An Authorization Framework for Preserving Privacy of Big Medical Data via Blockchain in Cloud Server

Hemanth Kumar N P[1]
Assistant Professor, Dept. of CSE
East Point College of Engineering and Technology
Bengaluru, India

Dr. Prabhudeva S[2]
Director, Dept. of MCA
Jawaharlal Nehru National College of Engineering
Shivamogga, India

***Abstract*—In recent years, cloud-based medical record sharing has greatly improved the process of researching the disease and patient diagnosis. However, since cloud systems are centralized, there is serious concern about data security and privacy. Blockchain technology is viewed as a promising method of dealing with privacy issues and data security because of its exclusive features of distributed ledgers, secrecy, verifiability, and enhanced security. The literature review has shown significant works on integrating blockchain technology with cloud system for managing and sharing healthcare data. It has been analyzed that previous works are primarily dependent on the centralized data storage approach, which raises privacy concerns. The previous works also do not emphasize handling big medical data and lack the reliability of the end-to-end security features system. This paper has presented an authorization framework for ensuring data security and privacy preservation using blockchain technology with IPFS as decentralized file storage and sharing system. The proposed study devises a proof of replication algorithm using smart contracts to provide a better access control mechanism. The implementation of the proposed framework is based on the symmetric encryption and Ethereum blockchain platform. The study outcome illustrates the efficiency and availability of the proposed scheme compared to the typical cloud-based blockchain method.**

***Keywords*—Medical data; cloud; blockchain; data sharing; access control; security; privacy preservation**

I. INTRODUCTION

Healthcare is a scientific field that refers to a multidimensional system that includes a range of services primarily concerned with preventing, diagnosing, and treating disease and injuries related to human health. Healthcare professionals need a variety of information to provide better patient treatment, such as a patient's medical history, clinical evidence (imaging and laboratory tests), and private and personnel information [1]. The traditional method of storing such medical records for patients was to handwrite notes or printed papers that are easy to lose and difficult to hold for the long run [2]. The advancement of information and communication technology (ICT) has enabled the creation of electronic medical records (EMR), which are easy to keep for a long time compared to handwritten notes or paper-based records. The boom in the digitization of medical records has led to the formation of big medical data, which can be used for a variety of purposes in the healthcare sector [3].

Medical experts and healthcare institutions need to compare and analyze big medical data containing similar or related clinical features to examine the disease and seek better treatments [4]. Furthermore, most often, the patients may not be able to recall in detail their past medical conditions and symptoms. EMR sharing can assist physicians in learning more about their patients' health, therefore enhancing accuracy in diagnosis and leading to an effective decision-making process towards increasing the success rate of the treatment. Despite these benefits, two major problems exist, i) the storage of big medical data and ii) the security aspects related to EMR sharing across a healthcare organization. The deep learning market report [5] states that approximately 90% of EMR generated in hospitals are images. Due to large EMRs being generated, more storage space and computing power are required to perform analytical tasks to benefit healthcare services for patient well-being. On the other hand, sharing EMRs raises a significant and noteworthy concern regarding privacy protection, data security, and interoperability [6]. Firstly, electronic medical records contain personal and sensitive information, so privacy protection is the shield of a patient's reputation. Next, only authentic or trustworthy medical data can be used to make accurate decisions on diagnosis and treatment. Conversely, the forged EMR reduces the importance of clinical aspects and mislead in effective treatment planning. Moreover, interoperability can enable patients to control access to their EMRs and enhance sharing of EMRs across healthcare facilities.

Cloud computing has emerged as a promising solution that offers flexible storage management and efficient sharing of big medical data in response to these problems. The cloud computing ecosystem adopts different cryptographic primitives to secure medical data or EMRs and access control mechanisms to ensure privacy-aware data sharing [7]. Despite the high emphasis on data security and privacy protection, few serious security concerns still remain. Firstly, cloud systems are assumed to be trusted when it comes to storing, managing, and distributing data. In this regard, the design of cloud systems for data management and sharing heavily relies on the involvement of third-party mechanisms, which are prone to leak, theft, and tampering due to lack of surveillance [8]. Unfortunately, there is no standard verification mechanism for existing schemes, and also there is no effective countermeasure to penalize a misbehaving server or cloud entity.

In recent years, blockchain technology has appeared as a potential solution to address security loopholes and ensure reliable sharing of EMRs using a distributed ledger [9]. Since the blockchain is open and transparent, blockchain-based medical data sharing can help patients to have better access control and monitor the use of medical data [10]. However, despite the many benefits of blockchain technology, most of the current work suffers from the following three major issues: i) How to design an effective mechanism for verifiable security of EMR in the blockchain. Another problem is that the blockchains are not meant for big EMR or medical data (like high dimensional medical images) as they are not scalable quickly. When data is stored on the blockchain, the information is available to everyone. Therefore, the second problem can be highlighted as follows: ii) How to ensure that only the authenticated person can access the EMR data; and iii) How to design a computationally-efficient mechanism, block structure, and fault-tolerant mechanism that can ensure system reliability and availability for a longer run. All these factors are not adequately addressed in the existing literature.

As a motivation, the proposed research work emphasizes developing a big medical data authorization framework that can maintain an effective balance between security, privacy, and scalability. Apart from this, it also ensures full-proof security for the seemly management and sharing of big medical data in the distributed environment. The major contribution of the proposed research work is highlighted as follows:

- The proposed study employed an Ethereum blockchain platform on an amazon cloud system to explore the effectiveness of smart contracts in user validation and access management.
- The study implements InterPlanetary File System (IPFS) as a peer-to-peer network system for storing medical, thereby eliminating issues associated with a centralized system.
- The study considers high dimensional medical images as big data further split into multiple and equal shards. Thereby, the study handles scalability issues in the blockchain system.
- A mechanism of key-pairing is employed using an asymmetric encryption algorithm to each shard of medical data. This mechanism ensures data security so that data is not visible to the user or IPFS nodes that store the data.
- A novel proof of algorithm is developed that handles each transaction related to data storage access and facilitates better management of data sharing.

The remaining sections of the proposed manuscript are described as follows: Section II presents the related work and highlights some significant issues explored based on the review analysis; Section III presents the proposed system model and methodology adopted; Section IV presents the system design implementation; Section V presents the result analysis and discusses the performance of the proposed system, and finally Section VI provides a conclusive remark on the entire contribution and findings of the proposed study.

II. Review of Literature

This section briefly reviews the previous works on secure storage and sharing of medical data or EMR in digital healthcare systems using blockchain.

The first attempt towards medical data sharing using Ethereum blockchain is made by Yue et al. [11] to enable patients to control the access of their health-related information without compromising any privacy risk. Although this mechanism ascertains privacy preservation, it has some significant limitations: it does not provide access to patient family members, an essential concern in emergency situations. Apart from this, the model lacks the scalability feature. In a similar line of work, Jaiman and Urovi [12] suggested an access control framework for EMR sharing concerning patient consent. This work emphasizes joining data use ontology and access matrix that holds information about the data requestor. Considering these features, the data owner, i.e., patient or doctors makes EMR sharing rules, monitors its usage, and updates the access policy at any time. The existing approaches have not considered the efficiency factors in sharing medical data that continuously stream from bio-sensors and monitoring devices. In this regard, Shen et al. [13] presented an efficient scheme that combines peer-to-peer networking techniques with blockchain and digest chain to bring efficiency and flexibility in the sharing of medical data. The authors have introduced a scheme of the authorized network of participants to enable secure sharing of EMR between different healthcare departments, pharmacies, and patients. However, this approach has a security loophole because the storage of data in on-chain is vulnerable to scalability and privacy issues. Similarly, Dagher et al. [14] suggested a privacy-preserving scheme concentrating on the interoperability and access of the EMR using blockchain. This scheme adopted smart contracts stored on the Ethereum model, which holds hashing key of the EMR for ensuring interoperable. On the other hand, an advanced cryptography mechanism is employed to ensure secure sharing of EMR. Although this scheme is subjected to the high storage cost, it can't be applied to process big medical data and is open to vulnerability as it reveals information about the transactions. The above approaches lack off-chain policy and scalability, which is addressed in the proposed work using IPFS technology.

Zhang et al. [15] also aimed to address the issue of scalability by implementing the Ethereum model of blockchain technology. In this work, the ciphered data is stored on the on-chain system, which refers to the original medical records, while original medical records are kept over the off-chain system. The work of Madine et al. [16] has attempted to address the risk of single-point failover for the EMR stored on the cloud. The authors have applied the Ethereum model for devising smart contracts to enable patients to have full control over their data in various ways, such as transparency, traceability, undisputable, and security. IPFS system is employed as decentralized storage, and a re-encryption mechanism is adopted to provide data security. A blockchain-oriented digital healthcare system is presented in the work of Huang et al. [17] to attain robust security and traceability of the EMR sharing over the cloud system. An attribute re-encryption technique is applied to achieve complete access control for data

providers. An approach of a lightweight data sharing scheme is provided by Su et al. [18] using blockchain technology to achieve data privacy in the healthcare sector. The authors have applied an interleaving encoder to encrypt the medical data. Xu et al. [19] have given the mechanism of the health chain to support the requirement of privacy preservation for big medical data using blockchain and encryption to ensure a fine-grained access mechanism. This work also focuses on the key-management process to efficiently revoke or update keys for the authorized data requester.

Another significant work given by Wang and Song [20] adopted an attribute-based encryption technique combined with a blockchain mechanism for the cloud-assisted EMR sharing system. However, this approach failed to solve a problem related to the storage of big medical data and its sharing in an optimal way. Margheri et al. [21] introduced an information management model for monitoring EMR using modular blockchain system-based smart contracts. The work carried out by Guo et al. [22] modeled a hybrid system to enable better access control of EMR using blockchain. This scheme provides tamper-proof features by managing customized access policy, and off-chain nodes are employed to enforce attribute access mechanism on EMR data. Rajput et al. [23] suggested an approach of EMR access policy in an emergency based on Hyperledger fabric and composer. The authors have derived some customized policies under a smart contract for accessing emergency conditions for a specific time duration. The work carried out by Roehrs et al. [24] gave a prototype of the distributed blockchain to enhance the replication of data across nodes by combining medical records using blockchain and open interoperability. Another work of Guo et al. [25] have devised an attribute-oriented signature technique with blockchain, where a patient approves an EMR while providing no information other than evidence that he or she has substantiated it. The work carried out by Chen et al. [26] presented a conceptual design of a data storage system based on blockchain and a cloud system to manage personnel healthcare data. The study claimed that this storage system is independent of a third party and also does not allow a single party to have complete control over the access and processing of the data. However, none of the existing studies have provided standard management and sharing schemes for the medical data for pharmaceutical scientists. The work of Fan et al. [27] presented a clinical information management system using distributed ledger and consensus mechanism without much depending on the network resources. In a similar direction, a recent work carried out by Bataineh et al. [28] developed a security model for IoT-based healthcare systems using Ethereum Blockchain for surgical process management. Different from previous works, an application of fuzzy logic with blockchain is considered in another recent work by Zulkifli et al. [29] to provide an adaptive security mechanism for IoT-oriented healthcare.

For the management and sharing of medical data, different authors attempted to suggest solutions from different perspectives. There has been a huge development towards blockchain-based healthcare since 2016, which was the initial year till 2022. The analysis shows that most previous works adopted Ethereum or Hyperledger Fabric blockchain platform, and most of them provided conceptual and experimental approaches. Table I summarizes the above-discussed literature concerning blockchain platforms and solution types.

It has been identified that many previous works failed to ensure maximum requirements of the security such as privacy, data security, access control, interoperability, storage, scalability, and system cost analysis for secure sharing of medical data or EMRs. Although, most of the existing works have emphasized efficient access control and privacy preservation mechanisms using attribute-based re-encryption before introducing medical data to the blockchain system. Few existing works suggested the implementation of smart contracts, and some have employed a chain-coding scheme for privacy-preserving of EMR. On the other hand, the hospital continuously generates massive data as the number of people coming to the healthcare center is countless. Among the literature which is being reviewed, it has been found that most of the works have not considered the issue associated with big medical data storage. The authors in [9] have considered the issue. However, it lacks details on the storage services. Moreover, there are few relevant works [14],[16],[23] [22] that have been considered IPFS as a medium of data storage. The solutions given by these works can overcome the big data issues; however, it needs more optimization to handle a considerable amount of EMR data especially high dimensional medical data. In particular, the previous scheme for sharing and storing EMR using blockchain is still in its infancy stage, involves high cost, lacks scalability, and needs more effort in the design and development. Table II highlights the finding of the review analysis.

TABLE I. SUMMARY OF THE PREVIOUS WORKS BEING REVIEWED

Citations	Blockchain Platform	Solution Type
[11]	Not Defined	Implementation
[12]	Ethereum	Experimental
[13]	Not Defined	Experimental
[14]	Ethereum	Experimental
[15]	Ethereum	Implementation
[16]	Ethereum	Experimental
[17]	Not Defined	Conceptual
[18]	Bitcoin	Conceptual
[19]	Doc-chain	Experimental
[20]	Not defined	Conceptual
[21]	Hyperledger Fabric	Implementation
[22]	Hyperledger Composer	Experimental
[23]	Hyperledger Fabric	Experimental
[24]	Customized	Experimental
[25]	Not Defined	Conceptual
[26]	Not Defined	Conceptual
[27]	Not Defined	Conceptual
[28]	Ethereum	Implementation
[29]	Hyperledger Fabric	Experimental

TABLE II. HIGHLIGHTS OF THE RESEARCH ISSUES

Issues	Addressed By	Not Addressed By
Privacy and Security	[11-12], [14-23], [25-29]	[13], [24]
Accessibility and Interoperability	[12-15], [17-24], [26], [27]	[11], [16],[25]
Scalability	[14-16] [18-19] [21] [23] [26-27]	[11-14], [17], [20], [22], [24-25], [28-29]
Cost Analysis	[12], [18-19], [25], [28-29]	[11] [13-17], [20], [23-24], [26-27]
Big medical data	[14], [16], [19], [22-23]	[11-13], [15-18], [24-26]

III. SYSTEM MODEL

The proposed study presents a novel model for medical data authorization towards security and privacy preservation using blockchain methodology. The study considers big data as a medical image. The rationale behind considering the medical image as input data to the system is that the medical images are of high dimensional data associated with significant storage and scalability, which is a big data problem. Fig. 1 depicts the high-level architecture of the system model for big medical data based on blockchain that includes data sharing and its management over the distributed healthcare environment.

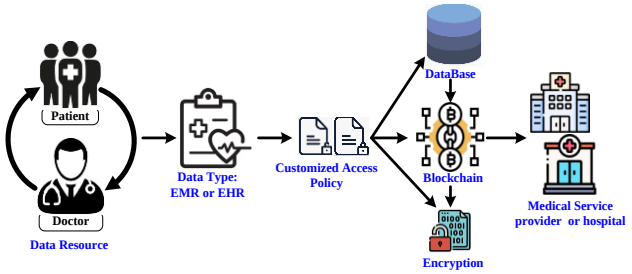

Fig. 1. High-level Architecture of the System Model.

A. System Model Components

As shown in Fig. 1, there are five main entities in the system model viz. i) data resource, ii) Data Type, iii) Access policy, iv) Database (medical data storage), and Security (encryption and blockchain), and v) medical service provider.

1) Data resource: The major medical data sources are the patient and the doctors. When patients interact with the physician, they share information consisting of previous records of the patient's health, drug history, current health issues, and other physiological symptoms.

2) Data type: The electronic health record (EHR) or electronic medical record (EMR) is built based on initial health information collected during the interaction between the patient and the doctors, after which the patient often undergoes other clinical examination such as laboratory tests, pre-operative assessment, and medical imaging. All these data are digital and stored in an electronic storage system. These digital records termed EMR or EHR, provide a holistic and long-term view of patient health.

3) Customized access policy: The patient he/she is the owner of their sensitive or private information in the EMR. However, in real-world scenarios, medical data is available with both patients and healthcare organizations. Therefore, both entities can act as a medium of storing medical data to a cloud server. Therefore, a customized access policy is built for accessing the data by the doctors or the healthcare organization upon having the valid reason to share their data to the other healthcare organizations (data requester).

4) Database and security: This component of the system model involves the core part concerning data storage, security, and privacy. The data is stored in a distributed manner, either on a centralized or decentralized server. The integration of encryption with blockchain technology offers higher security privay preservation and ensures the authenticity of the customized access policy. Blockchain is a system where there is no central authority to share the data, but even then, anyone can trust that the data is genuine.

5) Medical service providers: The medical service providers refer to healthcare facilities such as ad-hoc clinics, clinical laboratories, radiological centers, and hospitals interesting in accessing the medical data are treated as data requesters. According to the customized access policy, the data is provided to the data requester to decide the best strategy for surgery and treatment. Also, depending on the context, the patient can act as a data requester who requests to access the data stored on the cloud server. For example, sharing medical data to other healthcare facilities or organizations can faciliate better diagnosis, medical research, policy defining and effective treatment no matter where the patient is treated in the world if the healthcare records are available independent of time and place.

B. Need of the System

The proposed system model can address the following issues:

1) When the data is stored in a centralized server, there is a risk of a security issue, privacy leakage, and identity theft either by the creators of the system or by hackers.

2) The big medical data are not meant for blockchain as they are not scalable quickly.

3) When data is being stored on the blockchain, the data is available to everyone.

4) The solution on the blockchain is often subjected to higher costs as it requires a large processing time for the successful execution of the transactions.

C. Solution Strategy

1) The blockchain is a no-trust system where the code takes care of the issues associated with centralization and privacy.

2) The study considers the high dimensional medical image as the big data, which is further split into shards as 3 data copies of equal size and which handles the issue of big data into the blockchain. Therefore, the data become scalable.

3) The data is encrypted using an asymmetric algorithm that ensures data security and prevents unauthorized access.

4) The data is stored on a miner's device. The proof of replication algorithm is proposed to ensure that the miners are storing the data. While recovery, this algorithm ensures the file recovery with the least bandwidth.

5) In this work, the Ethereum blockchain provides secure and distributed sharing of medical records over the unsecured channel.

IV. System Implementation

The proposed study devices an authorization framework for medical data, which provides access to data requester in a distributed environment using blockchain. The study considers a high-dimensional medical image of the patient-generated from the radiological department. Here, the patient must create a customized access policy and smart contract in the Ethereum blockchain. On the other hand, the radiologist is accountable for uploading data to the IPFS network. The uploading of this data requires sending accept file storage requests by IPFS to blockchain clients, and similarly, access to storage is also required for retrieving the data on their mobile phone or computer. The data can be retrieved by either patient or the doctors in the other hospital to download previous medical records and upload current or new medical reports. The IPFS network is a peer-to-peer network system, where the user as storage provider needs to create an account to become a blockchain client and provide storage of their computing nodes. Due to bandwidth and storage constraints, the user here uses a cloud server as a storage point rather than storing on the local computing nodes. Upon accepting request regarding accept file storage, the data is encrypted using asymmetric encryption (RSA) as primary level security for data protection at the location of the data storage provider (IPFS). The encrypted data is further introduced with the blockchain module. In this module, the sharing of the secret key information is carried out over blockchain, which automatically changes the private and public key pair so that the user has complete controller over either allowing or revoking access to the other user. Fig. 2 presents the process flow of the proposed authorization system for secure uploading and sharing of big medical data.

- The data service provider interacts with the IPFS network using the web3.0 interface and connects to the backend.
- The data service provider requests to accept file storage, blockchain client interacts with IPFS in the backend.
- IPFS network at cloud server verifies the blockchain client identity and reliability (bandwidth requirement, storage capacity) with proposed proof of replication algorithm (PoRA) and allows the file storage.
- On the blockchain client storage node, the medical data is encrypted and sent towards IPFS.
- MD5 hashes of medical data (image) are given to distributed hash tables (DHT) for protecting its integrity.
- The file storage transaction is verified by the transaction pool linked to Ethereum (Eth) blockchain. Gas price is paid at this time (miners in Eth1.0 validators in eth 2.0 in study 2.0 is used).
- Once the transaction gets updated, the response is given to the blockchain client, and the storage is used for the purpose of the application (storing patient data).
- Data requester, which can be patient or doctor either stores a new data or requests for access of a data.
- The request is forwarded to PoRA.
- PoRA checks the access policy (11) via smart contracts (PoRA is part of smart contracts).
- Access to storage is given to the Data requester.

A. Blockchain

In the present study, the Ethereum blockchain is being used, and the focus of this study is to build a medical image authorization system in a distributed environment. All transactions are logged in the blockchain. The transactions are of four types.

Fig. 3 presents a basic architecture of the medical data blocks over the blockchain (B.C.) with different components and transaction (T) records. The miner id refers to the identity of the transaction validators produced based on the proof of the work on the validator's computing devices or nodes. The second component, namely timestamp, refers to the first block created and the last transaction. The third component is the data hash value (H) SHA256 of the blocks for each transaction numerically represented as follows:

$$\mathrm{H12} = f(\mathrm{H1} + \mathrm{H2}) \rightarrow f(\mathrm{T1} \cdot \mathrm{H}) + (\mathrm{T2} \cdot \mathrm{H}) \quad (1)$$

Where, $f(\cdot)$ denotes hashing function, and the validation and integrity check of the blocks are done based on the previous H of the block.

The next component, namely data_id is the unique identifier of the medical data that belongs to the particular patient id, whereas patient id denotes the unique identifier of the patient or the data owner. The data stored under the cloud server establishes trust between the requester or provider and the IPFS network. A smart contract is basically a computer program that runs on the blockchain. The smart contract includes all operations of the access policy used to validate and accept requests for data access. The operation involved in the smart contract is available to all blockchain users. The collection of smart contracts controls the entire storage and access mechanism in the proposed system. Hence, there won't be a centralized authority for limiting access and storing the files. The transactions done over a blockchain are organized in accept file storage, new file storage requests, give access to a file, and revoke access to a file.

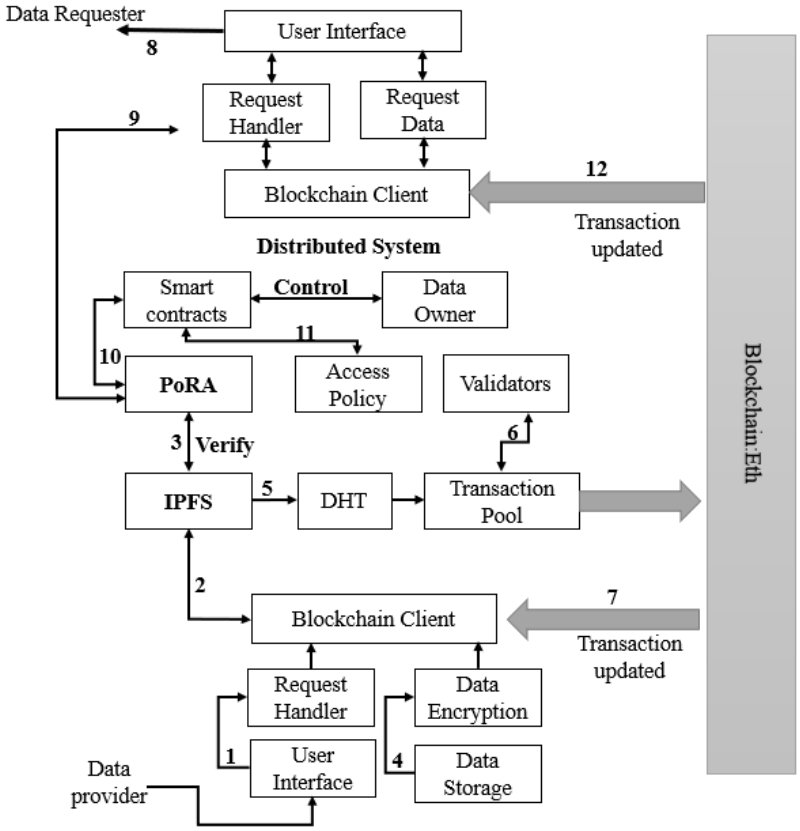

Fig. 2. System Implementation Process Flow for Secure Data Uploading and Sharing.

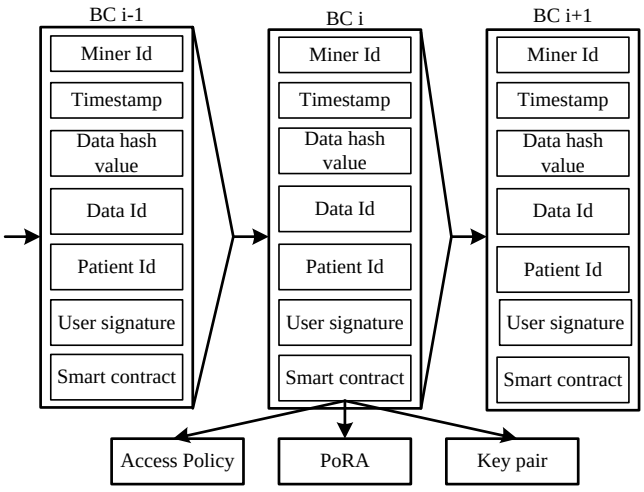

Fig. 3. Schematic Representation of Data Blocks.

B. Accept File Storage

The accept file storage transaction is subjected to the interaction of data service provider and data storage provider in the IPFS network. The data service provider needs a registration to have an account of blockchain, so that they can initiate a file storage request for uploading medical data to the IPFS distributed system meant for storing, sharing and accessing files. File storage request is typically accepted by the storage nodes. The storage nodes mean a user which is a part of blockchain client that facilitates their storage space either on the local machine or cloud server. In the current study, the storage nodes of the IPFS network are the cloud server. Upon receiving the request, the blockchain client interacts with the IPFS network and confirms the storage request, (i.e., the user is ready to provide its storage space). The IPFS verifies the file storage request by creating accept file storage transaction. In this process, the IPFS validates the identity of storage provider and checks are reliability in terms of storage and bandwidth capacity with PoRA using smart contracts and allow the file storage. The mechanism of file storage consists of few core operations as shown in Fig. 4.

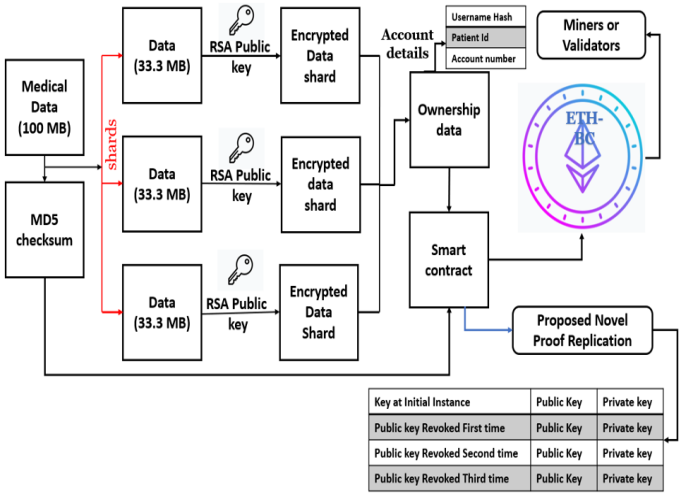

Fig. 4. Data Uploading and Storage Process.

1) In Fig. 4, the process of data storage on the blockchain client computing nodes in the IPFS network. The study considers medical data of size 100 MB, which is further split into the three shards with equal size. The ideology behind shards is that sharing and storing big data or high-dimensional data on the blockchain is challenging. Therefore, creating a set of equal-sized sub-images of the original image can provide a promising solution towards efficient file sharing operation in the blockchain. The study used an IPFS network, a decentralized file system based on the BitTorrent protocol, and the Distributed Hash Table (DHT). The IPFS system does not require a central server, and data can be stored in a network of distributed storage nodes, resulting in no single point of failure, higher storage throughput, and enhanced storage access and information retrieval mechanism over existing cloud storage systems. SHA256 hash values are calculated and stored in DHT to verify the integrity of medical images. At the same time, each shard is encrypted using the RSA asymmetric encryption algorithm and stored in IPFS storage nodes. As part of the study, smart contracts were integrated with the IPFS network to manage data access efficiently.

2) The patient metadata needs to be stored on the blockchain for the transaction; even though the medical data is stored on IPFS nodes, the metadata includes patient identification (I.D.), account numbers, smart contract details, hash values, and timestamps, among others. Mining pool transactions are conducted periodically by validators or blockchain miners. The fastest validators to verify a block of data will send the signature to the other validators for validation. After all, validators agree, valid blocks are added to the blockchain in consecutive order. Finally, the blockchain is synchronized by using the blockchain client to receive the identical copy of the blockchain. Furthermore, by storing hashes on the blockchain instead of raw data, the proposed model eliminates the risk of data leakage, ensuring high levels of data security.

C. New File Storage Request

This module of blockchain transactions is subjected to data access and storage updating with new data. The data requesters requested to access data or update the newly generated medical records of a particular patient in existing storage units of the IPFS network. A data requester is a patient or doctor in a medical facility. If the data requester is a patient, the system checks the blockchain record, verifies the patient's credentials, and allows access to the data store. On the other hand, if the data requester is a doctor or medical institution, the system using smart contracts verifies the requester's credentials. It authorizes access according to the access policy set by the patient. Fig. 5 shows the process of a new file storage request.

The data provider's entire process of data access and storage updating can be described in the following manner.

1) Data storage access request (accomplished by data provider): In order to access data on IPFS storage nodes in the cloud server, the data requester requires to execute a data access request transaction. If the data requester is a blockchain client, a key-sharing mechanism consisting of a pair of private and public keys is used for transaction authorization and its verification to access the blockchain. In addition, data requesters are required to provide a patient I.D. and account number when requesting access to the data. The transaction will then forward to PoRA for access request verification (refer to points 8 and 9 in Fig. 2).

2) Approval to access storage (accomplished by PoRA) : The data access request transaction is forwarded to the PoRA for the purpose of auditing and authorization. Then, the PoRA verify for this request based on the customized access policy and key sharing in the smart contract. Upon confirmation and authorization of the data requester's public key in the smart contract and access policy, permission is granted to the data requester to access the data storage on the IPFS network (refer to point 10 in Fig. 2). In addition, PoRA analyzes the authorized request identity, which contains the access request information related to the specific patient-ID and account number. It then searches for the hash value of the quarried medical data in the DHT to get the requested data on the IPFS storage node in the cloud server. The IPFS network then provides the requested data file in the ciphered form. PoRA uses a patient's private key to decrypt the ciphered medical data (image) and return the original plain medical data to the data requestor (refer to point 11 in Fig. 2). Here, transactions will be clustered and introduced into the transaction pool for mining. All validated transactions will be enumerated and linked to the blockchain to be shared with data requesters.

3) Updating data storage (accomplished by validators): As long as the requester has access to the data collection, such as patients, doctors, or healthcare providers can access the data of their interest. Also, they can request new file storage access to upload newly generated medical data to existing data storage nodes via a blockchain client (refer to point 12 in Fig. 2).

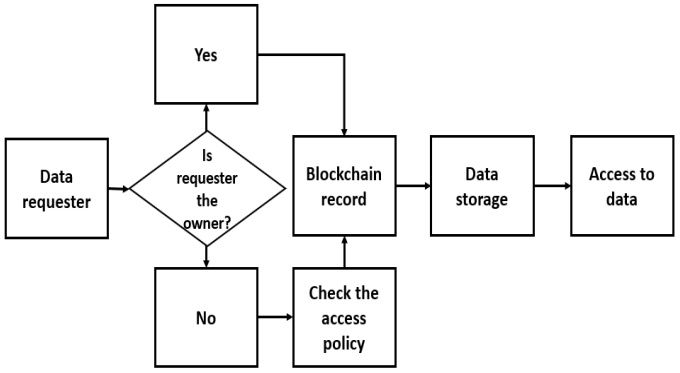

Fig. 5. Data Access and Storage Updating.

D. Give Access and Revoke Access to Data Storage

Upon verifying the data requester by the proposed PoRA, the requestor starts to make a transaction to access medical data by providing the unique identifier and account details of his patient to provide effective care accordingly. Once accessing data storage is completed, the transaction is updated and added to the blockchain by validators and acknowledged to all blockchain clients in the network. In this way, the patient can monitor their data sharing, thereby ensuring its ownership. However, to make the network reliable and ensure data security and privacy in the longer run, it is required to revoke the public and private key by updating to a new one after the purpose of data storage access is finished. Apart from managing the access control and sharing of medical data, the proposed PoRA also manages the key sharing process by securely revoking and updating the key that is known to the data requesters at each instance of data storage is accessed. Thus, eliminating the chances of key compromising and ensuring the trustworthiness of the network. The core procedure of the proposed PoRA algorithm is discussed in Algorithm 1.

Algorithm-1: Proof of Replication Algorithm (PoRA)

Input: file(F), Hash(H), Owner details(D), Access List(L)

Output: revoke key

1. L′ = Old Copy of Access List
2. for each user (U) in access List(L)
3. if U not in L':
4. P, P′ = newKeyPair () # P (public) & P′ (private) key
5. [F1, F2, F3] = sharding (F)
6. Fe←ENC (F1, F2, F3, P)
7. Transfer (F1, F2, F3)
8. Share private key to U
9. For each user U' in old access list(L')
10. If U' not in L:
11. P, P’ = newKeyPair () #
12. Run F_e = ENC (F, P)
13. For U in L
14. Share private key with U

The algorithm takes input values as a medical data file (F), hash value (H) of the file, owner details (D), access list (L). After successful execution, it returns key revocation for each instance of access completion. The smart contract goes through the entire list when the access needs to be modified. L is the new access-list, whereas L' is the old access-list (line:1-2). In the new access list, if a user U is not present in the old access list, then the user's access needs to be given (line:3). This is done by generating a new keypair (P & P') (line:4). In the next step, the medical data file is divided into shards and encrypted using function ENC( ), an asymmetric algorithm that takes shards and encryption key P(line:5-6). The encrypted file is then transferred to the IPFS storage nodes along with the private key to the user U (line:6-7). Further, the algorithm considers the users U' subjected to the old access list (L') and do not belong to the new access-list L. The proposed algorithm PoRA again generates a new key pair that encrypts the file and shares the user's private key. In this way, at every instance of access and its completion, the proposed algorithm revokes the previous access list and re-encrypts the file in the new list with a new keypair (), which is further updated to the authorized user or doctor.

V. Results

The design and development of the proposed system are carried out using Nodejs programming language installed on windows 10 Machine Intel(R)Core (T.M.)i7 16.0 GB RAM. Image dataset is used to validate the proposed model, and asymmetric encryption is used to protect the image. Ganache Simulator is used to simulate blockchain on the Local Machine. Kovan Testnet is used to simulate the entire program over the testing network with Keth (Ethereum with no real-world value) in order to test the gas price and processing time. Table III highlights configuration details of the system implementation.

A. System Parameters

TABLE III. System Model Configuration Details

Operating System	Windows 10
CPU	Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz 2.59 GHz
RAM	16.0 GB
Programming Language	Nodejs
Compiler	Solodity 0.6.0(London)
Test Framework	Mocha 6.2.0
Ethereum Platform	Ganache
Performance Testing	Kovan Testnet
Storage nodes	AWS nano
Local network speed	100 mbps
Hard disk type	Solid State Drive
GPU	Nvidia GTX

B. Result and Analysis

Table IV presents the numerical outcome obtained for the proposed algorithm PoRA. The performance of the proposed scheme is compared with the cloud-based blockchain system. It must be noted that 1 unit of gwei is equal to 1 nano Ethereum. The parameter time for storage and recovery represents how fast a file can be uploaded and recovered under the ideal scenario. The ideal scenario is represented by the simulation parameters, i.e., 100 MB of the data file. Since the storage is decentralized, any one of the nodes is always available.

The performance analysis of Fig. 6 shows that the proposed system PoRA outperforms the cloud-based blockchain system. The graph trend exhibits that the time for storage is less, i.e., 0.8 milliseconds in the proposed system PoRA, compared to the cloud-based system, i.e., 1 millisecond. It is because the smart contract used in the proposed authorization scheme has a smaller number of steps.

TABLE IV. COMPARATIVE ANALYSIS

Parameters/Techniques	Cloud-Blockchain	PoRA [proposed]
Time for data storage 100 MB	1ms	800us
Time for Accessing data 100MB	5ms	2ms
Minimum gas price	10382 gwei	5202 gwei
Typical gas price	20481 gwei	8642 gwei
Uptime	99%	100%
Downtime	1%	0%(negligible)
Data loss	No loss	no loss
Data integrity with MD5 hash	97%	99%
Storage space	limited by cloud	limited by number of nodes [Virtually unlimited]
File system	Linux file system	distributed file system
Maximum size of individual file	2 G.B.	Unlimited (Maximum file size 2 G.B.)
Main storage type	Centralized (Only Storage, cloud	Decentralized

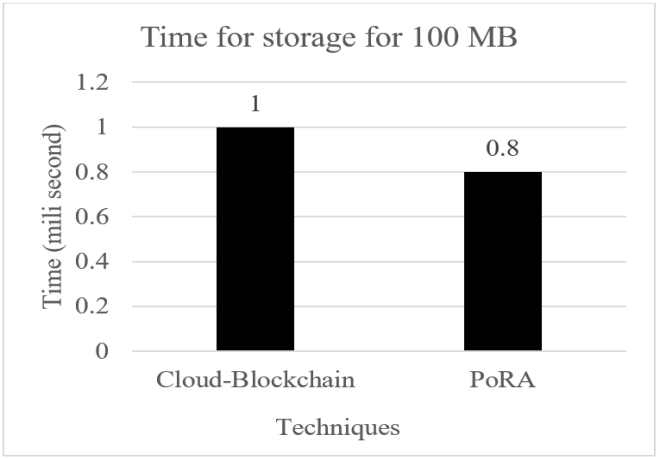

Fig. 6. Analysis of Storage Time.

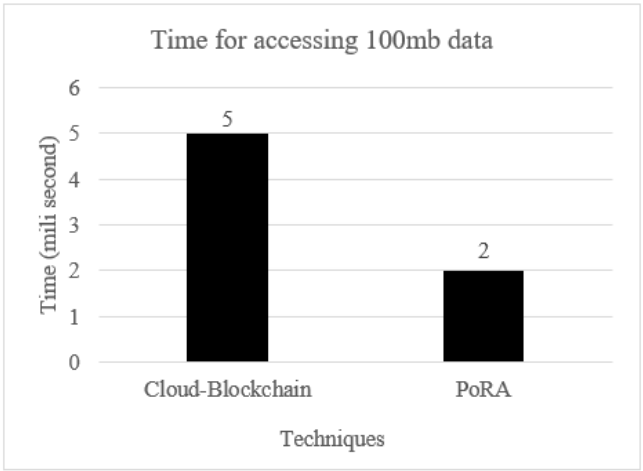

Fig. 7. Analysis of Access Time.

In Fig. 7, the performance analysis is shown for the time required for accessing the storage and recovering the file. The graph trend exhibits that the overall time is less for accessing the 100 MB of the file compared to the cloud-based system. The reason behind this is that, in the proposed system, the IPFS network is used as data storage nodes. The file is not stored in a central server but in a decentralized system where the system recognizes the nearest unit for the storage. As a result, this also reduces latency and the storage and recovery times of the data file. Apart from this, since more than one cloud is involved here, the storage space is virtually unlimited. The file size is limited by the md5 hash algorithm in the existing system since it is stored as a whole. However, in the proposed system, it is split into shards. Hence the file size is unlimited.

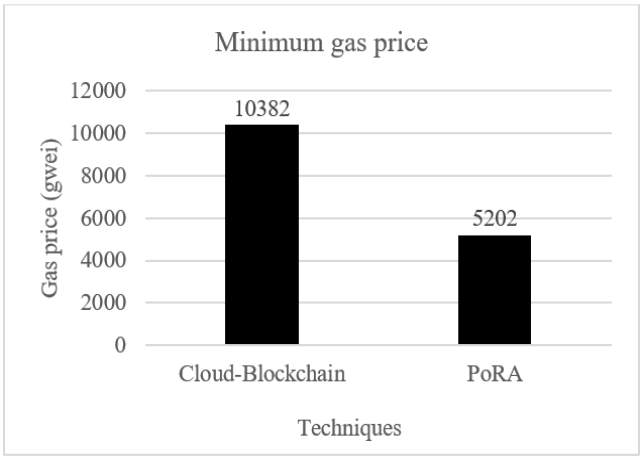

Fig. 8. Analysis of Minimum Gas Price.

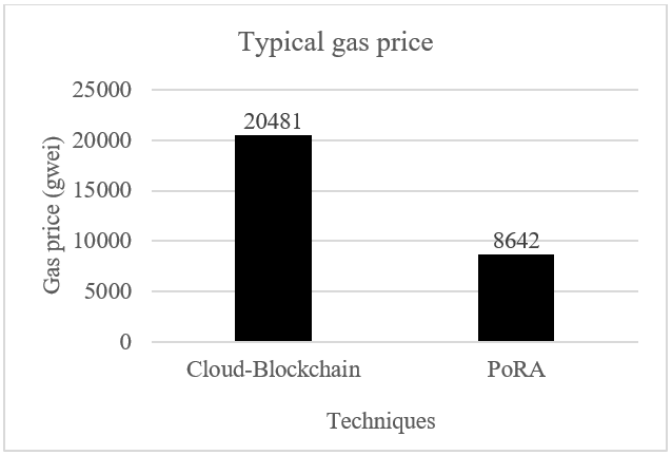

Fig. 9. Analysis of Typical Gas Price.

The blockchain client or any user pays the gas price in order to get his request accepted quickly by the blockchain. The gas price is decided by the number of times the smart contract updates a value in the blockchain. Fig. 8 and 9 present the performance analysis regarding the minimum gas price. The graph trend of both Fig. 8 and 9 exhibits that the proposed system has almost half of the gas price compared to the existing system. This means that using the proposed system is subjected to cost efficiency. In addition, the study has also carried out performance analysis regarding uptime and

downtime analysis. The analysis shows that the uptime is 100% in the proposed system, with a running time of 48 hours. The uptime might be lesser if the software runs for more hours. Also, data integrity is higher in the proposed system since the proposed design ensures the integrity of every shard of data rather than the entire file at once.

C. Discussion

The proposed system offers better flexibility compared to the existing approaches. The introduction of smart contract in the system design eliminates the trusted third entity's involvement in carrying out actions related to data processing. Therefore, cost-efficiency is introduced in the data-sharing process between organizations. The data scalability issue is narrowed down to a significant extent by introducing a mechanism of shards, which also leads to quick execution and fast processing with less delay. However, it may happen that when the user increases, the model faces some scalability issue over time. Future work will explore the scalability problem with the dynamic environment like IoT, and more optimization will be introduced to the model.

In the proposed system, the data owners, especially patients, are allowed to see their data being shared. They can change or customize the access policy for their data. They can also make a request regarding deleting and removing medical records if they want. In this way, the model ensures the agreement with privacy protection laws. The security of data is ensured with the encryption mechanism, and during the transaction, it will not reveal the patient information. The security and privacy features are the prime aspects in the proposed system modeling. Also, in future work, the study further explores the effectiveness of other variants of cryptographic approaches for securing the data-sharing platform.

VI. Conclusion

The core aim of the proposed work is to facilitate an efficient authorization framework based on blockchain technology for secure access of the patient medical data stored on the on-cloud server. Initially, the study has conducted a critical analysis of the existing literature and identified significant challenges associated with the previous works. To address the existing challenges, this work has suggested a novel approach of proof of replication algorithm to ensure the validation of each transaction on the blockchain and the key revocation mechanism. The study has established a reliable system using smart contracts to achieve an efficient access control and privacy-aware data sharing mechanism. The implementation of the proposed authorization framework study employed the Ethereum blockchain platform and IPFS network for the data storage on the AWS nano. Employing IPFS storage nodes in a cloud server eliminates the issues associated with a centralized system; the data is split into shards and stored in encrypted form, which leads to compensating the issue of scalability and data security. The outcome analysis shows the effectiveness of the proposed system regarding storage and access time, gas price, data integrity, and security compared to the typical cloud-blockchain-based systems. The proposed system can be applied in real-time scenarios to provide data security, privacy, and better medical data sharing process management. In the future, the proposed work can be extended with more optimization approaches in system computational requirements, customized smart contracts and scalability enhancement for the IoT-assisted big data applications.

References

[1] Hassanalieragh, M., Page, A., Soyata, T., Sharma, G., Aktas, M., Mateos, G., Kantarci, B. and Andreescu, S., 2015, June. Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges. In 2015 IEEE international conference on services computing (pp. 285-292). IEEE.

[2] Raghupathi, W. and Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. Health information science and systems, 2(1), pp.1-10.

[3] Sood, S.P., Nwabueze, S.N., Mbarika, V.W., Prakash, N., Chatterjee, S., Ray, P. and Mishra, S., 2008, January. Electronic medical records: A review comparing the challenges in developed and developing countries. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008) (pp. 248-248). IEEE.

[4] Topol, E.J., 2015. The big medical data miss: challenges in establishing an open medical resource. Nature Reviews Genetics, 16(5), pp.253-254.

[5] Report Linker "Deep Learning Market: Focus on Medical Image Processing", 2020–2030. Available online: https://www.reportlinker.com/p05987918/Deep-Learning-Market-Focus-on-Medical-Image-Processing (accessed on 16 December 2020).

[6] Bhartiya, S., Mehrotra, D. and Girdhar, A., 2016. Issues in achieving complete interoperability while sharing electronic health records. Procedia Computer Science, 78, pp.192-198.

[7] Zhang, R. and Liu, L., 2010, July. Security models and requirements for healthcare application clouds. In 2010 IEEE 3rd International Conference on cloud Computing (pp. 268-275). IEEE.

[8] Chinnasamy, P. and Deepalakshmi, P., 2022. HCAC-EHR: hybrid cryptographic access control for secure EHR retrieval in healthcare cloud. Journal of Ambient Intelligence and Humanized Computing, 13(2), pp.1001-1019.

[9] Shi, S., He, D., Li, L., Kumar, N., Khan, M.K. and Choo, K.K.R., 2020. Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. Computers & security, 97, p.101966.

[10] Sookhak, M., Jabbarpour, M.R., Safa, N.S. and Yu, F.R., 2021. Blockchain and smart contract for access control in healthcare: a survey, issues and challenges, and open issues. Journal of Network and Computer Applications, 178, p.102950.

[11] Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. Journal of medical systems. 2016; 40(10):218.

[12] V. Jaiman and V. Urovi, "A consent model for blockchain-based health data sharing platforms," IEEE Access, vol. 8, pp. 143734–143745, 2020.

[13] B. Shen, J. Guo, and Y. Yang, "MedChain: Efficient healthcare data sharing via blockchain," Appl. Sci., vol. 9, no. 6, p. 1207, Dec. 2018.

[14] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," Sustain. Cities Soc., vol. 39, pp. 283–297, May 2018.

[15] Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. Fhirchain: applying blockchain to securely and scalably share clinical data. Computational and structural biotechnology journal. 2018; 16:267– 278.

[16] M. M. Madine, A. A. Battah, I. Yaqoob, K. Salah, R. Jayaraman, Y. Al-Hammadi, S. Pesic, and S. Ellahham, "Blockchain for giving patients control over their medical records," IEEE Access, vol. 8, pp. 193102–193115, 2020.

[17] H. Huang, X. Sun, F. Xiao, P. Zhu, and W. Wang, "Blockchain-based eHealth system for auditable EHRs manipulation in cloud environments," J. Parallel Distrib. Comput., vol. 148, pp. 46–57, Feb. 2021.

[18] J. Fu, N. Wang, and Y. Cai, "Privacy-preserving in healthcare blockchain systems based on lightweight message sharing," Sensors, vol. 20, no. 7, p. 1898, Mar. 2020.

[19] J. Xu, K. Xue, S. Li, H. Tian, J. Hong, P. Hong, and N. Yu, "Healthchain: A blockchain-based privacy preserving scheme for large-scale health data," IEEE Internet Things J., vol. 6, no. 5, pp. 8770–8781, Oct. 2019.

[20] H. Wang and Y. Song, "Secure cloud-based EHR system using attributebased cryptosystem and blockchain," J. Med. Syst., vol. 42, no. 8, p. 152, 2018.

[21] Margheri A, Masi M, Miladi A, Sassone V, Rosenzweig J. Decentralised Provenance for Healthcare Data. International Journal of Medical Informatics. 2020; p. 104197.

[22] H. Guo, W. Li, M. Nejad, and C.-C. Shen, "Access control for electronic health records with hybrid blockchain-edge architecture," in Proc. IEEE Int. Conf. Blockchain (Blockchain), Jul. 2019, pp. 44–51.

[23] A. R. Rajput, Q. Li, M. T. Ahvanooey, and I. Masood, "EACMS: Emergency access control management system for personal health record based on blockchain," IEEE Access, vol. 7, pp. 84304–84317, 2019.

[24] Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. Analyzing the performance of a blockchain-based personal health record implementation. Journal of biomedical informatics. 2019; 92:103140.

[25] R. Guo, H. Shi, Q. Zhao, and D. Zheng, "Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems," IEEE Access, vol. 6, pp. 11676–11686, 2018.

[26] 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, p. 5, Jan. 2019.

[27] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "MedBlock: Efficient and secure medical data sharing via blockchain," J. Med. Syst., vol. 42, no. 8, p. 136, Aug. 2018.

[28] M. R. Bataineh, W. Mardini, Y. M. Khamayseh and M. M. B. Yassein, "Novel and Secure Blockchain Framework for Health Applications in IoT," in IEEE Access, vol. 10, pp. 14914-14926, 2022.

[29] Z. Zulkifl et al., "FBASHI: Fuzzy and Blockchain-Based Adaptive Security for Healthcare IoTs," in IEEE Access, vol. 10, pp. 15644-15656, 2022.