Review

A Review of Blockchain-Based Secure Sharing of Healthcare Data

Peng Xi, Xinglong Zhang, Lian Wang, Wenjuan Liu and Shaoliang Peng *

The College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
* Correspondence: slpeng@hnu.edu.cn

Abstract: Medical data contains multiple records of patient data that are important for subsequent treatment and future research. However, it needs to be stored and shared securely to protect the privacy of the data. Blockchain is widely used in the management of healthcare data because of its decentralized and tamper-proof features. In order to study the development of blockchain in healthcare, this paper evaluates it from various perspectives. We analyze blockchain-based approaches from different application scenarios. These are blockchain-based electronic medical record sharing, blockchain and the Internet of Medical Things and blockchain-based federal learning. The results show that blockchain and smart contracts have a natural advantage in the field of medical data since they are tamper-proof and traceable. Finally, the challenges and future directions of blockchain in healthcare are discussed, which can help drive the field forward.

Keywords: blockchain; healthcare data; privacy; security

1. Introduction

Healthcare data is relevant to everyone. It records physical information about our bodies. It is important for the diagnosis and treatment of diseases [1]. With the rapid development of artificial intelligence, medical data has become a great asset. It can help us build artificial intelligence diagnostic models and assist doctors in diagnosis. Although the recording of medical information has evolved from the initial paper records to electronic medical records (EMR), which are more convenient for data access and storage, more attention needs to be paid to protecting the privacy of data [2]. Many hospitals and institutions have reduced data transfer and sharing in order to avoid data privacy leakage, which has led to the formation of data silos as medical data is scattered among various medical institutions [3].

Health care data privacy and security also lead to other problems. For example, for security, patients need to be re-examined every time they go to a new hospital. This behavior wastes energy and money. In order to protect patient privacy, medical data cannot be shared with scientific institutions, which prevents medical development. These have prompted the search for secure data storage and transmission methods, and blockchain is widely used, because of its decentralized, tamper-proof nature, for sharing medical data [4]. Innoplexus combines artificial intelligence and blockchain to enable continuous scanning of global life science data [5]. The system provides data to research institutions and pharmaceutical companies. BlockRx is a platform that has been successfully used in real-world applications [6]. The platform combines blockchain technology and iSolve’s advanced digital ledger technology. The platform integrates medical data from biomedical and research institutions. BlockRx has been put into practical application and has achieved great development.

There have been several papers that summarize blockchain-based models. Jin et al. analyze the privacy sharing of medical data through the type of blockchain used in the model [7]. The review divides blockchains into two categories: permissionless and permissioned. Then analyzes the advantages and disadvantages based on the blockchain types.
Leili et al. have analyzed a number of papers published between 2016 and 2020 [8]. This article focuses on healthcare application situations and does not focus primarily on the comparison and summary of models. Some blockchain-based healthcare approaches are summarized by Saha et al., but they do not compare these approaches [9]. Israa et al. conducted a model analysis with a unique perspective, looking at both the benefits and threats that technology poses to patients [10]. Hasselgren et al. performed a statistical analysis of the published papers. However, this review did not summarize the techniques [11]. Xu et al. mainly analyze the application of blockchain in oncology medical data, such as drug traceability and oncology data sharing [12]. This study has limitations because only oncology data are analyzed. In this paper, we analyze blockchain-based methods for sharing healthcare data, dividing the technologies into three cases in terms of application scenarios: blockchain-based healthcare data storage and access, blockchain and internet of medical things(IOMT) and blockchain-based federal learning. Each technology is also compared and summarized, and finally, compared with the traditional cryptography-based data sharing methods.

This paper surveys the latest blockchain-based technologies for sharing and storing medical data and summarizes these technologies. The subsequent structure of this paper is organized as follows: Section 1 introduces the research background and the significance of this paper. Section 2 introduces the basic concept of blockchain. Section 3 introduces the scenarios of blockchain application in medical data sharing, categorizes them into three cases and introduces them in detail. Section 4 analyzes the challenges and opportunities encountered in blockchain in medical data. The fifth section concludes the paper.

2. Blockchain
2.1. Background of Blockchain

Since Bitcoin was proposed by Satoshi Nakamoto, blockchain has been widely noticed. Bitcoin is still one of the representatives of cryptocurrency [13]. The chain structure, Merkle tree and hash algorithm together guarantee the blockchain’s tamper-evident nature [14]. All nodes maintain the same ledger together, ensuring decentralization. Because of asymmetric encryption and authorization technology, the transaction information stored on the blockchain is public, and the account identification information is highly encrypted. Access is only possible with the authorization of the data owner, which ensures data security and personal privacy. The structure of the blockchain is shown in Figure 1.

![Blockchain structure](image)

**Figure 1.** Blockchain structure.

2.2. Classification of Blockchain

Blockchain can be divided into the public chain, private chain and consortium chain according to the way of participation [15]. A public chain is, as the name implies, completely
public and accessible to anyone. Because the data on the chain cannot be altered, public
chains are considered to be fully decentralized. The consortium chain is only limited to
authorized members to participate, and the read and write permissions and participation
accounting permissions on the blockchain are formulated according to the rules of the
alliance. The private chain is only used in private organizations, and the read and write
permissions on the blockchain and the permissions to participate in bookkeeping are
formulated according to the rules of the private organization. Participating nodes are few
and severely restricted [16]. Table 1 compares different types of blockchains.

<table>
<thead>
<tr>
<th></th>
<th>Decentralization</th>
<th>Throughput</th>
<th>Cost</th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public chain</td>
<td>high</td>
<td>low</td>
<td>high</td>
<td>poor</td>
</tr>
<tr>
<td>Consortium chain</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>great</td>
</tr>
<tr>
<td>Private chain</td>
<td>low</td>
<td>high</td>
<td>low</td>
<td>great</td>
</tr>
<tr>
<td>Hybrid chain</td>
<td>-</td>
<td>-</td>
<td>low</td>
<td>great</td>
</tr>
</tbody>
</table>

2.3. Consensus Algorithm

As a decentralized peer-to-peer system, the nodes receive the transactions in a different
order [17]. Therefore, consensus algorithms are needed to ensure that the nodes agree
on the transactions. Proof of work (POW) is the first successful decentralized blockchain
consensus algorithm. Practical Byzantine Fault Tolerance (PBFT) was proposed to solve the
Byzantine problem [18]. It makes sure that the blockchain can still function properly with
some faulty or malicious nodes.

2.4. Smart Contracts

Smart contracts are computer protocols that disseminate, validate or enforce contracts
in an informational manner [19]. Smart contracts do not require third-party authentication
and successful transactions are traceable and irreversible. A computer program is used
to write a legally valid contract, and this contract can be executed automatically. A smart
contract is a code deployed on the blockchain that ensures that transactions are safe and
secure without third-party oversight [20]. The smart contract process is shown in Figure 2.

1. Decentralized: The execution of smart contracts does not need to rely on the participa-
tion or intervention of third-party organizations, and the supervision and arbitration
of contracts are performed by computers;

2. Untamperable: Once a smart contract is deployed, all contents cannot be modified.
This is somewhat like a contract in the traditional world, which cannot be modified
once it is signed;

3. Low cost: Since smart contracts do not require supervision by a third-party intermedi-
ary, once a breach of contract occurs, the code is enforced and has a much lower cost
compared to traditional contracts;

4. Open and transparent: Once deployed successfully, a smart contract will run ac-
cording to the design code and can be viewed by anyone, with a high degree of
transparency [21].
3. Blockchain in Healthcare Data

There are many situations in which blockchain is used in medical data sharing, and now there are three types according to the application scenarios. The first is blockchain-based data security storage and access. The second is the use of blockchain combined with IOMT. The third is the use of blockchain to replace the central institution of federal learning.

3.1. Blockchain-Based Data Security Storage and Access

The emergence of EMR has brought convenience as well as privacy issues. Subject to security issues, medical data cannot be shared freely. Some blockchain-based models have been proposed [22].

A blockchain-based ‘medichain’ model was proposed by Rahul et al. [23]. This model uses the blockchain as a database to store the complete case information of the patient in the block. The transaction records are hashed to store the obtained hash values in the Merkle tree to ensure the security of the data and prevent tampering, thus reducing errors in clinical decision-making. To address the problem of a wide range of sources and diverse structures of medical data, the data of all fields are combined into a single hyperfield stored in the proposed framework. The method uses on-chain storage. However, the blockchain is less scalable. On-chain storage is also expensive. Wu introduces a patient-oriented privacy-preserving access control model into the process of access control of private information in healthcare systems [24]. Then, blockchain technology is used to build a private information storage platform, and standard cryptographic algorithms are used to realize information transmission. In this process, the privacy information is also secured by a file authorization contract to further prevent the theft of medical privacy information. The model proposes a fine-grained privacy-preserving access control method that grants different privileges to users by judging their types. EMR information is stored in the cloud database and hosted by a third-party cloud service organization. When data are stored on the cloud, a hash of that data is generated. Then the hash is stored on the blockchain. When the data in the cloud is tampered with, it can be compared by the hash value on the chain. In this model, the consensus algorithm is POW, which requires a lot of invalid computations by the nodes. Liu et al. propose a lightweight blockchain-based model for sharing and protecting medical data [25]. The model uses proxy re-encryption technology to enable data sharing among physicians in different hospitals. The hash function used is hard to collide. Therefore, stored medical information is almost impossible to be tampered with. The traditional delegated proof of interest is improved to obtain a new consensus algorithm that is more secure and reliable. A disease-matching mechanism is designed where patients suffering from the same disease can communicate with each other. After mutual authentication, session keys can be set between patients. This mechanism can help patients to exchange disease information. The private chain is fast in transactions but less decentralized. It is more suitable for applications within companies or institutions. It is not applicable when there are many patients and hospitals. A hybrid chain-based EHR sharing scheme
is proposed by Yu et al. to store the private part of the electronic case in the federated chain and the non-private part in the public chain [26]. Only licensed users can access the private part, and the non-private part can be shared with scientific institutions for medical development. The model also uses off-chain storage, and only data hashes are stored on the chain to prevent data tampering, and smart contracts can automatically manage the EMR request, approval and usage process. The hybrid chain approach applied by the model is very novel. However, no attributes are granted to the nodes, and coarse-grained access control is used. Zou et al. have designed a new chain structure to avoid the forking problem and proposed a trust-based consensus mechanism to resist Byzantine attacks [27]. Medical institutions can accumulate trust points through continuous mining in exchange for EMR. The proposed reputation system needs to accumulate reputation scores through a large number of invalid calculations. A large amount of energy is consumed in order to obtain voting rights. Shahnaz et al. propose a blockchain-based fine-grained access system that grants different access rights to patients, doctors, nurses and administrators [28]. The access to electronic cases is recorded in the model proposed in [29], and a searchable encryption method is used for the data on the chain to search for information without decryption. This method protects the privacy of the data and ensures the speed of the query. The method also uses role-based access control management. The comparison of different models is shown in Table 2. As can be seen through the table, almost all of these models use off-chain storage. This is due to the small capacity of the blockchain, which limits the capacity of data storage. This is an issue that needs to be addressed in the future. In addition, fine-grained access control is also an advantage of blockchain-based approaches.

Table 2. Comparison of methods.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Blockchain Types</th>
<th>Storage Methods</th>
<th>Data Encryption</th>
<th>Access Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>[23]</td>
<td>public</td>
<td>On-chain storage</td>
<td>no</td>
<td>coarse grained</td>
</tr>
<tr>
<td>[24]</td>
<td>public</td>
<td>Off-chain storage</td>
<td>yes</td>
<td>fine grained</td>
</tr>
<tr>
<td>[25]</td>
<td>private</td>
<td>On-chain storage</td>
<td>yes</td>
<td>coarse grained</td>
</tr>
<tr>
<td>[26]</td>
<td>hybrid</td>
<td>Off-chain storage</td>
<td>no</td>
<td>coarse grained</td>
</tr>
<tr>
<td>[27]</td>
<td>public</td>
<td>Off-chain storage</td>
<td>yes</td>
<td>coarse grained</td>
</tr>
<tr>
<td>[28]</td>
<td>public</td>
<td>Off-chain storage</td>
<td>no</td>
<td>fine grained</td>
</tr>
<tr>
<td>[29]</td>
<td>public</td>
<td>Off-chain storage</td>
<td>yes</td>
<td>fine grained</td>
</tr>
</tbody>
</table>

3.2. Blockchain with IOMT

The IOMT includes various medical devices that use computer networks to connect and detect parameters of patients’ signs. IOMT brings great advantages for patients’ disease management, and the detection of physical signs can detect diseases as soon as possible to seek medical treatment [30]. However, there are many IOMT products in the market without unified management standards, and it is prone to information leakage [31]. Blockchain brings a solution for the security of medical IOMT [32]. Figure 3 depicts the blockchain and IOMT.
Chen et al. developed an IOMT-based data acquisition system in order to achieve secure storage and sharing of medical data [33]. The system can collect data from multiple medical devices simultaneously to achieve real-time collection of patient health records during surgery. The system is designed as a cloud server-based anonymous medical data sharing scheme with a proxy re-encryption algorithm. This approach improves the security of private medical data sharing. The system is implemented based on Hyperledger Fabric, a permissioned blockchain architecture, with a two-channel structure deployment architecture and medical chain code designed for data management and access control. The kafka consensus algorithm is used in this method. This consensus algorithm can tolerate the failure of half of the nodes, but it cannot tolerate malicious nodes. This makes the system more vulnerable to attacks. A novel blockchain-based secure authentication technology was proposed by Jafar to enhance the security of sensitive medical data transmitted between patients and hospitals [34]. Lamport Merkle Digital Signature (LMDS)
performs signature generation and verification here to enable secure transmission of sensitive medical data in cloud server-based medical IoT networks. Smart contracts enable transacting parties (i.e., patients and physicians) to set conditions and automate operations through the cloud server, reducing the work of third parties. Smart contracts also have different addresses and accounts on the blockchain so that each IoT device can view and execute its instructions, thus reducing communication overhead. Alqaralleh et al. present a novel deep learning and blockchain-assisted secure image transmission and diagnosis model for IoMT [35]. The proposed model includes several processes, in particular, data collection, secure transactions, hash value encryption and data classification. In the primary stage, patient details are collected using IoT tools and then encrypted using the GO-FFO algorithm. In addition, the hashes in the blockchain are encrypted and compressed by NIS-BWT technology. Finally, the classification process is performed using the DBN model. The improved encryption algorithm, although more secure, takes longer to encrypt and decrypt than other algorithms. An API interface is provided in the system proposed by Suyel [36]. This interface generates and maintains health data between the healthcare provider and the patient. In addition, smart contracts are fully used in the proposed system to prevent malicious behavior by setting secure rules through smart contracts. The method uses only simple authentication. If fine-grained properties can be given to the nodes. This could make the model more perfect. Hu et al. propose that many blockchain-based IOMT studies now focus on cryptographic algorithm verification [37]. More efforts should be focused on invalid signatures at times to reduce the chance of verification failure. The comparison of blockchain-based IOMT models is shown in Table 3. Blockchain smart contracts play an important role in IOMT. Smart contracts do not require third-party participation and can automatically perform set tasks when conditions are met. Usually, the model uses cryptographic algorithms to enhance security.

Table 3. Comparison of blockchain-based IOMT models.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Blockchain Types</th>
<th>Data Encryption</th>
<th>Smart Contract</th>
<th>Key Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>[33]</td>
<td>fabric</td>
<td>yes</td>
<td>no</td>
<td>Re-encryption, anonymous sharing</td>
</tr>
<tr>
<td>[34]</td>
<td>public</td>
<td>yes</td>
<td>yes</td>
<td>Lamport merkle digital signature</td>
</tr>
<tr>
<td>[35]</td>
<td>public</td>
<td>yes</td>
<td>yes</td>
<td>Encryption after data classification</td>
</tr>
<tr>
<td>[36]</td>
<td>public</td>
<td>yes</td>
<td>yes</td>
<td>unique data certificate storage efficient digital verification mechanism</td>
</tr>
<tr>
<td>[37]</td>
<td>fabric</td>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

3.3. Blockchain with Medical Federal Learning

The privacy of healthcare data prevents data-driven machine learning from working [38]. Federated learning is a new AI technique that protects data privacy when building AI models. Federated learning enables multiple nodes to learn a model together publicly, and only the gradients and losses are transmitted between nodes, and not the data itself, which can protect the data well. However, the nodes need to transmit the data to the central institution for the next computation. Blockchain can be a good alternative to the central institution and avoids the dishonesty of the central structure [39]. Figure 4 shows a federated learning model where the blockchain replaces the central authority.
Rahman et al. use blockchain and off-chain techniques to protect the source data itself from tampering and unauthorized access [40]. The central gradient aggregator is replaced by the blockchain. At the end of aggregation, the hash of the encrypted global model is stored in the blockchain for further sharing. Malicious FL nodes may introduce toxic models, which may add bias to the training data. A monitoring mechanism is introduced to query the history and authenticity of the training process. Malicious nodes can be detected. The malicious nodes can be reduced by reputation score. Sharing computing power and data among nodes is an ideal situation, which is difficult to achieve in reality. Because the computing power and data between nodes are not the same. It is difficult to reach a consensus among nodes without profit. It is fairer to reward tokens based on the workload of nodes. A new fully decentralized healthcare architecture is proposed by [41]. The model is used to share distributed EMR between federal hospitals based on blockchain and MEC. To facilitate EMR sharing, they designed a new decentralized EMR storage on MEC servers using the InterPlanetary file system (IPFS) platform. In particular, an access authentication mechanism was developed using a blockchain-based smart contract to authenticate access at the edge of the network without any central authority. In the consensus phase, only some miners are selected for Byzantine protocol verification in each round. Although this approach reduces the overhead, it increases the risk of being attacked. The authors of [42] propose a blockchain-smart healthcare-based FL framework. This model provides the governance of the entire training process. An adaptive differential privacy algorithm is proposed to add an additional security layer to FL. The algorithm adapts the noise according to the training process, balancing privacy and model accuracy. Finally, an efficient consensus protocol based on gradient verification is designed to encourage reliable IoT devices and edge nodes to contribute their data and computational power to federation learning. Blockchain replaces centralized institutions that may be risky. This avoids the situation where a central node is evil, and each transaction is recorded on the chain, which enables timely detection of malicious nodes and provides oversight. Blockchain also has unique economic properties that can motivate nodes to participate in model learning by posting tokens.
3.4. Traditional Methods Based on Cryptography

Data encryption is the traditional method of data protection, and this section lists some ways to protect data privacy using encryption algorithms. Finally, the traditional methods are compared with blockchain-based methods.

A lightweight encryption algorithm with a shorter secret key computation time has been proposed by Hasen et al. [43]. This algorithm solves the problem that traditional encryption algorithms are not applicable to medical image data, and the algorithm obtains a lower signal-to-noise ratio. Yang et al. propose the use of a plaintext encryption method, which embeds private data into medical images [44]. The correlation with the original image is intuitively difficult to see in the plaintext encrypted image, which reduces the chance of being attacked. David et al. optimized the traditional homomorphic encryption model [45]. First, edge computing is used to speed up plaintext encryption. Then, avoiding the use of complex centralized encryption algorithms reduces the high computational and communication overhead.

There are some problems with the traditional approach. For example, the risk of secret key leakage is greater when there are more organizations to share it with, and traditional cryptographic algorithms have no way to achieve fine-grained access control. The blockchain-based approach enables fine-grained access control through smart contracts, and many image data have some distortion after encryption and decryption.

4. Discussion of Potential Challenges

Some challenges of blockchain-based healthcare data sharing.

1. Blockchain capacity: Almost all models have adopted off-chain storage of original data, such as cloud and IPFS, and on-chain storage of data hashes to prevent data tampering. Increasing blockchain capacity and scalability in the future is a top priority.
2. Throughput: Throughput and latency are big factors that limit the development of blockchain. Bitcoin can only process seven transactions per second, and each transaction takes 1 h to determine. Ether has improved the throughput but still cannot fully meet the demand and needs further improvement.
3. Consensus algorithm: The consensus algorithm is an essential part of blockchain. A proper consensus algorithm can improve the security of blockchain and also reduce the transaction latency to increase the throughput. However, only a small number of models have improved the consensus algorithm. Optimizing the consensus algorithm according to the specific usage scenario can better improve the applicability of the model.
4. Anonymity: Anonymity is a double-edged sword. It protects the privacy of nodes but also brings additional risks to medical data. It makes it impossible to know the true identity of the nodes accessing the data. Especially in the public chain, it is unable to reject nodes trying to join.
5. Retrieval: Most of the models use encryption algorithms to improve security. However, ciphertext retrieval is more complex compared to plaintext retrieval. Almost all models ignore the workload of ciphertext retrieval. Only a very few papers have noticed and solved this problem.
6. Encryption algorithms: The encryption of medical data can increase the security of the data, but both encryption and decryption require a large amount of computing power. It is urgent to develop encryption algorithms with high security and low computational power.

5. Conclusions

The decentralized, traceable and tamper-proof nature of blockchain has made it of great interest in the field of healthcare data. This paper summarizes and analyzes blockchain-based medical data sharing from a unique perspective. This paper summarizes three application scenarios of healthcare data sharing and compares them with traditional approaches. Compared with the traditional cryptography-based model, the blockchain-based
model is more secure and intelligent because smart contracts play an important role. However, blockchain technology is suffering from issues, including low throughput and low scalability. These have limited the development of blockchain in healthcare data sharing. In the future, sharding, cross-chain and consensus algorithms are technologies that need to be focused on.

**Author Contributions:** Conceptualization, P.X. and X.Z.; methodology, P.X. and X.Z.; software, L.W. and W.L.; validation, P.X. and X.Z.; formal analysis, P.X.; investigation, P.X. and X.Z.; resources, L.W. and S.P.; data curation, L.W. and S.P.; writing—original draft preparation, P.X. and X.Z.; writing—review and editing, P.X., X.Z. and S.P.; visualization, L.W. and W.L.; supervision, S.P.; project administration, P.X. and S.P.; funding acquisition, S.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by NSFC Grants U19A2067; Science Foundation for Distinguished Young Scholars of Hunan Province (2020JJ2009); National Key R&D Program of China 2017YFB0202602, 2018YFC0910405, 2017YFC1311003, 2016YFC1302500; Science Foundation of Changsha Z20209420652, kq2004010; JZ2019524029, JH20199142034; The Funds of State Key Laboratory of Chemo/Biosensing and Chemometrics and Peng Cheng Lab.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**
