[i]
11:    previous_block = chain[i − 1]
12:    if block.previous_hash !=previous_block.hash then
13:          invalid block
14:    end if
15: end for
```

Data integrity is ensured by a chain growth and data verification system. One block is formed and then linked to the free end of the blockchain while fresh healthcare and medical data are uploaded to it. Once a block is altered, cryptographic connections in blocks will be destroyed so that the entire chain will be disrupted. Additionally, this blockchain enables users to check the accuracy and integrity of their data [37]. Furthermore, the parameter transmission mechanism in terms of the blockchain between the global model and local models guarantees privacy preservation without raw data transfer. Therefore, combining blockchain with federated learning can address privacy and security problems caused by the transmission and sharing of data to improve the reliability of the presented system.

3.2 Data Management Layer

This layer mainly refers to peer-to-peer (P2P) networks, consensus mechanisms, and smart contracts.

3.2.1 Peer-to-Peer Network

Blockchain technology is utilized to manage the healthcare and medical data and the parameters generated from the model training in FL that are appended to a distributed ledger across a P2P network. The blockchain nodes have to concur on transactions using a series of consensus mechanisms, in a P2P network manner. The ability of blockchain systems to spread information among connected peers depends on a P2P network. All network master peers must agree on knowledge of information consensus set such as blocks and transactions depending on certain functions in the consensus mechanisms and receive the required information in terms of gossip or flooding protocols, turning the blockchain network into an unstructured P2P network [38].

In the P2P network, healthcare participants, including doctors, pharmacists, patients, etc., naturally connect to create a network structure. Each node in this network stores valuable structural information on the blockchain. This empowers various healthcare and medical data users, such as medical institutions and individuals, to interact seamlessly, exchange data, and access public information through the P2P network. For example, physicians download the latest data from the blockchain onto their mobile devices and update the diagnosis and treatment information for patients on the blockchain. Healthcare workers collect medical data and distribute updated data, including test

results, notifications, policies, regulations, etc., from the blockchain. Individuals use mobile terminals to submit and receive data in a P2P network.

3.2.2 Consensus Mechanism

The healthcare and medical blockchain utilizes Practical Byzantine Fault Tolerance (PBFT) [39], which is more appropriate for the complicated structures of healthcare and medical data, to effectively reduce energy consumption. This improved consensus mechanism has five phases: 1) A group of master nodes is selected to be responsible for the validation of local gradients and block generation. The masters are dynamically elected at the beginning of each round based on the performance scores of nodes from the previous round. The remaining client nodes in the network perform local training on their respective data and generate local updates, sending $< REQUEST, o, t, c >$ to the master node and signs. The o indicates the specific operation requested. t is the timestamp, and c is the identifier of the client node. The *REQUEST* contains the message body and message summary $d(m)$. 2) After receiving the request, the master nodes validate the request by using their data as a validation set and assign it a sequence number n. Each master runs the learning model using the local updates and measures the validation accuracy on their dataset, and assigns the accuracy to each local client as the score s. Next, the master node broadcasts $<< PRE\ PREPARE, v, n, d >, m, s >$ to other common client nodes, and then signs and logs. v, d, and m are the view number, the client message digest, and the content of the message, respectively. 3) If the *ith* node receives the Pre-Prepare message, the scores from various master nodes are collected and combined. The median of these scores is calculated to determine the final score for each local update client. Then, the *ith* node sends signed $< PREPARE, v, n, d, i, s' >$ to master nodes and logs the Pre-Prepare and Prepare. This approach mitigates the influence of outliers and ensures a fair assessment. 4) The main nodes verify the correctness of Prepare and broadcast. If node i receives verified Prepare broadcast from all masters, it sends a signed $< COMMIT, v, n, d, i >$ message to master nodes and records the commit message in the log. 5) Finally, if master nodes receive commit messages and the score is more than a standard value, they accept the request operation o and return $< REPLY, s' >$ to the local update client, uploading the parameters that are from the client node to blockchain or downloading the updated parameters from blockchain. If the update client receives all identical reply messages, the request reaches the consensus of the whole network.

At the end of each round, a new master node group is elected based on the scores of nodes from the previous round. Nodes with higher scores are more likely to be elected, ensuring that the masters are composed of reliable and high-performing nodes.

3.2.3 Smart Contract

Smart contracts are utilized in a blockchain to build a secure technology infrastructure that allows for the consistency of healthcare and medical data, improving the quality of services, user coordination, and the effectiveness of data collaboration.

Smart contracts allow trusted transactions that are traceable and irreversible without a third party [40]. A smart contract is created among all users, including hospitals, healthcare providers, clinics, other medical institutions, patients, individuals who need healthcare, etc. This contract proliferates through P2P networks and is deposited on a blockchain. Then the smart contract executes automatically. The users can completely control the healthcare and medical information encrypted and stored in the blockchain using smart contracts. By setting access permissions through smart contracts, users can realize effective and secure point-to-point data sharing, avoiding data leakage

and tampering. Additionally, for the federated learning model training, the smart contract makes it returns the aggregated parameters from a server to local models in an automaticmanner.

3.3 Data Application Layer

The data application layer presents the healthcare and medical data services (shown in Fig. 2 and Table 1), consisting of electronic health record (EHR), electronic medical record (EMR), clinical research and development, biomedical development, neuroscience research, genome engineering, medical image identification, etc. This system is created as a consortium blockchain, where some pre-chosen nodes that are visible to all the participants can participate in the consensus process, and the other nodes that are allowed by the designated participants need permission to take part in it. To fully demonstrate the application services, we specifically concentrate on a representative case, the data collaboration services of clinical research and development.

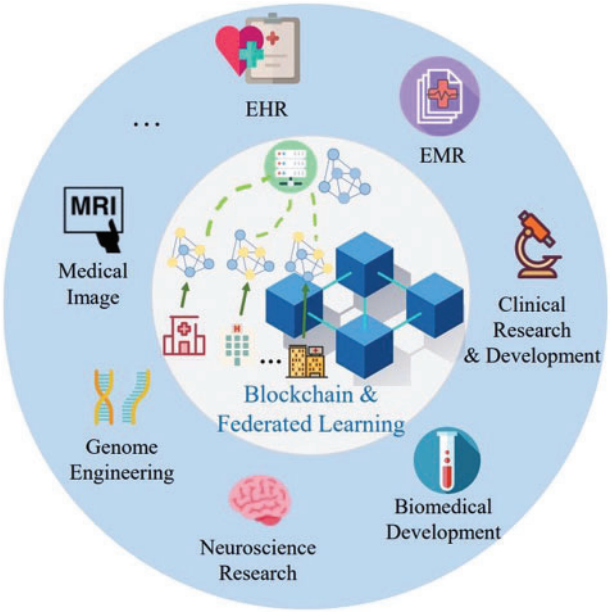

Figure 2: Illustration of data collaboration services in FL-HMChain

Table 1: Healthcare and medical applications based on FL-HMChain

Applications	Summary
Electronic Health Record (EHR)	EHR refers to the various personal healthcare information during the life cycle. This kind of service can authorize the users registered in FL-HMChain to access and obtain the EHR information at any time for personal health supervision or healthcare research, such as chronic disease management, health indicator monitoring, healthcare prediction, etc.

(Continued)

Table 1 (continued)

Applications	Summary
Electronic Medical Record (EMR)	EMR consists of various data from the clinical information system, such as diagnosis data, treatment data, examination data (images), testing data, etc. The users (such as physicians, researchers, etc.) registered in FL-HMChain can be authorized to acquire the data for diagnosis and treatment assistance and analysis, such as intelligent diagnosis, disease prediction, etc.
Clinical Research and Development	It mainly refers to the EMR data and the clinical activities covering various users, such as physicians, patients, pharmacists, technicians, radiologists, researchers, etc., who can obtain their customized services, such as clinical decision support for physicians, lifestyle advice for patients, data-driven predictive analysis for researchers, etc.
Biomedical Development	It is a significant engineering domain referring to biomedical research data, such as medical and healthcare records, transaction records, sensor information, consent forms, ambient temperature, etc. Biomedical researchers can be authorized by FL-HMChain to access integrated information related to their experiments.
Neuroscience Research	It is a complicated scientific domain to acquire and analyze human brain activity data. A lot of neuroscience data is generated by brain stimulation, brain study, brain thinking, brain re-enactment, etc. FL-HMChain authorizes neuroscientists to access and utilize the data for scientific research.
Genome Engineering	It concerns gene sequencing, including genotyping, genetic variants, genome analysis, interpretation, etc., which can produce a great number of personal genetic data to prevent and treat diseases in advance. FL-HMChain provides open-source access for advanced genomic analyses.
Medical Image	Through distributed and trusted blockchain technology, FL-HMChain can provide a more secure, reliable, and better shared medical image data access service. It can provide a more solid foundation for registered users, such as physicians, to accurately realize the diagnosis and treatment of various diseases.

To fully demonstrate the application services, we specifically concentrate on a representative case. Clinical data collaboration mainly focuses on research institutions and academia. According to Fig. 3, the clinical data, which mainly includes the diagnosis and treatments, prescriptions, examinations, test data, etc., are extracted from the EMR. This data is encrypted and stored in the FL-HMChain. Clinical research and development refer to the diagnosis and treatment related to different participants, such as the physician, pharmacist, radiologist, lab technician, patient, researcher, etc. Textual and image formats of clinical data are authorized to be transferred between the blockchain and participants. In this FL-HMChain system, some analysis models, such as disease prediction, medicine recommendation, etc., will be applied to assist in clinical diagnosis and treatment. For privacy preservation, federated learning is introduced to realize distributed and parallel model training without sharing the raw data. To further guarantee data security, the parameters generated by the local or global models can be stored in the blockchain. Through FL-HMChain-based data access, different participants can acquire their customized services, such as lifestyle suggestions for patients, data-driven predictive analysis for researchers, clinical decision support for physicians, etc.

Figure 3: Illustration of clinical data collaboration services

4 A CNN-Based Federated Learning Model in Healthcare and Medical BlockChain

To verify the effectiveness and reliability of FL-HMChain, we design a Convolutional Neural Network-based Federated Learning Model in the Healthcare and Medical Chain (FL-CNN-HMChain) (shown in Fig. 4), which consists of the following components.

4.1 CNN Model Constructing

In this federated learning framework, a Convolutional Neural Network (CNN) model with six layers is constructed for the local model and server aggregation training. In this Convolutional Neural Network (CNN) model, the first convolutional layer has a kernel size of 3×3. The second convolutional layer has a kernel size of 3×3, and it has a maximum pooling layer with a kernel size

of 2 × 2 and a stride of 2. The maximum pooling layer down samples the input feature map to reduce the spatial size of the feature map while retaining its most salient features. Both the third and fourth convolutional layers have a kernel size of 3 × 3. The fifth convolutional layer has a kernel size of 3 × 3, and it has a maximum pooling layer with a kernel size of 2 × 2 and a stride of 2. Each convolutional layer uses BatchNorm2D to speed up convergence and combat overfitting.

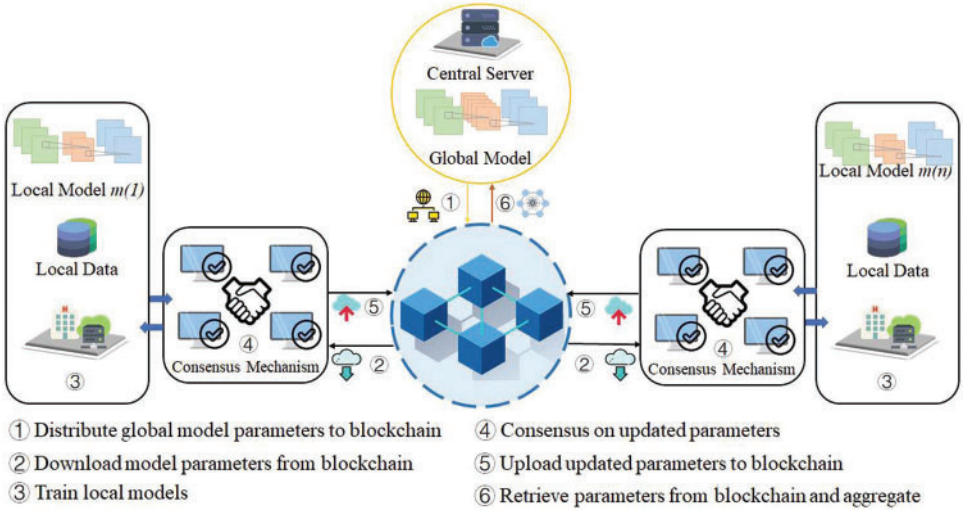

Figure 4: The framework of FL-CNN-HMChain

The final fully connected layer outputs the prediction results. The activation function in each layer is set to ReLU. Stochastic gradient descent (SGD) is taken as the optimizer, updating the model parameters iteratively to minimize the defined objective function. The loss function we used is the Cross Entropy, which is denoted as follows:

$$Loss = -\frac{1}{N}\sum_{i\in n}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{1}$$

where N represents the total sample, y_i represents the *ith* label, and p_i represents the *ith* predicted result.

4.2 Local Model Training (Client Update)

The clients send the request to master nodes to download the parameters from the blockchain to train their local Convolutional Neural Network (CNN) models after the central server dispatches the global Convolutional Neural Network (CNN) model's parameters (weights) to the blockchain. Firstly, the server randomly selects k clients from K clients in the training of the local model with a certain proportion P. The numbers of training rounds and epochs per round are defined as R and E, respectively. Each round is defined as $r \in R$ and each epoch is defined as $e \in E$. For further verification and aggregation, each client trains the local Convolutional Neural Network (CNN) model in parallel and then uploads its trained local model's weights w in each round to the blockchain based on the local medical images. w trained and updated on each local model is based on a federated averaging algorithm [41]. Client k calculates the local data gradient:

$$g_k = \nabla l(w_r; b) \tag{2}$$

where b denotes the batch size. ∇ is the computing gradient and $l()$ is the loss function. Then the local model will be updated.

$$w_{r+1}^{k} \leftarrow w_r - lrg_k \tag{3}$$

where lr is the learning rate in the local model.

4.3 Server Aggregation Training (Server Update)

After the clients upload the trained local model's parameters (weights) w_r to the blockchain, the server downloads them and uses the federated averaging algorithm to execute the global aggregation based on these parameters and train the global model. Then, the server updates the global model's weights w_{r+1} to the blockchain for the next round of training of all selected clients' local models. The server aggregation is based on a weighted average of the number of samples from each client:

$$w_{r+1}^{k} \leftarrow \sum_{k=1}^{k} \frac{n_k}{n} w_{r+1}^{k} \tag{4}$$

where k denotes the kth selected client, n_k and n refer to the quantity of samples on the kth client and on all the chosen clients, respectively. Finally, the loop will stop until all communication rounds are completed.

The pseudocode of the CNN-based federated learning model in the healthcare and medicine chain is as in Algorithm 2.

Algorithm 2: CNN-Based Federated Learning

Input: Build local CNN_k model for each clients. Define the maximum communication round R and the training epoch number E for each round. Divide image data into k clients. // each round $r \in R$, and each epoch $e \in E$.

Output: The global model's parameters

1: Initialize the global parameters w_1 to the blockchain
2: // Global model training
3: **for** $r = 1$ in range $(1, R)$ **do**
4: $k = K \times P$ // The server randomly select k clients with a certain proportion P
5: // Local model training
6: **for** each client k **do** // in parallel
7: Download global parameter w_r from the master nodes
8: **for** $e = 1$ in range $(1, E)$ **do**
9: Local parameter is w_{r+1}^{k} updated by Eq. (3) in CNN_k
10: **end for**
11: Upload the local parameter w_{r+1}^{k} to the master nodes
12: **end for**
13: Server downloads all local parameters w_{r+1}^{k} from the blockchain
14: Global parameters w_{r+1} are aggregated by Eq. (4)
15: Server uploads the global parameters w_{r+1} to the blockchain
16: **end for**

5 Experiments

5.1 Experiment Design

The experiments were run on a 24 GB RAM, 3.7 GHz Intel (R) Core (TM) CPU. Based on four data sets from MedMNIST v2 [42], we designed a series of experiments to verify the performance of the FL-CNN-HMChain model, including the model comparison and parameter sensitivity test. Two evaluation metrics, Area Under Curve (AUC) and Accuracy (ACC) are used to verify the prediction results of medical images.

5.2 Data Sets

We use the 2D image data sets from MedMNIST v2, including PneumoniaMNIST, BreastMNIST, TissueMNIST, and OrganAMNIST, to compare the FL-CNN-HMChain model with a general Convolutional Neural Network (CNN) model. MedMNIST v2 contains large-scale standardized biomedical images. All the 2D images are preprocessed into 28×28 pixels with known labels.

The PneumoniaMNIST is based on 5856 pediatric chest X-ray images, which are split into 4708 training data, 524 validation data, and 624 test data. It has pneumonia and normal classes. BreastMNIST focuses on breast ultrasound, which includes 780 samples with a split of 546 training data, 78 validation data, and 156 test data. It has malignant and benign (normal) classes. TissueMNIST contains 236, 386 human kidney cortex cell samples with a split of 165,466 training data, 23,640 validation data, and 47,280 test data. It has 8 classifications: collecting duct and connecting tubule, distal convoluted tubule, glomerular endothelial cells, interstitial endothelial cells, leukocytes, podocytes, proximal tubule segments, and thick ascending limb. OrganAMNIST refers to 58,850 samples of abdominal computed tomography (CT), including 34,581 training data, 6491 validation data, and 17,778 test data. It has 11 classifications: bladder, femur-left, femur-right, heart, kidney-left, kidney-right, liver, lung-left, lung-right, pancreas, and spleen.

5.3 Experimental Results and Analysis

We used the VGG16 as the cooperation, which is a convolutional neural network with 16 layers for various computer vision tasks, particularly image classification. The parameter set of FL-CNN-HMChain is shown in Table 2. Compared to the Convolutional Neural Network (CNN) model, the parameters referring to Epoch, Learning Rate, and Batch Size are set the same as in the FL-CNN-HMChain model. Table 3 and Fig. 5 show FL-CNN-HMChain performs the CNN-based model. FL-CNN-HMChain can get better evaluation values of 0.960, 0.875, 0.896, and 0.989 compared to the Convolutional Neural Network (CNN) with 0.920, 0.776, 0.861, and 0.976, and VGG16 with 0.541, 0.500, 0.495, and 0.469 on Area Under Curve (AUC) for PneumoniaMNIST, BreastMNIST, TissueMNIST, and OrganAMNIST. It also obtains better results of 0.875, 0.795, 0.620, and 0.868 compared to Convolutional Neural Network (CNN) with 0.837, 0.731, 0.504, and 0.808, and VGG16 with 0.625, 0.731, 0.321, and 0.185 on Accuracy (ACC). It has a greater performance compared with VGG16. Furthermore, the Area Under Curve (AUC) values of the FL-CNN-HMChain are improved by 4%, 9.9%, 3.5%, and 1.3% on PneumoniaMNIST, BreastMNIST, TissueMNIST, and OrganAMNIST than the Convolutional Neural Network (CNN) model, and the Accuracy (ACC) values of the FL-CNN-HMChain are improved by 3.8%, 6.4%, 11.6%, and 6%. The combination of federated learning with Convolutional Neural Network (CNN) and blockchain can not only effectively

improve the model performance but also achieve privacy preservation in healthcare and medical data collaboration.

Table 2: Parameters of FL-CNN-HMChain model

Parameters	Value
Communication rounds	50
Clients	100
Client fraction	10%
GCN layers	5
Activation functions	ReLU or Softmax
Learning_rate	$1e{-}2$
Epochs	3
Batch size	128
Optimizer	SGD
Loss function	Categorical_crossentropy

Table 3: Model comparison on AUC and ACC with different data sets

Model	Metric	PneumoniaMNIST	BreastMNIST	TissueMNIST	OrganAMNIST
VGG16	AUC	0.541	0.500	0.495	0.469
	ACC	0.625	0.731	0.321	0.185
CNN	AUC	0.920	0.776	0.861	0.976
	ACC	0.837	0.731	0.504	0.808
FL-CNN-HMChain	AUC	**0.960**	**0.875**	**0.896**	**0.989**
	ACC	**0.875**	**0.795**	**0.620**	**0.868**

To verify the FL-CNN-HMChain model's performance, we compare the presented model to a general Convolutional Neural Network (CNN) model on two evaluation metrics: Area Under Curve (AUC) and Accuracy (ACC). Additionally, we carry out parameter sensitivity tests by tuning the parameters of Learning Rate and Batch Size to investigate the impact of different FL-CNN-HMChain's parameters. The experimental results on Area Under Curve (AUC) and Accuracy (ACC) are referred to in Table 4. By setting 0.1, 0.01, 0.001, and 0.0001 to Learning Rate and verifying them on four data sets, we find that FL-CNN-HMChain performs best on Area Under Curve (AUC) with 0.960, 0.875, 0.896, and 0.989, and on Accuracy (ACC) with 0.875, 0.795, 0.620, and 0.868 when Learning Rate = 0.1. Similarly, we set 32, 64, 128, and 256 to Batch Size. When Batch Size = 128, FL-CNN-HMChain acquires the best results of 0.875 and 0.989 on BreastMNIST and OrganAMNIST for Area Under Curve (AUC) and obtains 0.875, 0.795, and 0.620 on PneumoniaMNIST, BreatMNIST, and TissueMNIST for Accuracy (ACC), respectively.

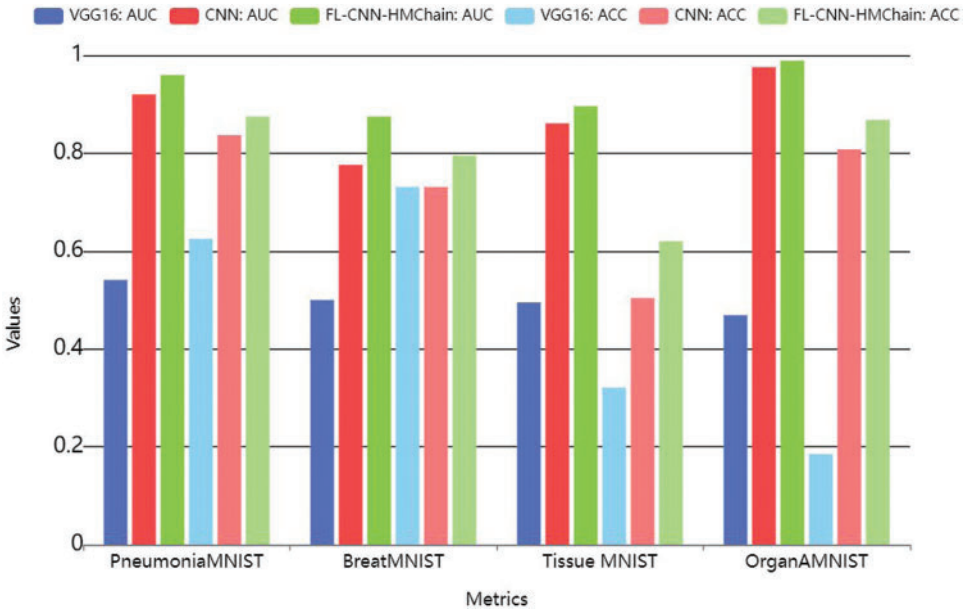

Figure 5: Model comparison on AUC and ACC with different data sets

Table 4: Parameter sensitivity test

Metric	Data set	Learning rate				Batch size			
		0.1	0.01	0.001	0.0001	32	64	128	256
AUC	PneumoniaMNIST	**0.960**	0.953	0.931	0.497	0.959	**0.965**	0.960	0.956
	BreastMNIST	**0.875**	0.825	0.737	0.594	0.820	0.820	**0 .875**	0.817
	TissueMNIST	**0.896**	0.890	0.867	0.773	0.875	0.878	0.896	**0.898**
	OrganAMNIST	**0.989**	0.986	0.881	0.706	0.987	0.989	**0.989**	0.989
ACC	PneumoniaMNIST	**0.875**	0.825	0.777	0.625	0.827	0.853	**0.875**	0.857
	BreastMNIST	**0.795**	0.788	0.731	0.731	0.769	0.769	**0.795**	0.744
	TissueMNIST	**0.620**	0.606	0.563	0.451	0.588	0.598	**0.620**	0.614
	OrganAMNIST	**0.868**	0.839	0.486	0.291	0.855	0.868	0.868	**0.872**

6 Conclusions

Based on blockchain and federated learning techniques, we presented the healthcare and medical service system (FL-HMChain) for multi-source, heterogeneous, and cross-regional data sharing and collaboration. The distributed data storage and management strategies in terms of blockchain are designed to ensure the security and reliability of data transmission, processing, and access. On the basis of the master node election mechanism consensus, it significantly reduces the risk of fake local training parameters being uploaded. This system can also provide effective and reliable applications, such as clinical research, biomedical development, neuroscience research, genome engineering, etc. Moreover,

a clinical data collaboration service has been presented in detail. Furthermore, to verify the effectiveness and reliability of the FL-HMChain, a CNN-based Federated Learning model with Blockchain, termed FL-CNN-HMChain, is proposed to realize medical image identification. The experimental results demonstrate that the presented model realizes the average 4.7% and 7% improvements in Area Under Curve (AUC) and Accuracy (ACC) compared to the baseline Convolutional Neural Network (CNN), respectively. In addition, this model realizes privacy preservation in healthcare and medical data collaboration services, providing support for sensitive information security, efficiency, and reliability in data sharing and analysis. Its implementation can significantly advance the field of healthcare informatics and contribute to improving patient care and medical research. In future research, we will focus on more reliable and accurate prediction models for healthcare and medical data, besides medical images, further optimizing their performance to ensure reliability and accuracy in medical data analysis.

Acknowledgement: The authors express sincere thanks to the financial support provided by the funding support from the Science and Technology Projects of the National Archives Administration of China and the Fundamental Research Funds for the Central Universities, Central China Normal University.

Funding Statement: We are thankful for the funding support from the Science and Technology Projects of the National Archives Administration of China (Grant Number 2022-R-031), and the Fundamental Research Funds for the Central Universities, Central China Normal University (Grant Number CCNU24CG014).

Author Contributions: All the authors made a great contribution to this work: Fang Hu: Methodology, formal analysis, writing—original draft. Siyi Qiu: Methodology, modeling, writing. Xiaolian Yang: Methodology, modeling, writing. Chaolei Wu: Methodology, modeling, validation. Miguel Baptista Nunes: Methodology, conceptualization. Hui Chen: Methodology, supervision, writing—original draft, review & editing. All authors reviewed the results and approved the final version of the manuscript.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, "Blockchain-based medical records secure storage and medical service framework," *J. Med. Syst.*, vol. 43, no. 1, pp. 1–9, 2019. doi: 10.1007/s10916-018-1121-4.

[2] P. P. Ray, D. Dash, K. Salah, and N. Kumar, "Blockchain for IoT-based healthcare: Background, consensus, platforms, and use cases," *IEEE Syst. J.*, vol. 15, no. 1, pp. 85–94, 2020. doi: 10.1109/JSYST.2020.2963840.

[3] W. N. Price and I. G. Cohen, "Privacy in the age of medical big data," *Nature Med.*, vol. 25, no. 1, pp. 37–43, 2019. doi: 10.1038/s41591-018-0272-7.

[4] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, "Big data service architecture: A survey," *J. Internet Technol.*, vol. 21, no. 2, pp. 393–405, 2020.

[5] W. Viriyasitavat, L. Da Xu, Z. Bi, and A. Sapsomboon, "Blockchain-based business process management (BPM) framework for service composition in Industry 4. 0," *J. Intell. Manuf.*, vol. 31, no. 7, pp. 1737–1748, 2020. doi: 10.1007/s10845-018-1422-y.

[6] X. Guo, G. Liang, J. Liu, and X. Chen, "Blockchain-based cognitive computing model for data security on a cloud platform," *Comput. Mater. Contin.*, vol. 77, no. 3, pp. 3305–3323, 2023. doi: 10.32604/cmc.2023.044529.
[7] T. Zhang, L. Zhang, Q. Wu, Y. Mu, and F. Rezaeibagha, "Redactable blockchain-enabled hierarchical access control framework for data sharing in electronic medical records," *IEEE Syst. J.*, vol. 17, no. 12, pp. 1962–1973, 2022. doi: 10.1109/JSYST.2022.3186145.
[8] P. Kumar, R. Kumar, S. Garg, K. Kaur, Y. Zhang and M. Guizani, "A secure data dissemination scheme for IoT-based e-health systems using AI and blockchain," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*, Rio de Janeiro, Brazil, IEEE, 2022, pp. 1397–1403.
[9] J. Peng, J. Guo, F. Bao, C. Yang, Y. Xu and Y. Qin, "Multi-robot privacy-preserving algorithms based on federated learning: A review," *Comput. Mater. Contin.*, vol. 77, no. 3, pp. 2971–2994, 2023. doi: 10.32604/cmc.2023.041897.
[10] R. Wang and W. -T. Tsai, "Asynchronous federated learning system based on permissioned blockchains," *Sensors*, vol. 22, no. 4, pp. 1672, 2022. doi: 10.3390/s22041672.
[11] L. Ismail, H. Materwala, and S. Zeadally, "Lightweight blockchain for healthcare," *IEEE Access*, vol. 7, pp. 149935–149951, 2019. doi: 10.1109/ACCESS.2019.2947613.
[12] D. Azzolina, P. Berchialla, D. Gregori, and I. Baldi, "Prior elicitation for use in clinical trial design and analysis: A literature review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 4, pp. 1833, 2021. doi: 10.3390/ijerph18041833.
[13] A. Dubovitskaya *et al.*, "Action-EHR: Patient-centric blockchain-based electronic health record data management for cancer care," *J. Med. Internet Res.*, vol. 22, no. 8, pp. e13598, 2020. doi: 10.2196/13598.
[14] D. J. Skiba, "The potential of blockchain in education and health care," *Nurs. Educ. Perspect.*, vol. 38, no. 4, pp. 220–221, 2017. doi: 10.1097/01.NEP.0000000000000190.
[15] A. Shahnaz, U. Qamar, and A. Khalid, "Using blockchain for electronic health records," *IEEE Access*, vol. 7, pp. 147782–147795, 2019. doi: 10.1109/ACCESS.2019.2946373.
[16] L. Hang, E. Choi, and D. -H. Kim, "A novel EMR integrity management based on a medical blockchain platform in hospital," *Electronics*, vol. 8, no. 4, pp. 467, 2019. doi: 10.3390/electronics8040467.
[17] U. Bodkhe *et al.*, "Blockchain for Industry 4.0: A comprehensive review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020. doi: 10.1109/ACCESS.2020.2988579.
[18] S. Tanwar, K. Parekh, and R. Evans, "Blockchain-based electronic healthcare record system for Health-Care 4.0 applications," *J. Inf. Secur. Appl.*, vol. 50, pp. 102407, 2020.
[19] 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, pp. 1–31, 2023. doi: 10.1145/3570953.
[20] V. Chamola, A. Goyal, P. Sharma, V. Hassija, H. T. T. Binh and V. Saxena, "Artificial intelligence- assisted blockchain-based framework for smart and secure EMR management," *Neural Comput. Appl.*,vol. 35, no. 31, pp. 22959-22969, 2023. doi: 10.1007/s00521-022-07087-7.
[21] R. Haque *et al.*, "Blockchain-based information security of electronic medical records (EMR) in a healthcare communication system," in *Intell. Comput. Innov. Data Sci.*, Springer, 2020, pp. 641–650.
[22] C. -L. Chen, Y. -Y. Deng, W. Weng, H. Sun, and M. Zhou, "A blockchain-based secure inter-hospital emr sharing system," *Appl. Sci.*, vol. 10, no. 14, pp. 4958, 2020. doi: 10.3390/app10144958.
[23] U. Chelladurai and S. Pandian, "A novel blockchain based electronic health record automation system for healthcare," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 1, pp. 693–703, 2022. doi: 10.1007/s12652-021-03163-3.
[24] A. Lakhan, M. A. Mohammed, J. Nedoma, R. Martinek, P. Tiwari and N. Kumar, "DRLBTS: Deep reinforcement learning-aware blockchain-based healthcare system," *Sci. Rep.*, vol. 13, no. 1, pp. 4124, 2023. doi: 10.1038/s41598-023-29170-2.
[25] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma and J. He, "BlocHIE: A Blockchain-based platform for healthcare information exchange," in *2018 IEEE Int. Conf. Smart Comput.*, Taormina, Sicily, Italy, IEEE, 2018, pp. 49–56.

[26] G. H. Lee and S. -Y. Shin, "Federated learning on clinical benchmark data: Performance assessment," *J. Med. Internet Res.*, vol. 22, no. 10, pp. e20891, 2020. doi: 10.2196/20891.
[27] L. Huang, Y. Yin, Z. Fu, S. Zhang, H. Deng and D. Liu, "LoAdaBoost: Loss-based adaboost federated machine learning with reduced computational complexity on IID and non-iid intensive care data," *Plos one*, vol. 15, no. 4, pp. e0230706, 2020. doi: 10.1371/journal.pone.0230706.
[28] Z. Xue *et al.*, "A resource-constrained and privacy- preserving edge-computing-enabled clinical decision system: A federated reinforcement learning approach," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9122–9138, 2021. doi: 10.1109/JIOT.2021.3057653.
[29] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng and D. Liu, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," *J. Biomed. Inform.*, vol. 99, no. 9, pp. 103291, 2019. doi: 10.1016/j.jbi.2019.103291.
[30] W. Ding, M. Abdel-Basset, H. Hawash, S. Abdel-Razek, and C. Liu, "Fed-ESD: Federated learning for efficient epileptic seizure detection in the fog-assisted internet of medical things," *Inf. Sci.*, vol. 630, no. 3, pp. 403–419, 2023. doi: 10.1016/j.ins.2023.02.052.
[31] K. Guo, T. Chen, S. Ren, N. Li, M. Hu and J. Kang, "Federated learning empowered real-time medical data processing method for smart healthcare," *IEEE/ACM Trans. Computat. Biol. Bioinform.*, 2022.
[32] M. M. Salim and J. H. Park, "Federated learning-based secure electronic health record sharing scheme in medical informatics," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 617–624, 2022.
[33] D. Połap, G. Srivastava, and K. Yu, "Agent architecture of an intelligent medical system based on federated learning and blockchain technology," *J. Inf. Secur. Appl.*, vol. 58, no. 11, pp. 102748, 2021. doi: 10.1016/j.jisa.2021.102748.
[34] H. Jin, X. Dai, J. Xiao, B. Li, H. Li and Y. Zhang, "Cross-cluster federated learning and blockchain for internet of medical things," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15776–15784, 2021. doi: 10.1109/JIOT.2021.3081578.
[35] O. El Rifai, M. Biotteau, X. de Boissezon, I. Megdiche, F. Ravat and O. Teste, "Blockchain-based federated learning in medicine," in *Int. Conf. Artif. Intell. Med.*, Minneapolis, MN, USA, Springer, 2020, pp. 214–224.
[36] Z. Lian, W. Wang, Z. Han, and C. Su, "Blockchain-based personalized federated learning for internet of medical things," *IEEE Trans. Sustain. Comput.*, vol. 8, no. 4, pp. 694–702, 2023. doi: 10.1109/TSUSC.2023.3279111.
[37] V. Mothukuri, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and K. -K. R. Choo, "FabricFL: Blockchain-in-the-loop federated learning for trusted decentralized systems," *IEEE Syst. J.*, vol. 16, no. 3, pp. 3711–3722, 2021. doi: 10.1109/JSYST.2021.3124513.
[38] T. Neudecker and H. Hartenstein, "Network layer aspects of permissionless blockchains," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 1, pp. 838–857, 2018. doi: 10.1109/COMST.2018.2852480.
[39] S. Guo, Y. Qi, Y. Jin, W. Li, X. Qiu and L. Meng, "Endogenous trusted DRL-based service function chain orchestration for IoT," *IEEE Trans. Comput.*, vol. 71, no. 2, pp. 397–406, 2021. doi: 10.1109/TC.2021.3051681.
[40] L. Wu, X. Du, W. Wang, and B. Lin, "An out-of-band authentication scheme for Internet of Things using blockchain technology," in *2018 Int. Conf. Comput., Netw. Commun. (ICNC)*, Maui, Hawaii, USA, IEEE, 2018, pp. 769–773.
[41] B. McMahan, E. Moore, D. Ramage, S. Hampson, and A. B. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Stat.*, PMLR, 2017, pp. 1273–1282.
[42] J. Yang *et al.*, "MedMNIST v2–A large-scale lightweight benchmark for 2D and 3D biomedical image classification," *Sci. Data*, vol. 10, no. 1, pp. 41, 2023. doi: 10.1038/s41597-022-01721-8.