Taylor & Francis
Taylor & Francis Group

Computer Methods in Biomechanics and Biomedical Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gcmb20

Proposed association rule hiding based privacy preservation model with block chain technology for IoT healthcare sector

A. Yogeshwar & S. Kamalakkannan

To cite this article: A. Yogeshwar & S. Kamalakkannan (2022): Proposed association rule hiding based privacy preservation model with block chain technology for IoT healthcare sector, Computer Methods in Biomechanics and Biomedical Engineering, DOI: 10.1080/10255842.2022.2156287

To link to this article: https://doi.org/10.1080/10255842.2022.2156287

Published online: 29 Dec 2022.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING
https://doi.org/10.1080/10255842.2022.2156287

Proposed association rule hiding based privacy preservation model with block chain technology for IoT healthcare sector

A. Yogeshwar[a] and S. Kamalakkannan[b]

[a]Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India; [b]Department of Information Technology, Vels Institute of Science Technology & Advanced Studies (VISTAS), Pallavaram, Chennai, India

ABSTRACT

The purpose of this study is to improve healthcare system performance by utilizing cutting-edge computing technologies like blockchain and the Internet of Things. Blockchain-based data transfer, Association Rule hiding, and ideal key generation are the three primary aspects of the proposed work. Initially, data are altered using blockchain, then the data enter the Proposed Association Rule concealing stage. In this research a novel association rule concealment phase is implemented, which has three crucial processes: (1) data pattern mining using the improved apiori algorithm, (2) detection of sensitive data based on the improved apiori algorithm, and (3) a method for cleaning and restoring data. Using the generated optimal key, the sanitized sensitive data are recovered. Keys are critical to both the data sanitization and restoration procedures. Hence, a multi-objective hybrid optimization model is known as the Rock Hyraxes Updated Marriage in Honey Bee Optimization (RHUMBO) is employed. Then, the confidentiality of the suggested model's performance has been validated. From the experimental analysis the proposed model achieved 97% for Cleveland dataset at 90th learning percentage which is the best score. And the cost function of the suggested model is minimum (∼0.08 at 100th iteration).

ARTICLE HISTORY
Received 18 July 2022
Accepted 4 December 2022

KEYWORDS
Healthcare sector; Internet of Things (IoT); blockchain; proposed association rule hiding; RHUMBO (Rock Hyraxes Updated Marriage in Honey Bee Optimization)

1. Introduction

Since the healthcare industry is largely connected with people's social welfare and lives, it is instead a critical concern for both developed and developing economies. In the healthcare industry, research & development is ought to be an ongoing process since it will assist to raise the standard of living by combating numerous health conditions and diseases (Rathee et al. 2020; Chen et al. 2021; Zou et al. 2021). It has been simple to observe the improvement in the healthcare sector owing to technological innovation and recent achievements. The most advanced computer technology may significantly improve the capacities of the healthcare and medical sectors. The use of this cutting-edge computer technology helps doctors and healthcare staff recognize various illnesses early on (Chen et al. 2019; Sri and Bhaskari 2020; Arul et al. 2021). These advanced computer technologies can also dramatically enhances the accuracy of detecting illnesses in their initial stages.

IoT refers to the concept of linking everything to the World wide web (Mandala and Rao 2019; Shailaja and Rao 2019). Automobiles, household appliances, as well as other goods infused with electronics, and also software, sensors, actuators, and connections that allow these objects to interact, gather, as well as share data, are included with this class. For sensitive rules that discriminate against data based on community, sex, nation, etc, it offers privacy. (Ranjan and Kumar Paul 2018; Li et al. 2020; Stafford and Treiblmaier 2020). Thereby, in health care, patients are always been observed as well as cared for outside of the traditional clinical context *via* remote patient monitoring (RPM; in the home as an example). Patients can communicate with healthcare providers as needed. It also lowers medical expenditures and raises the standard of treatment. This is the primary reason why healthcare practitioners are looking at ways to make RPM available to the general public (Xia et al. 2017; Daraghmi et al. 2019; Akkaoui et al. 2020; Liu et al. 2020; Yang et al. 2020). Here, a smartphone with internet connectivity, an RPM application, and a tracking device designed expressly to track and transmit health data to smart contracts may be the main elements of an RPM system. (Wang et al. 2019; Garg

CONTACT A. Yogeshwar ayogeshwar818@gmail.com

et al. 2020). In RPM and the current smart city, initiative an increasingly important role is played by the Wearable gadgets as well as IoT. RPM and the current drive to build Smart Cities both heavily rely on wearable technology and the Internet of Things. A dynamic method is employed for Medical Image Encryption by utilizing Elliptical Curve Cryptography, which is mainly used for protecting data from healthcare sectors (Madine et al. 2020; Zhang et al. 2021; Sreekala and Varghese 2022).

Conventional methods rely upon cloud computing to store data offsite, and various organizations share information in a structured way. Privacy and security are indeed the two most significant problems in these technologies. For the sake of confidentiality, no data owner should deduce a specialized patient's personal information, including name, address, or contact information, from such a query of stored medical records. Patients will indeed be unwilling to disclose their medical information if personal confidentiality is not sufficiently secured throughout the data exchange (Chen et al. 2021). Patients as well as health organizations must believe that even if the data is not secured, the cloud provider will not disclose it (Wang et al. 2018; Zhuang et al. 2020). The encrypted data, on the other hand, would obstruct the sharing procedure. Cloud-based data solutions suffer from single points of failure, susceptibility, and complexity. As cloud-based information sharing is dependent on third-party companies, there's also a risk of security breaches, leakage, manipulation, or misuse (Guo et al. 2019; Li et al. 2019; Wang et al. 2019; Biswas et al. 2020; Zerka et al. 2020). While previous cryptography-related solutions successfully solved some cloud-related difficulties, the single point of failure problem remains unsolvable. Health information management, on the other hand, has a centralized access control system that is typically centered on responsibilities (Son et al. 2020). The RBAC (role-based access control) paradigm necessitates the creation of complicated rules to limit the overall access of various categories of user information.

The major contribution of this research work is:

- To perform data sanitization and restoration efficiently, an Association Rule hiding-based privacy model is proposed, which is based on a multi-objective optimal key generation approach
- To collect data patterns and identify sensitive data, an improved apiori algorithm is employed in which the input medical data pattern and the sensitive data are mined.
- Introduces multi-objective hybrid optimization model via Rock hyraxes Updated Marriage in Honey Bee Optimization (RHUMBO), which generates data sanitization and restoration optimally.

Rest of the paper is organized as: The literature works undergone in Iot-blockchian healthcare model is addressed in Section 2. Section 3 depicts the PROPOSED association rule hiding based privacy preservation model: an overview. Section 4 manifests the information regarding PROPOSED association rule hiding based privacy preservation mode. The illustration of the proposed work is highlighted in Section 6. The results acquired with the proposed work are comprehensively portrayed in Section 7. This paper is concluded in Section 8.

2. Literature Review

2.1. Related works

In Chen et al. (2021) have introduced a novel medical information system model for safer medical information transmission. Here, the blockchain was used for data dissemination. The data IoT from the user end was collected using the IoT. On the basis of the proxy re-encryption algorithm as well as the cloud servers, a new "anonymous medical data sharing scheme" was introduced for enhancing the confidentiality of the users' information during its transmission. In terms of "permissioned blockchain architecture Hyperledger Fabric, and a dual-channel Fabric deployment architecture," they have implemented the system.

In Zou et al. (2021) have projected a novel approach referred as SPChain ("blockchain-based medical data sharing and privacy-preserving eHealth system"). The EMRs of the patients were stored using the devised special key blocks and micro blocks, Therefore this information was quite easier to retrieve. In a privacy-preservation technique, the data of the patients were shared using SPChain using the proxy re-encryption schemes. This model had the attained least computational time in case of data storage as well as retrieval.

In Huang et al. (2020) have proposed a blockchain-based privacy-preserving scheme for transferring the patient's history (data) in a secure manner. This research work had focused on the data sharing that takes place between multiple entities like People, academic institutions, and unreliable cloud platforms. More particularly, the proxy re-encryption technology has been included and this had restricted the research institutions from unauthorized access.

In Rathee et al. (2020) have introduced a framework built on a reliable blockchain for transferring patient health records (including multimedia data) in a secure manner. The proposed model exhibited a higher success rate of 86%.

In Arul et al. (2021) have proposed the MMSDDF based on blockchain in IoMT for secure authority and ability to access patient records. This has been mainly devoted to ensuring security in the IoMT devices. Towards the healthcare application network, the Blockchain's key was applied, and this created a warning notification during unauthorized access.

In Chen et al. (2019) have projected a novel storage scheme with the intention of managing the personal medical data of the users. This proposed model was based on the blockchain and cloud storage. The projected model had restricted third-party access.

In Amir Latif et al. (2020) have introduced a blockchain-based healthcare system framework based on smart contracts. According to the proposed hypothesis, the changes occurring in the healthcare system could be represented in the blockchain by utilizing the concepts and tools of a public ledger. This approach was said to be secure, as it had preserved the confidentiality of the users data from third-party access.

In Sri and Bhaskari (2020) have investigated as well as analyzed blockchain-based encryption of patient data over a shared network. Employing the consensus approach, the Proof of Word and Interoperability for data retrieval and accessibility were verified. This paradigm was claimed to have a centralized source of trust and to have produced superior trade-offs.

In Luong and Park (2022) offered a zero-knowledge succinct non-interactive argument of knowledge (zk-SNARK). It was a privacy-preserving healthcare IoT system that was based on blockchain. The public blockchain's anonymity feature was employed to safeguard users' privacy. Without depending on a reliable third party, this approach also offered the authentication property against collusion assaults. The zk-SNARK approach was intended to identify instances where the same witness was utilized more than once during the anonymous authentication phase.

2.1.1. Settings

In Elhoseny et al. (2021) presented the AI-EPP method, which employed a graph-based optimal model to generate trustworthy nodes for data routing as well as a scalable and reliable system for data collection and transmission. Moreover, it completed registration and verification steps while boosting the legitimacy of the transmission with a cloud platform and symmetric digital certificates. Additionally, it guarantees the delivery of efficient services to the healthcare system by incorporating blockchain technology in distributed development with the little computational burden on nodes in the network.

In Hossein et al. (2021) designed modified BC networks and the Proof-of-Authority (PoA) consensus mechanism was implemented in healthcare systems to protect user privacy. This approach improved the system's scalability and performance. The blockchain-based approach allowed people to share their medical information with doctors while giving them control over their data. To increase scalability and throughput, the BC network nodes were segmented into several clusters, and each user was assigned to a specific cluster for the storage of his data and access policies. A hierarchical approach based on user-assigned IDs was devised to expedite search and access to certain clusters. Table 1 shows the advantages and drawbacks of the existing privacy preservation model.

2.2. Problem statement

The privacy and security of patient information is a primary concern with an EHS. According to the survey results, the conventional privacy presentation methods in the healthcare sector have the following implementation shortcomings: The database's secrecy is really not guaranteed in their research, absence of expertise, more vulnerable to rollback attacks, which expose private information, block generation speed is slow and so on (Huang et al. 2020; Sri and Bhaskari 2020; Arul et al. 2021; Chen et al. 2021; Zou et al. 2021). Authentication and key management are indeed the second categories of complexity that emerges in the EMR system. Several researchers use cryptographic processes to demonstrate that medical data is secure and reliable.

A "blockchain-based electronic medical record system" is presently being promoted by many researchers. The EMR system appears to be very well to the blockchain's inherent methodology. Smart contracts and consensus are two of the most basic features of the blockchain process. Only a few studies, suggested that smart contracts and consensus be used to verify and validate data. A number of existing EMRs are aimed towards medical practitioners. But the practitioners who are unfamiliar with patients do the necessary actions depending upon their requirements. The long-term viability of the EMR system will become the next consideration. The majority of patients

Table 1. Advantages and drawbacks of existing privacy presentation model in the healthcare sector.

Author [citation]	Methodology	Advantages	Drawbacks
Chen et al. (2021)	"lightweight medical data sharing scheme"	• lightweight and reliable approach	• Higher computational and communication cost.
Zou et al. (2021)	SPChain	• Provide high throughput and scalability • achieves low storage overhead	• suffers from rollback attacks and lead to privacy leakage • consumes huge computational time
Huang et al. (2020)	blockchain- based privacy-preserving scheme	• Less computing cost • Fewer startup nodes • low-cost computing power	• speed of block generation is slower
Rathee et al. (2020)	hybrid framework	• ensures security and • transparency of patient's record	• lower success rate (86%) • suffers from falsification attack, worm hole attack • higher transaction time
Arul et al. (2021)	MMSDDF	• achieves high accuracy ratio	• Higher latency range • Lower response time
Chen et al. (2019)	storage and sharing scheme	• does not depend on any third-party • no single party has absolute power to affect the processing	• need to and privacy of patients' medical • data
Amir Latif et al. (2020)	remix IDE	• promise of total privacy gain	• consumes huge cost • lack of database • interoperability
Sri and Bhaskari (2020)	consensus mechanism	• shows better trade off	• No consideration on data security and consistency
Luong and Park (2022)	zk-SNARK	• No need third party support • Have good anonymous authentication	• Proofs production phase needs a lot of processing power • Device management phase attained poor performance and high expense
Elhoseny et al. (2021)	AI-EPP method	• Lower computing power • Provide better scalability and reliability	• Lack of expertise.
Hossein et al. (2021)	Modified BC networks and PoA	• Shows better scalability and throughput	• No consideration on cluster management optimization • Problem in assigning the best cluster

continued to increase at a rapid pace. The standard EMR-based system appears to be incapable of coping with the aforementioned conditions.

To overcome the aforementioned issue, a scalable framework is necessary. A privacy model and a multi-objective hybrid optimization model are proposed in this research. The next section provides a detailed description of our proposed methodology.

2.3. Research gap

There are still many issues to be resolved even though certain extant privacy presentation methods are lightweight and trustworthy. The projected work must also take into account issues including lack of expertise, increased vulnerability to rollback attacks that reveal private information (A rollback attack is a particular type of cryptographic assault that forces a computer network or communications protocol to switch from a modern, high-quality operation method to a more traditional, lower-quality one that is often offered for backward compatibility with older hardware. This attack is cause when an unscrupulous person updates the present secure storage data with an earlier version that is still valid), slow block generating speed, weak patient privacy, and vulnerable to fraud and worm gap attacks.

2.4. Objectives

- To enhance the performance of the healthcare sector, computing technologies like blockchain and IoT are implemented.
- To efficient cleaning of data and restore data, a privacy preservation model is introduced.
- To generate data sanitization and restoration optimally a multi-objective hybrid optimization model is proposed.

3. Proposed Association Rule hiding based Privacy Preservation Model: an overview

3.1. Proposed methodology

An innovative model for preserving patient privacy was presented in this research. The projected work includes three primary aspects namely Proposed Association Rule hiding, optimum key generation, and blockchain-based data transmission. The steps followed in the suggested model are shown below (Figure 1):

Step 1: The acquired raw medical data D from wearable sensors (IoT devices) is initially stored on cloud storage. The data dissemination is maintained via the blockchain

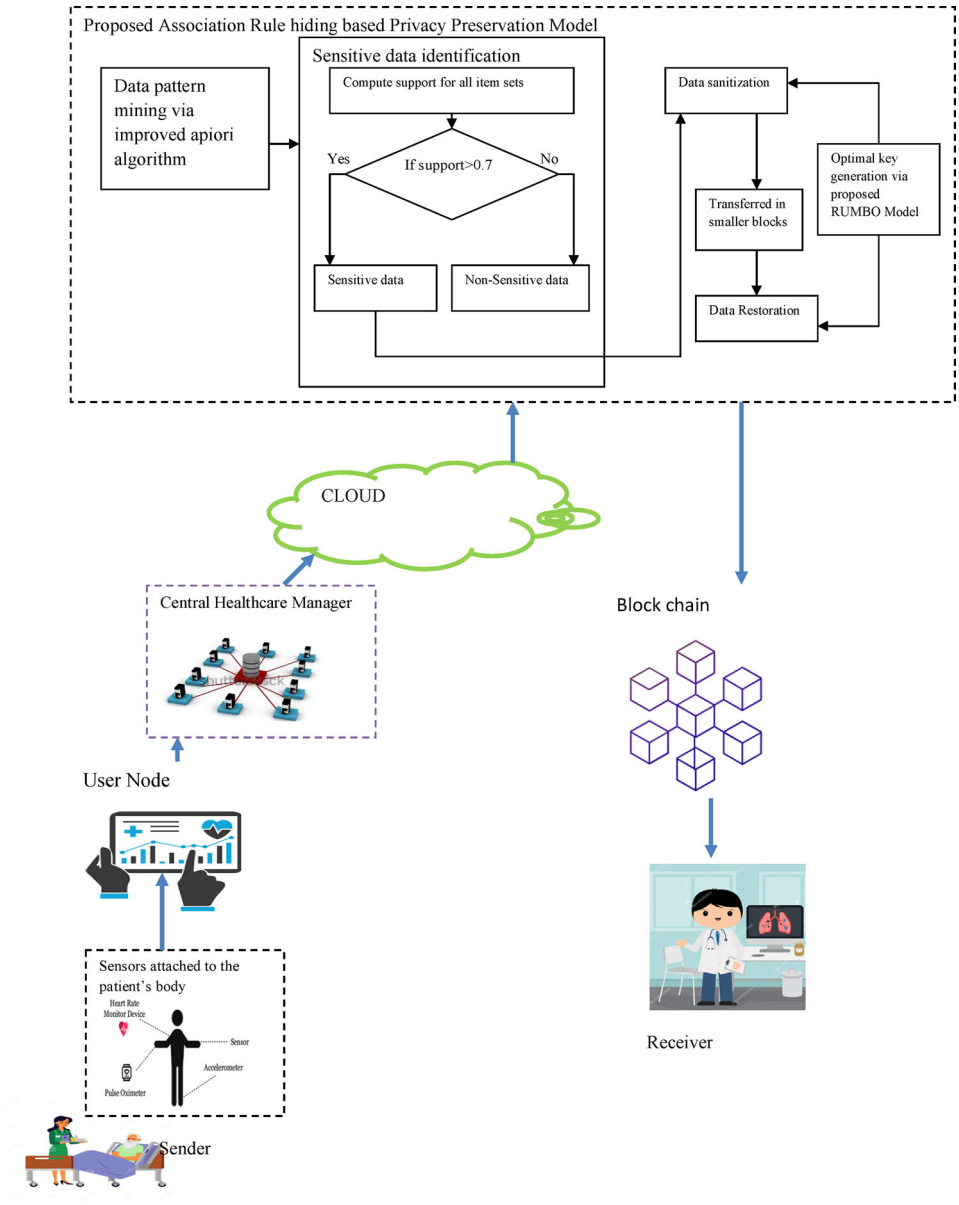

Figure 1. Framework of the proposed approach.

Step 2: The data is subjected under the proposed association rule concealing phase before the placement of blockchain.

Step 3: Three significant steps are there in the Proposed Association Rule hiding phase: (i) Data Pattern mining using the enhanced apiori algorithm, (ii) sensitive data identification based on the improved apiori algorithm's computed support value, and (iii) data sanitization and restoration technique. Initially, the suggested enhanced apiori algorithm is used to mine the data patterns of input medical data (acquired from the user node). The sensitive data S is then detected from the data patterns using the improved apiori algorithm's derived support value. The sensitive data is next subjected to a sanitization step, during which the sensitive sanitized data $S\prime$ is hidden using the proposed optimum key Key_2.

Step 4: This sensitive data S' was cleaned up for data distribution to be maintained via blockchain.

Step 5: The sanitized sensitive data S^* is recovered at the receiver end by the produced optimum key Key_2.

Step 6: The suggested multi-objective hybrid optimization strategy is utilized to manufacture the keys optimally because they are crucial for both data sanitization and restoration. The optimum key is built using a multi-objective mechanism that takes into account goals like the ratio of concealment, the degree of change, and the ratio of information

preservation. The proposed hybrid optimization model is referred to as the RHUMBO. This RHUMBO is the conceptual amalgamation of the standard MBO and RHSO. The hybridization hence provides optimum data sanitization and data restoration. Thus, the sensitive data will get transferred in a secure manner. Finally, the efficiency of the projected model is validated for security.

3.2. System model

The privacy protection of medical data in the Internet of Things is introduced in this study using a new blockchain platform. The framework consists of three main phases: optimum key generation, suggested data sanitization, and data restoration. Initially, the data owner O stores his/her data D (say EHR) onto a particular cloud C (say hospital's private cloud) *via* the IoT. In addition to O, there is N count of other cloud users $U_i; i = 1, 2, \ldots, N$ in C. All these users have utilized an IoT device like PDA (Personal Digital Assistance), mobile phone; wearable sensors laptop, or body sensor networks' to upload their data onto the cloud C. Since a huge count of users shares the same cloud resources as well as the same medical blockchain for data transmission; the security of users' data becomes a major challenge. More particularly, in case of EHR, there is a huge chance of misusing the data of other users, and this might be a life-threading situation to the concerned user. Therefore, there is an urgent need to design a novel privacy preservation platform that would enhance the functionality of healthcare systems. By taking up this challenge, a new IoT blockchain-based privacy preservation framework for medical data is developed in this work. The proposed work contains three main phases: (a) Proposed Association Rule hiding (b) Optimal Key selection and (c) blockchain based data transfer. As per the proposed work, the original data D of O is sanitized using the proposed association rule hiding approach, wherein a new Improved apiori algorithm is introduced. This improved apiori algorithm overcomes the problem of frequent database scanning and generating massive candidate sets. As a part of this, the sensitive and non-sensitive data are identified based on the computed support level for each item sets $I_1, I_2, I_3, I_4, I_5, I_6$ belonging to O. Now, the sensitive item sets being identified at the end of proposed data sanitization approach is maintained *via* blockchain. Sensitive data is delivered to the receiver at the opposite end of the blockchain, who then decodes it with the produced optimum key. The block chain is a core technology that has attained a considerable attention among the industry and academia. The security of the blockchain is completely based on the underlying data encryption. Moreover, the success of the data encryption model is completely depend on the generated optimal key. Therefore, in this research work, we've generated an optimal key by the meta-heuristic algorithms. This optimal key generation is based on a multi-objective function that encapsulates the three major objectives like: hiding ratio, degree of modification, and information preservation ratio. The proposed hybrid optimization model is the amalgamation of the Marriage in Honey Bee (MBO) and Rock Hyraxes Swarm Optimization concepts (RHSO). Thereby, the sensitive data will get transferred in a secured manner.

4. Proposed IoT-BC-based architecture

Sensor data can be delivered to the cloud using the Internet of Things application platform ThingSpeak. Real-time data gathering, processing of data, and visualizations are among the features of ThingSpeak. The software can also be used to combine, analyze, and visualize live data streams. Using web services, issues are alerted, instantly visualize real-time data, and upload data to ThingSpeak from the devices. The proposed IoT-BC-based privacy preservation framework for medical data includes the following elements:

a. User or Data Owners (say patient)
b. Sensors attached to the patient's body (IoT Device Node)
c. Patient's smartphone or PDA (User Node)
d. Central Healthcare Manager
e. Cloud Storage within the Cloud Service provider to manage and store data
f. Hospitals and health centers
g. Blockchain network
h. Mining and Miners

4.1. User or data owner's (say patient)

Suppose N count of patients $A_1, A_2, \ldots, A_N$ of a hospital H_1 uploads their data onto the cloud C_1 using the IoT. At the same time, the M count patients $B_1, B_2, \ldots, B_M$ in the hospital H_2 have uploaded their data onto the corresponding cloud C_2. In addition, there are two health centres H_3 and H_4 with their corresponding cloud storage C_3 and C_4, respectively. In H_3 and H_4, the patients are symbolized as $E_1, E_2, \ldots, E_S$ and $F_1, F_2, \ldots, F_S$.here, R, S denotes the count of patients in H_3 and H_4, respectively.

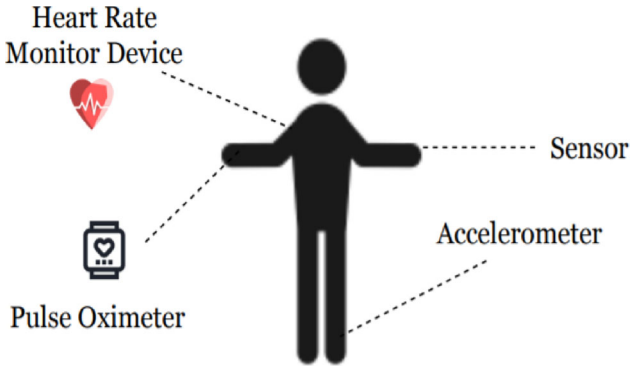

Figure 2. Wearable devices (IoT).

4.2. Sensors attached to the patient's body (IoT devices)

Each patient reportedly maintains a sizable number of sensors that are used to obtain data on their physiological state, including blood pressure, blood sugar levels, heart rate, sleep patterns, and calories burned. In fact, the sensors are batteries powered low energy, low storage and resource-constrained devices, and lower processing capability. Owing to these limitations, the collected data (say D_i) from the sensors are transmitted to the more powerful devices like smartphones of PDA using short-range communication like Bluetooth or Zigbee. These powerful devices act as a gateway for transferring data to healthcare servers. There is one or more management nodes for each of the IoT nodes, and these management nodes periodically update the collected health-related data towards the user node. Moreover, the IoT devices being lower-power devices with fewer storage capabilities and computational power aren't directly connected towards the blockchain. The illustration of wearable devices connected is shown in Figure 2.

4.3. Patient's smartphone or PDA (user node)

When compared to the sensors, the smart phones do posses high processing power and battery life. In addition, these Patient's smart phones or PDA are able to carry out heavy tasks like packet transmission and cryptographic operations *via* long-range communications.

4.4. Central healthcare manager

The Central Healthcare manager can be a PC for storing the patient's data $D_i; i = 1, 2, \ldots N$. The Central Healthcare manager is responsible for performing the following operations:

- Information is received from smartphones and stored on cloud servers.
- Executes hashing operations and other cryptographic activities.
- Transferring the data and policy hash to the medical blockchain network.

4.5. Hospitals and health centers

Healthcare facilities and hospitals are in charge of managing user-specific data. They monitor the BC network as well as the miners. Hospitals and clinics get user data hashes from the Central Healthcare Manager and store them.

a. The BC network could be contacted by registering users (patients, medical personnel, etc.) and allocating each of them an HW (comparable to the wallet in the bitcoin system).
b. For each patient, cluster miners are allocated. Cloud Storage: cloud storage servers have been utilized for storing the IoT healthcare data D_i of the users since they cannot be directly forwarded to the blockchain.

4.6. Proposed Association Rule hiding based Privacy Preservation Model

The user's data D_i are grouped by the cloud storage into identical blocks by associating them with a unique block number. Since the blockchain is highly susceptible to security issues; an Association Rule hiding-based Privacy Preservation Model is employed in this work. This is described comprehensively in the upcoming section.

4.7. Block chain

The blockchain is a DLT that transmits data (say sanitized sensitive data S'). Blockchain technology is said to be "tamper-resistant, decentralized, and unable to change a published transaction later inside the user community that shares the ledger," as per the NIST. A distributed data structure called the blockchain records every transaction. The blocks are connected to one another to form a chain-like configuration. The chain's very first link is Genesis. A Block Header, Transaction Counter, and Transaction are all present in every block. It records data in a decentralized manner. A transaction is regarded as legitimate if the system has proved that sufficient computational work has been exerted by authorizing nodes. By using private and public keys, individuals may manage and own their data through blockchain interactions. It is not permitted for third-

party intermediaries to obtain and misuse data. On a Blockchain, the data can save a document or file's cryptographic identity. Also, blockchain storage is more affordable and secure. This technology is also called DLT. The following key areas represent the main applications of blockchain in healthcare:

- Managing electronic medical record (EMR) data
- Protection of healthcare data
- Point-of-care genomics management

Using the medical blockchain, hospitals and health centers across the globe can be interlinked. This allows timely diagnosis and emergency care even during the pandemic periods like COVID-19. In general, the hash corresponding to the data and access policies is stored in the blockchain. The blockchain-based sensitive data transaction aids in boosting the availability as well as the integrity of the data corresponding to the users. Interestingly, the storage of these sensitive data within the blockchain prevents the data from falling into single-point failure or Dos Attacks.

4.8. Miners and mining

The process of adding a new block to the blockchain is referred to as mining. Every block in the chain is identified by a hash that appears in the header. Each header contains the address of the block that came before it in the chain. The blockchain is the leading technology for the healthcare system since data within blocks cannot be deleted or changed. As per our proposed framework, there is a massive count of users in each of the hospitals as well as health care centers. The new transactions are validated by the miners, and they are stored in the BC ledger.

5. Proposed Association Rule hiding based Privacy Preservation Model

The Proposed Association Rule hiding-based Privacy Preservation Model includes four major processes: (a) data pattern mining using improved apiori algorithm, (b) Sensitive data identification, (c) data sanitization and restoration model with Optimal key generation.

The steps followed in each of the phases are manifested below:

5.1. Data pattern mining via improved apiori algorithm

Initially, from the received user data D_i from the smartphones, the data patterns (itemsets) are mined for every transaction *via* the improved apiori algorithm. Among the available association rule mining techniques, the Apriori algorithm is said to be the most significant one (Jayasri and Aruna 2022). However, the standard apiori algorithm suffers from two major shortcomings:

a. From the candidate generation, the itemset size is utmost larger.
b. Consumes a huge time in computing the support values, wherein the itemset database has to be scanned again and again.

With the intention of overcoming these difficulties, we've introduced an improved apiori algorithm. The improved apiori algorithm applied within the Association Rule Hiding Model also aids in finding the sensitive items within the collected data *via* the smartphones.

The improved apiori algorithm overcomes the problem of frequent scanning of the database, and therefore the time consumed for encryption as well as decryption minimizes. Set Minimum support level = 10

To acquire a comprehensive view on the proposed improved apiori algorithm, let's suppose that there is ten transactions and six itemsets within D_1. Table 2 manifests these transactions and item sets.

Table 2. Items and their transactions.

Transactions (TID)	Item set
T1	I_1, I_2
T2	I_2, I_3
T3	I_1, I_2
T4	I_1, I_2, I_3, I_4
T5	I_1, I_2, I_3
T6	I_1,I_3,I_5
T7	I_2, I_3
T8	I_3, I_4
T9	I_4, I_6
T10	I_3, I_5

The steps followed within the improved apiori algorithm are depicted below:

Step 1: All the transactions are browsed with the intention of identifying each 1-itemset that corresponds to each transaction T_k. (here, k denotes the count of transactions). For each of the items, the count of TIDs is computed, and then the 1-itemsets with minimal support value are discarded; and $L_1 = \{\{I_1\}, \{I_2\}, \{I_3\}, \{I_4\}, \{I_5\}\}$.

Step 2: Create the 2-itemsets candidate by means of joining L_1 with L_1. Then, the support of the candidate items is computed using the overlapping strategy, and afterward, all the 2-itemsets with minimal support value are discarded $L_2 = \{\{I_1I_2\}, \{I_1I_2\}, \{I_2I_3\}, \{I_3I_4\}, \{I_3I_5\}\}$.

Step 3: The frequent k-item sets are generated (here the 3-item sets).

a. In prior to the candidate itemsets G_h (here, G_3). The L_{h-1} is pruned (here, L_2). within L_{h-1}, the count of occurrence of all items. Then, discard the itemsets that have numbers less than $h-1$ (here, $h=2$). In L_2, the itemsets I_4 and I_5 have occurred only once (i.e. it is lower than 2). So, the 2-itemsets I_4 and I_5 are discarded, and now we acquire $L_2' = \{\{I_1I_2\},\{I_1I_2\},\{I_2I_3\}\}$
b. With L_{h-1} L_2', the L_{h-1} (L_2') is joined, only when their prior $h-2$ elements are similar. Moreover, when the L_{h-1} (L_2) does not have the subset of G, then all the itemsets within $h-1$ subsets are discarded. Finally, the 3-itemsets candidates G_3 are acquired as: G_3=$\{I_1I_2I_3\}$
c. For $h-$ candidate itemsets (here G_3), the support is computed using the overlapping strategy. This computation is based upon the L_h and L_1 (L_1 and L_2). For I_1I_2, the TID is $\{T_1, T_3, T_4, T_5, T_6\}$ and the TIDs for L_3 is $\{T_2, T_4, T_5, T_4, T_6, T_7, T_8, T_{10}\}$. They both are combined, and the outcome is $\{T_4, T_5, T_6\}$. So, for the 3-itemset candidate $I_1I_2I_3$, the support value is 3. The itemsets that do possess a support level below the minimal value is deleted. Finally, the frequent h-candidate itemsets are acquired (here, $L_3 = I_1I_2I_3$).

Step 4: In $L_3(L_h)$, the count of itemsets is computed. If the count is less than that $3(|L_k| > k)$, the algorithm is terminated, and all frequent items are returned. In addition, if the count $L_3(L_k)$ of $>3(|L_k| > k)$, repeat step 3 and step 4. As per our example, the count of itemsets in L_3=1, so the algorithm terminates.

5.2. Sensitive rule identification using improved apiori algorithm

The sensitive rules are mined using the improved apiori algorithm. Within the itemset T_i, the underlying relations amongst the itemsets are identified. These associations are identified using the three major measures (a) support, (b) lift, and (c) confidence (Jayasri and Aruna 2022) (https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/).

Support sup() : The support provides information about the frequency of the appearance of itemsets during the transaction T. Let there be two item sets I_1, I_2. The association rules A are denoted as $I_1 \Rightarrow I_2$, and the set of transactions is denoted as T for the database D. Within the transaction, the support value of I_1 can be computed as the ratio of the count of transactions undergone with itemset I_1 to the overall count of transactions (https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/). Mathematically, the support I_1 $Supp(I_1)$ is computed as per Eq. (1).

$$Supp(I_1) = \frac{I_1 \subseteq T}{|T|} \tag{1}$$

Our contribution resides in the support computational phase. As per our contribution, when the support of the itemset is greater than 0.7 (threshold value), then it is said to be sensitive data, and these sanative data alone are subjected to the sanitization phase. The itemsets with support values less than 0.7 (threshold) are said to be non-sensitive, and these data are passed onto the blockchain directly.

$$\text{Proposed} - \begin{cases} if\ Supp(I_1) > 0.7 & Sensitive\ Data \\ if\ Supp(I_1) \le 0.7 & Non - Sensitive\ Data \end{cases} \tag{2}$$

Confidence: "Confidence is an indication of how often the rule has been found to be true". With respect to the transaction T, the confidence value for $I_1 \Rightarrow I_2$ is shown in Eq. (3) (https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/).

$$con(I_1 \Rightarrow I_2) = \frac{Supp|[I_1 \cup I_2]|}{Supp|I_1|} \tag{3}$$

Lift: The lift of a rule is depicted in Eq. (4) (https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/).

$$lift(I_1 \Rightarrow I_2) = \frac{Supp|[I_1 \cup I_2]|}{Supp(I_1) \times Supp(I_2)} \tag{4}$$

The identified sensitive data S acquired from D every transaction T are subjected to a sanitization phase for hiding the sensitive data S.

5.3. Data sanitization and restoration phase

Data sanitization is the process of concealing sensitive data in the cloud environment with the goal of preventing security breaches (i.e. unauthorized access by malicious users). The reverse of the data sanitization step is the data restoration phase. In the data sanitization as well as restoration phase, there is an urgent requirement to generate an optimal key for preserving the privacy of the data (Balashunmugaraja and Ganeshbabu 2020). In this research work, we took up this challenge and resolved it by introducing a new hybrid meta-heuristic optimization model.

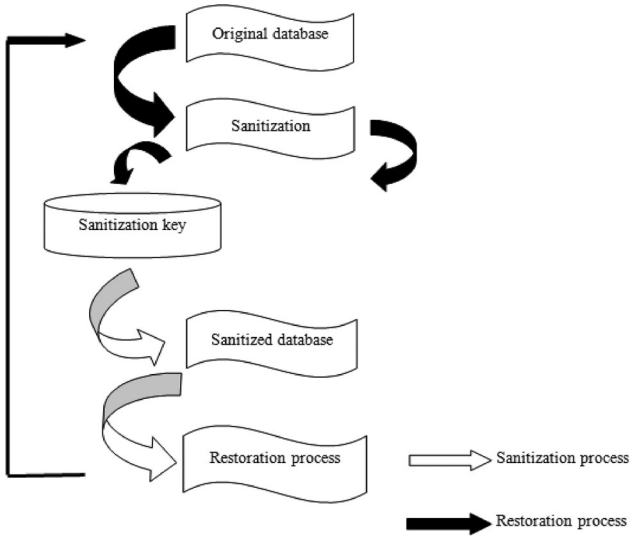

Figure 3. Data sanitization and restoration.

The flow diagram of the sanitization process is manifested in Figure 3.

The steps followed in the data sanitization and restoration phase are manifested below:

Data Sanitization

1. The generated optimal key matrix Key_2 with the proposed hybrid optimization (pruned key matrix) Key_2 and the Sensitive data S of the original database D are converted to binary form. This Key_2 has been generated on the basis of the multi-objective function.
2. The XOR function is carried out between the binarized key matrix Key_2 and Sensitive data S. This can be mathematically given as $Out = Key_2 \otimes S$. To the computed value Out a unit step input (1) is added with the intention of avoiding the probability of getting '0 '(i.e. to avoid data loss). So, the final sanitized sensitive data S' is denoted as $S' = Out + 1$.In case of sanitization, the rule hiding process is included S for acquiring the sanitized database S', wherein the sensitive information is hidden.

This sanitized sensitive data S' is passed onto the blockchain.

Data Restoration: The receivers at the other end of the blockchain acquire the sensitive data S' of the data owners by decrypting the sensitive data S' using the generated optimal key Key_2.

1. The sanitized sensitive data $S\prime$ and generated optimal key Key_2 are binarized. The binarized database from the binarization block is reduced from the unit step input.
2. The binarized key matrix Key_2 and binarized S' database are XOR-ed, and as a consequence, the restored database S^* is acquired. The restored database S^* will be the original sensitive data S. This mechanism of restoring the original sensitive data at the receiver end is mathematically shown in Eq. (5).

$$S^* = (S' - 1) \oplus Key_2 \tag{5}$$

5.4. Optimal key generation using the proposed multi-objective hybrid model (RUMBO)

In both the data sanitization as well as restoration phases, a pre-dominant role is played by the generated optimal key (Balashunmugaraja and Ganeshbabu 2020). This optimal key generation mechanism based on the multi-objective function is furnished comprehensively in this section.

Defined Multi-Objective Function: The Prime objective of this research work for generating the optimal key is depicted mathematically in Eq. (6).

$$Fitness = \min(Obj) \tag{6}$$

Here,

$$Obj = W_1\left[\frac{C_1}{\mathrm{Max}(C_1)}\right] + W_2\left[1 - \frac{C_2}{\mathrm{Max}(C_1, C_2)}\right] + \\ + W_3\left[\frac{C_3}{\mathrm{Max}(C_3)}\right] + W_4(1 - C_4) \tag{7}$$

In which, C_1, C_2, C_3 and C_4 are the cost functions; and W_1, W_2, W_3 and W_4 are the weight functions selected between 0 to 1.

The hiding failure ratio:

$$C_1 = \frac{Freq_{S'}}{Freq_S} \tag{8}$$

Here, $Freq_{S'}$ and $Freq_S$ points to the frequency of the sensitive itemsets in the sanitized data S' and the frequency of the sensitive itemsets in the original sensitive database S.

The information preservation rate:

$$C_2 = \frac{Freq_{NS}}{Freq_S} \tag{9}$$

Here, $Freq_{NS}$ denotes the frequency of the non-sensitive itemsets in the sanitized data S'.

Degree of modification:

$$C_3 = Dist(S, S') \tag{10}$$

Here, $Dist$ denotes the Euclidean distance between the original sensitive database S and sanitized

sensitive database S'. Distance between each itemset of sanitized sensitive data and original sensitive data:

$$C_4 = \frac{1}{A * B} \sum_{i=1}^{A} \sum_{j=1}^{B} \frac{S_{ij} - S'_{ij}}{\max(S_{ij}, S'_{ij})} \tag{11}$$

Here, A, B symbolizes the length of the original and sanitized sensitive database.

Based on the computed multi-objective function *Obj*, the key chromosome *Key* is subjected to the proposed hybrid optimization model. The key length (chromosome length) is set as $\sqrt{N''_T}$

The initial step for optimal key generation is the solution transformation, wherein the Key (*Key*) is transformed into a new format (Key_1) using the Kronecker method with the matrix dimension $\sqrt{N''_T} * T_{\max}$. The column-wise Kronecker product is symbolized as $\otimes$. For illustration, let *Key* = 0,2,1. Onto *Key* the row-wise duplication is carried out and the key matrix Key_1 is generated with $\sqrt{N''_T} * T_{\max}$ matrix dimension as shown in Eq. (12). In Eq. (13), the $\sqrt{N''_T}$ is allocated for row matrix and $T_{\max}$ is allocated for column matrix. Using size $\sqrt{N''_T} * T_{\max}$, the reconstruction of *Key* takes place (Balashunmugaraja and Ganeshbabu 2020).

$$Key_1 = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \tag{12}$$

with dimension $N_T * T_{\max}$, the key matrix Key_1 is attained by Khatri–Rao product using the expression shown in Eq. (14).

$$Key_2 = Key_1 \otimes Key_1 \tag{13}$$

$$\text{i.e. } Key_2 = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \tag{14}$$

5.5. Rumbo algorithm

It has been proposed in the literature that hybridizing the two common optimization methods speeds up the convergence of the solutions and, as a result, expands the scope of the search. In this study, we present the RUMBO model, a novel hybrid optimization model that we developed by fusing the RHSO and the MBO, respectively. Generally speaking, the RHSO was created based on the swarms of rock hyraxes' cooperative feeding and food-finding behaviour. This RHSO model has been suggested as one of the best approaches for solving the complex optimization problems. Moreover, with RHSO, the solutions were able to acquire the global search by means of neglecting the local search. On the other hand, the MBO was based on the inspiration acquired from the mating flight of honey-bees MBO is also significant in solving the optimization issues with higher convergence and at the same time, the solutions do not get trapped into the local optimal. Interestingly, on hybridizing these both techniques, the convergence speed of the solutions will automatically increase, and the global search ability of the solutions can be increased. The input to the proposed RUMBO model is the generated key. The steps followed in the proposed RUMBO model are depicted below:

Step 1: Initialize the population (*pop*) of N search agents. The current search agent is denoted as *itr* and the maximal iteration is denoted as $\max^{itr}$

Step 2: The fitness of every N search agent is computed using Eq. (6).

Step 3: Among N search agents, the best one is selected as the leader *Leader*

Step 4: Set $itr = 1$

Step 5: While $itr < \max^{itr}$ do

Step 6: Our contribution resides in this phase. Here, *Leader*'s position is updated based on the newly formulated expression given in Eq. (15)

$$Leader = (2 * rand_1[0,1] - 1) + X_{old}(Leader_{pos}, j) \tag{15}$$

Here, $rand_1[0,1$ is a random number generated between 0 to 1. In addition, X_{old} is the oldest position of the search agent and j points to the "diminution".

Step 7: The other participants in the search space are updated after the leader has been updated. According to the newly proposed expression presented in Eq. (16), this updated mechanism operates. The newly proposed expression is based on the position update of the MBO model. In addition, the levy flight function $Levy(\beta)$ is added. This levy flight (a random walk) is a strategy practiced by search agents to find the optimal solutions within the search space.

$$X(i,j) = \left(-\alpha \times (X_{pos}^{best} - X_{pos}^{worst}) + Levy(\beta)\right) \tag{16}$$

Here, X_{pos}^{best} and X_{pos}^{best} are the best and the worst position of the search agents.

Step 8: The fitness of every N search agent is computed as per Eq. (6).

Step 9: Among N search agents, the best one is selected as the leader *Leader*

Step 10: Update the angle *ang* (a random number in between [0,360]) in every iteration based on the upper *UB* as well as lower bound *LB* of the solutions in the search space.

$$dalta = rans[LB, UB] \tag{17}$$

$$ang = ang + dalta \tag{18}$$

Step 11: Modify the location of the solutions by using the recently proposed updated assessment illustrated in Eq. (21), with levy flight strategy.

$$Leader = \alpha \times (Leader_{pos}^{best} - Leader_{pos}^{worst}) + Levy(\beta) \tag{19}$$

Step 12: t=t+1
Step 13: End while
Step 14: Return the best solution as the leader

6. Illustration of the proposed work

Let there be two hospitals H_1, H_2, two health centers H_3 and H_4 that are connected to the medical blockchain. All these hospitals are said to have their own cloud C_1, C_2, C_3 and C_4, respectively. In each of the hospitals, there is a massive count of patients, and these patients are said to wear a wearable sensor for monitoring their heart rate, blood pressure, sleep pattern as well. In each of the hospitals, there's a countable number of doctors too. Let there be N count of patients $A_1, A_2, \ldots, A_N$ in a hospital H_1, and these patients upload their EHR data onto the cloud C_1 using the IoT. At the same time, the M count patients $B_1, B_2, \ldots, B_M$ in the hospital H_2 have uploaded their EHR onto the corresponding cloud C_2. In addition, there are two health centres H_3 and H_4 with their corresponding cloud storage C_3 and C_4, respectively. In H_3 and H_4, the patients are symbolized as $E_1, E_2, \ldots, E_S$ and $F_1, F_2, \ldots, F_S$. Here, R, S denotes the count of patients in H_3 and H_4, respectively. Let's suppose that the patient E_1, who has been admitted in the emergency unit of the health centre due to COVID-19 has to be monitored continuously by the cardiologist W_1 in the hospital H_1. Since the patient E_1 is being a VIP, there's a huge chance for unauthorized persons to hack his EHR and modify it. Our major contribution resides in preserving the privacy of the patient's EHR.

The patient E_1's sensor recorded data (EHR with heart beat rate, blood pressure level, weight, sleep pattern) are uploaded to the cloud C_1 corresponding to H_3 by using the user node (smartphone). This EHR E_1 includes both the sensitive as well as non-sensitive fields. The E_1 is subjected to the proposed associative rule hiding approach, wherein the associative rules are mined using the proposed improved apiori algorithm, and from which the sensitive fields are identified. Let the sensitive fields be heartbeat rate, and blood pressure level, and the non-sensitive fields are weight and sleep pattern. These sensitive fields have been identified based on the computed support value of the improved apiori algorithm. Then, these identified sensitive fields have been sanitized using the selected optimal key that has been generated with the proposed hybrid algorithm. This sensitive data passes into the blockchain and reaches the hospital H_1, Now, the cardiologist W_1 at the receiver end can only retrieve the EHR E_1 if he/she has the optimally generated key for performing the decryption process. Therefore, the security of the data is ensured, and no unauthorized users can retrieve it.

7. Result and discussion

7.1. Experimental setup

The proposed work has been implemented in PYTHON. The assessment has been done in terms of Convergence Analysis, CPA, key sensitivity, KPA, degree of modification, hiding ratio, information preservation, and privacy as well, as to ensure that the projected model (RUMBO) is superior to the current state-of-the-art methodologies. The traditional models considered in this evaluation are SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019), and RHSO, respectively. 70% of the acquired data are used to train the projected model, which validates the proposed work, and the rest 30% has been utilized for validation. The proposed method has been validated with the data collected from a benchmark dataset corresponding to heart disease

7.2. Dataset description

Heart Disease Data Set: The dataset is downloaded from the link "https://archive.ics.uci.edu/ml/datasets/heart+disease", accessed on 20-11-2021, which contains 76 attributes. By utilizing the subset of 14 of them, the other published experiments are defined. For analysis, the two datasets namely Cleveland and Switzerland datasets are being used. The number of instances of the Cleveland database is 303. The Switzerland database consists of 123 instances.

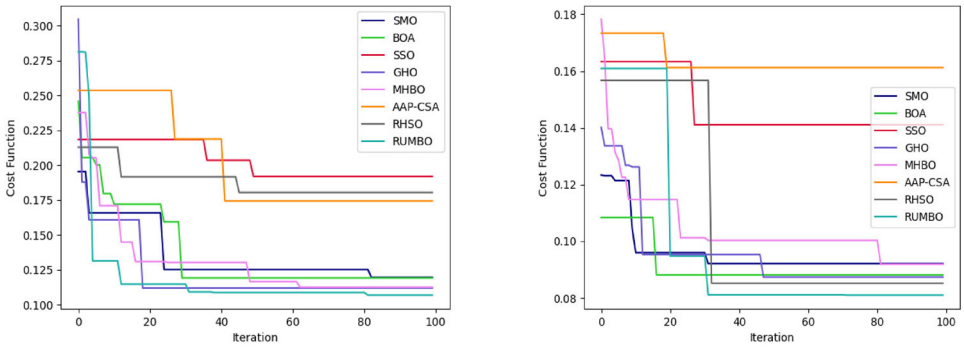

Figure 4. Convergence analysis of RUMBO model for (a) Cleveland and (b) Switzerland.

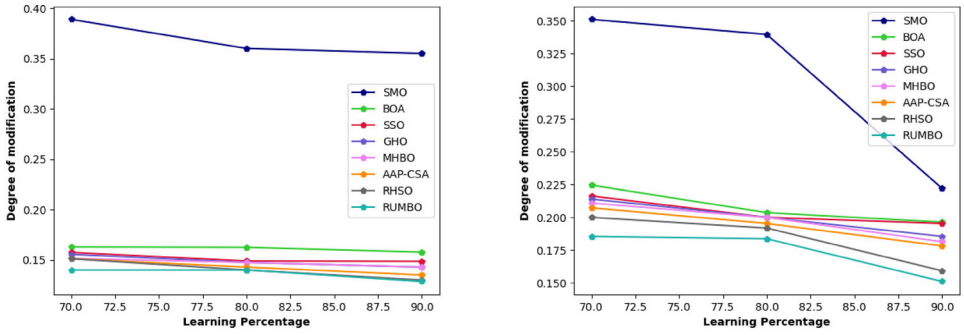

Figure 5. Degree of modification analysis of RUMBO model for (a) Cleveland and (b) Switzerland.

7.3. Convergence analysis

Convergence analysis is performed to check if the presented hybrid optimization model (RUMBO) is efficient over the existing SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019), and RHSO. The convergence analysis is performed by varying the count of iterations. The strategy with the lowest cost function is considered to be the best one because the defined fitness function is a minimizing function. The convergence analysis of 2 datasets that were acquired: Cleveland and Switzerland are shown in Figure 4. On observing the outcomings, the adopted scheme has attained the least cost function regardless of the iterations even with the highest number of iterations, i.e. it can achieve the defined objective function when working with huge iteration counts. Initially, from the Cleveland outcome, the cost function recorded by the adopted work as well as the conventional model is higher at the least iteration count (i.e. at 0–2). At the 3rd iteration count, the proposed work had undergone a steep fall in the cost function and it has reached the cost function as 0.125. The suggested work's cost function was then again reduced, and at the 100th iteration, the least expensive cost function—which is also the most advantageous one—was recorded. It was 0.110. Likewise, the presented work's cost function for the Switzerland dataset is highly applicable for maintaining the system's safety. The suggested approach using the Switzerland dataset first records a cost function of 0.16 at iteration 0. It's evident that the cost function for the projected research has considerably decreased (0.09) by the 20th iteration. Later, as the total number of iterations increased, the recommended work's cost function was reduced. The projected work's cost function, therefore has the lowest value at the 100th iteration, which is 0.08. Since the generated optimal key has achieved the defined objective function with a higher convergence speed than the existing model, it is said to be much more applicable for data privacy preservation.

7.4. Analysis on degree of modification

The determined Euclidean distance between the sensitive database S' that has been sterilized and the actual sensitive database S represents the degree of modification. This gap must be shorter, and as a result, there won't be any difference between the original sensitive database S and the sanitized sensitive database S'. Figure 5 illustrates the level of alteration detected by the proposed approach using the Cleveland and

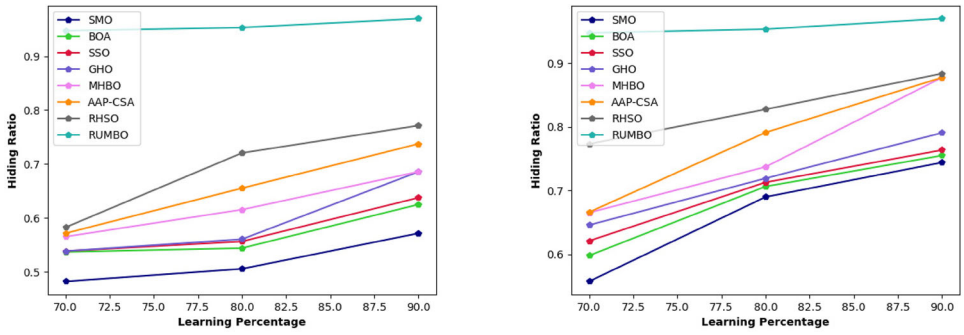

Figure 6. Analysis on hiding ratio of proposed work with (a) Cleveland and (b) Switzerland.

Switzerland datasets. The proposed work has, according to the recorded results, produced the least degree of modification on all three datasets. This is due to the introduction of the improved apiori algorithm for rule mining, wherein the unwanted rules were rejected. Moreover, here the sensitive data has been identified previously by computing the support values of the item sets, and therefore the sensitive data alone has been sanitized and restored. In addition, the obtained optimum key for data sanitization is also strongly convergent. Thus, these factors helped to reduce the degree of alteration. This reveals that the proposed technique is very acceptable for protecting the privacy of medical data because there is a reduced risk of data loss or data manipulation. The projected model's degree of change at the 70th position is 0.14, and this is the best value than SMO = 0.39, BOA = 0.16, SSO = 0.16, GHO = 0.16, MHBO = 0.15, AAP-CSA (Mandala and Rao 2019) = 0.15 and RHSO = 0.15. Also, our present model has reached the lowest Degree of Modification in the Switzerland dataset. This confirms that the projected approach is appropriate to transfer medical data through a blockchain.

7.5. Analysis on hiding ratio

The hiding ratios are employed to confirm if the planned work effectively uses the suggested optimal key to hide the sensitive data. The optimum key used is primarily responsible for this rise in the hiding ratio. Further, the proposed improved apiori algorithm was used to identify the sensitive features and also attained an increase hiding ratio. The results obtained from evaluating the hiding ratios of the proposed study using the Cleveland and Switzerland datasets are illustrated in Figure 6. The learning percentage was varied for this assessment from 70 to 90, correspondingly. For the Cleveland dataset, the hiding ratio of the projected model is 97%, at the 90th learning percentage which is the best score. At the same 90th learning percentage, the hiding ratio of the adopted model for the Switzerland database is 23%, 22%, 21%, 18%, 9.5%, 9.5%, and 8.95 improved over the existing SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019) and RHSO, respectively.

7.6. Analysis on information preservation

Maintaining the information safe while sanitizing the data is essential. The information preservation proportion of the proposed and existing works is calculated for the Cleveland and Switzerland datasets in this research. The values generated are shown in Figure 7. The suggested scheme appears to have achieved the highest information preservation ratio among the three datasets. Only with the specified optimal key, which can't be recovered by unauthorized participants, can the preservation ratio increase. In case of Cleveland, the proposed work has attained the highest Information Preservation ratio as 94%, 96%, and 98% at learning percentage = 70, 80, and 90, respectively. Moreover, at the 90th learning percentage, the projected model has attained the highest Information Preservation ratio of 98%, which is 10%, 9.5%, 9.09%, 7.86%, 7.7%, 6.12%, and 4.07% improved over the existing SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019) and RHSO, respectively. The planned work has the highest ratio for information preservation. As a result, it is claimed that the proposed approach will play a vital role in protecting medical data while it is being transmitted.

7.7. Analysis on privacy

The privacy of the adopted model is evaluated to make sure the effectiveness of protecting sensitive data. For the projected model using the Cleveland and Switzerland datasets, a privacy analysis is conducted. Figure 8 exhibits the results that were obtained. The projected model has achieved the highest privacy for every variation in the learning Percent when observing the acquired results. With the

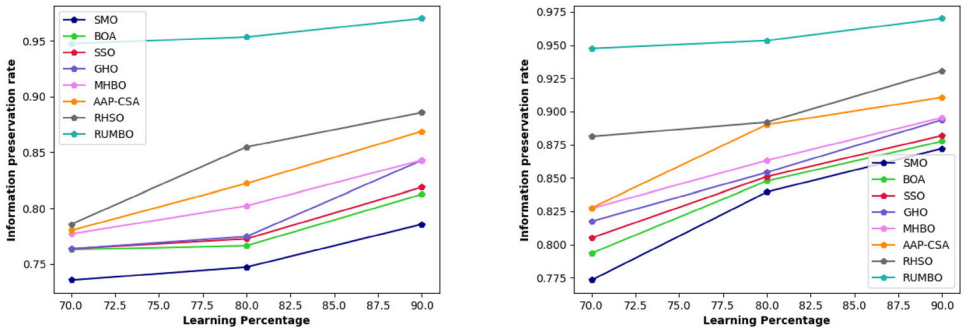

Figure 7. Analysis on information preservation of proposed work with (a) Cleveland and (b) Switzerland.

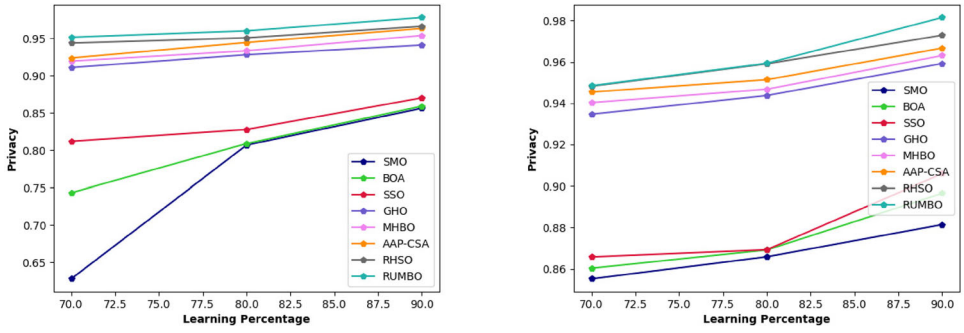

Figure 8. Analysis on privacy of proposed work with (a) Cleveland and (b) Switzerland.

Table 3. CPA analysis of RUMBO with Cleveland dataset.

Learning percentage	SMO	BOA	SSO	GHO	MBBO	AAP-CSA	RHSO	RUMBO
70	0.806154	0.997068	0.99894	0.956669	0.981425	0.972411	0.3411	0.163146
80	0.996087	0.996935	0.996539	0.95906	0.981425	0.996708	0.996915	0.423982
90	0.990362	0.847532	0.317496	0.165144	0.981425	0.169838	0.956258	0.141064

suggested hybrid optimization technique, this improvement in privacy credits only applies to the optimum keys that have been chosen. In the case of Cleveland, the proposed work has attained the highest value of 97% at 90th learning percentage. The privacy of the projected model is 10%, 9.5%, 9.09%, 7.86%, 7.7%, 6.12%, and 4.07% improved over the existing SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019) and RHSO, respectively for Cleveland dataset. Moreover, the privacy of the projected model for the Switzerland dataset is 10%, 9.5%, 9.09%, 7.86%, 7.7%, 6.12%, and 4.07% improved over the existing SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019) and RHSO, respectively at 90th learning percentage. The high privacy value recorded by the proposed work is 98%, and this has been recorded at the 90th learning percentage.

7.8. CPA analysis

The CPA analysis of the proposed work (Tables 3 and 4) is undergone by varying the learning percentage from 70-90. This evaluation has been done for Cleveland and Switzerland datasets. From the results, it is evident that the projected paradigm has the lowest CPA value across all three datasets. This clearly says that the key cannot be hacked by unauthorized users. At 70th learning percentage, the CPA value recorded by the proposed model is 0.163146, which is the least value when compared to SMO= 0.806154, BOA = 0.997068, SSO = 0.99894, GHO = 0.956669, MBBO = 0.981425, AAP-CSA = 0.972411, RHSO = 0.3411.

7.9. Key sensitivity

The key sensitivity analysis has been undergone for Cleveland and Switzerland datasets. the results acquired are shown in Tables 5 and 6.

7.10. KPA analysis

The proposed model is subjected to the KPA analysis, by altering the learning percentage. The datasets from Switzerland and Cleveland have both undergone this

Table 4. CPA analysis of RUMBO with Switzerland dataset.

Learning Percentage	SMO	BOA	SSO	GHO	MBBO	AAP-CSA	RHSO	RUMBO
70	0.998021	0.052751	0.059233	0.272955	0.942932	0.605931	0.050897	0.039659
80	0.134718	0.996308	0.199144	0.998511	0.942932	0.151097	0.112209	0.095377
90	0.176699	0.230913	0.196031	0.232907	0.942932	0.233247	0.253689	0.163969

Table 5. Key sensitivity of RUMBO with Cleveland dataset.

	Key sensitivity
key1	0.612374
key2	0.612372
key3	−0.25
key4	−0.61237
key5	0.408249

Table 6. Key sensitivity of RUMBO with Switzerland dataset.

	Key Sensitivity
key1	0.612374
key2	0.612372
key3	−0.25
key4	−0.61237
key5	0.408249

Table 7. KPA analysis of RUMBO with Cleveland dataset.

Learning Percentage	SMO	BOA	SSO	GHO	MBBO	AAP-CSA (Mandala and Rao 2019)	RHSO	RUMBO
70	0.996363	0.996483	0.953327	0.953576	0.912887	0.996208	0.996887	0.855204
80	0.997249	0.997052	0.955742	0.958087	0.912887	0.996954	0.996258	0.853758
90	0.14114	0.991239	0.853656	0.15501	0.912887	0.990835	0.854434	0.135305

Table 8. KPA analysis of RUMBO with Switzerland dataset.

Learning Percentage	SMO	BOA	SSO	GHO	MBBO	AAP-CSA (Mandala and Rao 2019)	RHSO	RUMBO
70	0.998985	0.028671	0.036339	0.29701	0.935586	0.299013	0.027378	0.023707
80	0.118605	0.993143	0.125442	0.999034	0.935586	0.120027	0.148433	0.111316
90	0.179123	0.167767	0.174986	0.168514	0.935586	0.164614	0.298365	0.158245

examination. Tables 7 and 8 show the results that were obtained, respectively. Analyzing the results reveals that the recommended work achieved higher KPA outcomes. This demonstrates that the predicted model is resistant to KPA attack.

8. Conclusion

This research introduced a novel privacy preservation model in the healthcare sector. The proposed work encapsulates three major phases: Proposed Association Rule hiding, optimal key generation, and blockchain-based data transfer. Initially, the collected raw medical data from the wearable sensors (IoT devices) are passed into the cloud using the user node (smartphone), and it has been stored in the cloud storage. In prior to the data being transferred *via* the blockchain, they are subjected to the Proposed Association Rule hiding phase. In the Proposed Association Rule hiding phase three important stages take place (i) Data Pattern mining with improved apiori algorithm, (ii) sensitive data identification based on the computed support value of the improved apiori algorithm, and (iii) data sanitization and restoration approach. The patterns of input medical data (collected from the user node) are initially mined using the proposed improved apiori algorithm. Subsequently, the sensitive data are identified from the data patterns based on the computed support value of the improved apiori algorithm. Then, the identified sensitive data are transferred to the next phase named, the sanitization phase, where the sensitive data are hidden using the proposed optimal key. This sanitized sensitive data has been subjected to the blockchain for data dissemination. At the receiver end, the sanitized sensitive data are restored using the generated optimal key. Since the optimal key generation plays a significant role in both the data sanitization as well as restoration end; they are optimally generated using the proposed multi-objective hybrid optimization model. Based on the multi-objective function the optimal key has been generated. The proposed hybrid optimization model has been referred to as RHUMBO. This RHUMBO has been the conceptual amalgamation of the standard Marriage in Honey Bee (MBO) and Rock Hyraxes Swarm Optimization (RHSO). Thus, the sensitive data will get transferred in a secure manner. Finally, the efficiency of the projected model has been validated for security. At the same 90^{th} learning percentage, the hiding ratio of the projected model for Switzerland database has been 23%, 22%, 21%, 18%, 9.5%, 9.5%, and 8.95 improved over the existing SMO, BOA, SSO, GHO, MHBO, AAP-CSA (Mandala and Rao 2019) and RHSO, respectively. However, despite adding

some complexity to solutions, the interdependence of computing paradigms is necessary to create efficient privacy-preserving systems. Future studies should focus primarily on smart contracts and mixing tactics, taking into account the integration of privacy protection strategies in these systems. Additionally, this suggested model may be expanded into a framework that can be utilized to develop a range of machine learning schemes that safeguard privacy in numerous encrypted data set components. Blockchain technology may potentially be used to handle smaller-scale financial transactions between patients and healthcare facilities. The remuneration of the services rendered by the providers under the new value-based healthcare models may be based on the level of well-being achieved rather than the quantity. Key management issues are not considered in this research which can be considered for future work.

Nomenclature

RHSO	Rock Hyraxes Swarm Optimization
IoT	Internet Of Things
HW	health wallet
MMSDDF	multi-modal secure data dissemination framework
DLT	Distributed Ledger Technology
RHUMBO	Rock Hyraxes Updated Marriage In Honey Bee Optimization
MBO	Marriage In Honey Bee
NIST	National Institute Of Standards And Technology
KPA	Known Plain Text Attacks
AAP-CSA	Adaptive Awareness Probability-based CSA
CPA	Chosen-Plaintext Attack
CSA	Crow Search Algorithm

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Akkaoui R, Hei X, Cheng W. 2020. EdgeMediChain: a hybrid edge blockchain-based framework for health data exchange. IEEE Access. 8:113467–113486.

Amir Latif RM, Hussain K, Jhanjhi NZ, Nayyar A, Rizwan O. 2020. A remix IDE: smart contract-based framework for the healthcare sector by using Blockchain technology. Multimedia Tools Appl. 81:1–24.

Arul R, Al-Otaibi YD, Alnumay WS, Tariq U, Shoaib U, Piran MD. 2021. Multi-modal secure healthcare data dissemination framework using blockchain in IoMT. Pers Ubiquit Comput; p. 1–13.

Balashunmugaraja B, Ganeshbabu TR. 2020. Optimal key generation for data sanitization and restoration of cloud data: future of financial cyber security. Int J Info Tech Dec Mak. 19(04):987–1013.

Biswas S, Sharif K, Li F, Alam I, Mohanty S. 2020. DAAC: digital asset access control in a unified blockchain based e-health system. IEEE Trans. Big Data. 8(5):1273–1287.

Li P, Xu C, Jin H, Hu C, Luo Y, Cao Y, Mathew J, Ma Y. 2019. ChainSDI: a software-defined infrastructure for regulation-compliant home-based healthcare services secured by blockchains. IEEE Syst J. 14(2):2042–2053.

Chen Y, Ding S, Xu Z, Zheng H, Yang S. 2019. Blockchain-based medical records secure storage and medical service framework. J Med Syst. 43(1):1–9.

Chen Z, Xu W, Wang B, Yu H. 2021. A blockchain-based preserving and sharing system for medical data privacy. Future Gener Comput Syst. 124:338–350.

Daraghmi EY, Daraghmi YA, Yuan SM. 2019. MedChain: a design of blockchain-based system for medical records access and permissions management. IEEE Access. 7: 164595–164613.

Data collected from https://archive.ics.uci.edu/ml/datasets/heart+disease [accessed 2021 Nov 20].

Elhoseny M, Haseeb K, Shah AA, Ahmad I, Jan Z, Alghamdi MI. 2021. IoT solution for AI-enabled privacy-preserving with big data transferring: an application for healthcare using blockchain. Energies. 14(17):5364.

Garg N, Wazid M, Das AK, Singh DP, Rodrigues JJ, Park Y. 2020. BAKMP-IoMT: design of blockchain enabled authenticated key management protocol for internet of medical things deployment. IEEE Access. 8:95956–95977.

Guo R, Shi H, Zheng D, Jing C, Zhuang C, Wang Z. 2019. Flexible and efficient blockchain-based ABE scheme with multi-authority for medical on demand in telemedicine system. IEEE Access. 7:88012–88025.

Hossein KM, Esmaeili ME, Dargahi T, Khonsari A, Conti M. 2021. BCHealth: a novel blockchain-based privacy-preserving architecture for IoT healthcare applications. Comput Commun. 180:31–47. https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/.

Huang H, Zhu P, Xiao F, Sun X, Huang Q. 2020. A blockchain-based scheme for privacy-preserving and secure sharing of medical data. Comput Secur. 99:102010.

Jayasri NP, Aruna R. 2022. Big data analytics in health care by data mining and classification techniques. ICT Express. 8(2):250–257.

Li CT, Shih DH, Wang CC, Chen CL, Lee CC. 2020. A blockchain based data aggregation and group authentication scheme for electronic medical system. IEEE Access. 8:173904–173917.

Liu X, Zhou P, Qiu T, Wu DO. 2020. Blockchain-enabled contextual online learning under local differential privacy for coronary heart disease diagnosis in mobile edge computing. IEEE J Biomed Health Inform. 24(8):2177–2188.

Luong DA, Park JH. 2022. Privacy-preserving blockchain-based healthcare system for IoT devices using zk-SNARK. IEEE Access.

Madine MM, Battah AA, Yaqoob I, Salah K, Jayaraman R, Al-Hammadi Y, Pesic S, Ellahham S. 2020. Blockchain

for giving patients control over their medical records. IEEE Access. 8:193102–193115.
Mandala J, Rao MC. 2019. Privacy preservation of data using crow search with adaptive awareness probability. J Inform Security Appl. 44:157–169.
Ranjan P, Kumar Paul R. 2018. A robust privacy preserving of multiple and binary attribute by using super modularity with perturbation. IJSRCSEIT. 3(8):121–132.
Rathee G, Sharma A, Saini H, Kumar R, Iqbal R. 2020. A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology. Multimed Tools Appl. 79(15-16):9711–9733.
Shailaja GK, Rao CV. 2019. Impact of opposition intensity on improved cuckoo search algorithm for privacy preservation of data. J Network Commun Syst. 2(4):33–41.
Son S, Lee J, Kim M, Yu S, Das AK, Park Y. 2020. Design of secure authentication protocol for cloud-assisted telecare medical information system using blockchain. IEEE Access. 8:192177–192191.
Sreekala M, Varghese P. 2022. An enhanced dynamic scheme for medical image encryption using elliptical curve cryptography. Int J Sci Res Comput Sci Eng. 10(5):1–9.
Sri PA, Bhaskari DL. 2020. Blockchain technology for secure medical data sharing using consensus mechanism. Materials Today: Proceedings.
Stafford TF, Treiblmaier H. 2020. Characteristics of a blockchain ecosystem for secure and sharable electronic medical records. IEEE Trans Eng Manag. 67(4):1340–1362.
Wang H, Wang Q, He D. 2019. Blockchain-based private provable data possession. IEEE Trans Dependable Secure Comput. 18(5):1.
Wang S, Wang J, Wang X, Qiu T, Yuan Y, Ouyang L, Guo Y, Wang FY. 2018. Blockchain-powered parallel healthcare systems based on the ACP approach. IEEE Trans Comput Soc Syst. 5(4):942–950.
Wang Y, Zhang A, Zhang P, Wang H. 2019. Cloud-assisted EHR sharing with security and privacy preservation via consortium blockchain. IEEE Access. 7:136704–136719.
Xia QI, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. 2017. MeDShare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access. 5: 14757–14767.
Yang X, Li T, Pei X, Wen L, Wang C. 2020. Medical data sharing scheme based on attribute cryptosystem and blockchain technology. IEEE Access. 8:45468–45476.
Zerka F, Urovi V, Vaidyanathan A, Barakat S, Leijenaar RT, Walsh S, Gabrani-Juma H, Miraglio B, Woodruff HC, Dumontier M, et al. 2020. Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (C-DistriM). IEEE Access. 8:183939–183951.
Zhang R, Xue R, Liu L. 2021. Security and privacy for healthcare blockchains. IEEE Trans Serv Comput. 24: 2169–2176.
Zhuang Y, Sheets LR, Chen YW, Shae ZY, Tsai JJ, Shyu CR. 2020. A patient-centric health information exchange framework using blockchain technology. IEEE J Biomed Health Inform. 24(8):2169–2176.
Zou R, Lv X, Zhao J. 2021. SPChain: blockchain-based medical data sharing and privacy-preserving eHealth system. Inform Process Manag. 58(4):102604.