SYSTEMATIC REVIEW OF BLOCK CHAIN WITH BIG DATA IN HEALTH CARE

Amrutha.BK
Research Scholar/Computer science, Anna University, Chennai

Abstract. Big data has played a significant influence in practically every domain. Big data in healthcare also covers the collection of patient and clinical records. It is very large or challenging to comprehend using conventional data processing methods. Large data management has a number of challenges for improved service quality, the most important of which are big data privacy and security. The interchange of personal health information by the health care enterprise is a key barrier to system interoperability and patient access to medical records. The decentralised nature and security of block chain technology provide big data services a great chance to grow. Systems built on the block chain have the potential to reduce or eliminate the costs and friction associated with present middlemen. Block chain technologies offer numerous prospects for health care, but they must grow further. It is explored how to handle and share individual patient records with other institutional data. There is an introduction to block chain and big data, as well as information on the drivers pushing their integration and a study of block chain services for big data in the health industry. Various broad topics have described how block chain technology has aided health care systems. This provides a clear depiction of the issues that may arise while combining block chain with health care big data.

INDEX TERMS: Block chain, Big data, Health care, IoT, Security, Privacy.

1 Introduction

Big data is in high demand, and data traffic has grown at an unpredictable rate over the past few years. Big data can be employed in various scientific and engineering fields, including data analytics, operations management, smart cities, etc., due to the lack of a clear definition. The most productive use of big data can lower costs across a range of sectors, including management, manufacturing, media, retail, transportation, and healthcare. Big data is the name given to a new generation of technologies and architectures that have been created with the intention of analyzing massive amounts of data and capturing their main characteristics. Big data is defined as data sets with a high volume and number of dimensions that can't be managed, stored, analyzed, or collected using traditional database methods. There are numerous difficulties and problems involved.

1.1 Big data Management

The task of managing organised and unorganised data presents the largest obstacle in respect to this massive amount of data. Big data cannot be managed using traditional procedures or technologies. This massive amount of data can be comprehended by employing novel fusion algorithms and artificial intelligence (AI) algorithms. Machine learning (ML) techniques such as neural networks and other AI techniques can be used to create automated decision-making. Big data is difficult to manage without the required software and hardware. With the correct archiving and analysis tools, big data information can be leveraged to power critical social infrastructure aspects and services. User-friendly big data visualisation will be a critical component of future societal advancement. Block chain-based systems have emerged as appealing choices for protecting the security and privacy of big data. Block chain mechanisms ensure security for effective data collecting in mobile ad hoc networks and mobile nodes. Because of its interaction with edge computing servers, it enhances data quality and fulfils all functions requested by IoT devices. All of these benefits contribute to the block chain's efficiency in handling big data system security characteristics and enabling all big data services and applications.

1.2 Health sector Information

Health care generates a large amount of data, such as lab test results, computer imaging reports such as MRI, CT scan, X-ray, previous medical reports, medication history, last appointment, and so on. When several patients are treated at the same hospital or medical facility, all of these statistics add up to a lot of data. Data evolution is occurring at rapid speed in the healthcare sector. The massive amount of healthcare data creates a number of issues, including security, privacy, access, and how this data might be available outside of the healthcare delivery firms who have it.

The integrity or redundancy of the data is assured by block chain technology, which also ensures that the data has not been updated or damaged due to a database issue. Furthermore, the implementation of block chain technology will facilitate the transportation of healthcare data among multiple nodes in a network that consume this data. These considerations, however, have an impact on the
cost, value, and quality of the data used. Because only network nodes have access to healthcare data and must first get approval from the patient’s own data, block chain technology can assist maintain the confidentiality and privacy of personal health information.

Sharing health information with other healthcare organisations is a significant impediment to system interoperability and patient access to medical data. Block chain is a new distributed infrastructure that allows for the integration of health records across multiple users and other members. The transparent information recording format of block chain could eliminate numerous data exchange strategies and time-consuming reporting tasks.

2 Methods

A systematic review is carried out based on the chosen reporting items for the associated topic. The main goal was to find original research articles about big data in health care, block chain in big data, safeguarding health records, the disadvantages of traditional goods, and the benefits of distributed ledger systems.

All articles and publications, as well as conference proceedings, that met the following criteria are included. During the process, items up to July 2022 were chosen and included. Excluded any duplicates and opinions that did not have a full-text available. To locate and gather health-related search terms, several literature surveys based on the stated topics were used.

The decision was taken based on the importance of blockchain technology in the big data and healthcare industries. The screening was done in bits and pieces. Initially, duplicate articles were avoided. Second, any article names that bore no resemblance to the target topic were deleted. Third, the summaries of the remaining articles were analysed. At this moment, only unique articles with concepts supported by proofs, technological advances, and so on are considered. The importance of the required information was then confirmed by reading the entire contents of the selected articles.

It was decided to create a uniform data abstraction form. It is peer evaluated to ensure uniformity in data contemplation. The results were aggregated and quantitatively examined. The missing content was removed. Inconsistencies and insufficient uniformity on the underlying block chain mechanisms were revealed based solely on the specified block chain platform, which can be misleading in light of technological improvements. The findings were organised first by big data in the health sector and block chain integration in big data, then by clinical and technical techniques.

As a result of the breadth of big data and block chain, more surveys have been released. Edge computing and 5G systems have both incorporated block chain. Using a historical approach, all of the data was gathered into charts.

3. Discussion

A comprehensive review summarises recent studies leveraging block chain technology in the healthcare sector. As this excerpt emphasises, the major components of a healthcare system are doctors or nurses, clinics, hospitals for the dispensing of drugs and other diagnostic or therapeutic equipment, and the money that goes with the first two. The doctors are professionals in a variety of fields, including psychology, physiotherapy, midwifery, dentistry, and medicine. Experts provide early inspection (for primary care), acute disease requiring experienced specialists (secondary care), sophisticated medical analysis and treatment (tertiary care), and truly excellent diagnostic or surgical techniques (quaternary care).

Health professionals are in charge of handling various types of data at each of these levels, such as the patient’s medical history (which includes details about diagnoses and prescriptions), medical and clinical data (such as details from radiography and laboratory tests), and other private or specific medical information. The majority of these patient medical records were retained in the form of typewritten reports or note transcriptions. Electronic medical records are the most commonly explored subject of block chain research. Hyperledger or Ethereum were used to build the vast majority of block chain platforms.

3.1 Health Care and Big Data

In the healthcare industry, huge data sources include vast databases of patient medical records, clinical reports, hospital records, exam results, and insurance policies. It can be used to forecast death rates, illness prevention, and treatment costs. Big data has evolved into an important database from which data may be acquired for illness control and treatment in many countries. The increased demand for big data analytics in healthcare has numerous benefits, including early disease detection. The term "big data" refers to the massive amounts of data generated when all documents are transformed to digital format, and these data are then combined and analysed by certain technology. This information will be used in the field of healthcare for the. In the field of healthcare, this information will be utilized for the general public (or for a specific person) and used to avoid potential pandemics, treat illnesses, lower expenses, etc.

Big data refers to massive amounts of information that are collected, stored, and evaluated. Traditional information storage systems struggle to manage large amounts of data. Due to significant improvements in medical technology, all medical records have been
transformed to computerised reports, and big data has emerged as a major force in the sector. Many argue that big data accelerates the current pipeline of medical management improvements rather than removing expert workers who report to intellectuals and knowledge specialists.

3.2 Importance of Data Collection in Health Care

As a result of selection and analysis of large data, doctors and health officials may make more informed decisions about medications and maintenance.

Clinicians with access to big data, for example, may be able to detect early warning signs of serious illness before they emerge. Early disease treatment is less difficult and less costly than late disease treatment.

Big data from health reports and Google maps have been used to create authoritative health maps, such as those that show underserved areas. Authorities and providers can utilise this information to find locomotive hospitals and other reserves.

Nursing homes and other health care facilities can use big data to diagnose a complete picture of a patient's history. Big data structures also enable health-care teams to associate data that would otherwise be kept in other clinics, hospitals, and so on. Big data ensures that patient statistics are combined, allowing for active and good contact between patients and the backbone that keeps the patient's comprehensive health records.

3.3 Types of Healthcare Data

Therapeutic reports are one type of healthcare input generated by medical organisations. According to the CDC's National Center for Health Statistics, over 85% of doctors who work in offices already accept electronic medical records.

Another type of patient-centered healthcare data is medical reports.

- Reports on dental and surgical procedures
- Observable information, such as a patient's food;
- Biometrics, such as a patient's blood pressure;
- Livelihood characteristics

Big data in health care can achieve visions at the institutional level also. Hospitals, dispensaries and autonomous providers may trail other forms of healthcare distribution data:

- Agendas of Staffs
- Patient expecting room time
- Medical referrals
- Allowance claim data
- Workplace data for employees, such as the number of patients seen each hour.
- Supply chain metrics, such as those used to order the appropriate quantities of personal data

The Institute of Medicine of the National Academies of Sciences, Engineering, and Medicine coined the term "electronic health records" in 2003 to represent reports gathered for the purpose of improving the healthcare business for the benefit of patients and physicians. Murphy, Hanken, and Waters define electronic health records (EHRs) as "digitised medical reports for patients any data relating to an individual's lifelong physical/mental health condition which resides in computerised system(s) used to capture, transmit, receive, store, retrieve, link, and manipulate multimedia data for the purpose of health and health-related services." EHRs have proposed a variety of conveniences for reviewing healthcare-related data. The fundamental advantage of EHRs is that they provide medical personnel with access to a patient's whole medical records.

Clinical histories, medical diagnoses, recommendations, information on known allergic reactions, demographics, and findings from various clinical laboratory comprise the data. It is possible to avoid ambiguities created by illegible handwriting. Analyzing the past reduces drug-affected reactions by reducing failures in treatment dose and frequency. For vaccines, aberrant laboratory results, cancer screening, and other routine checks, automatic reminders and prompts are set up to notify healthcare professionals. Improved communication makes it easier to connect patients and diverse healthcare specialists. EHRs also provide superior employee healthcare beneficiaries while controlling the rising expenses of health insurance benefits. Last but not least, EHRs can help to reduce problems and misunderstandings with billing and claims. EHRs and the internet, when used together, provide patients with unprecedented access to information.

For handling medical records, information technology is now more important than ever. The development, implementation, and usage of monitoring tools and related software can result in warnings and the broadcast of patient health data to medical specialists.
This device generates a large amount of data, which can be analysed to provide real-time therapeutic or scientific therapy that improves health outcomes.

3.4 Internet of Things

To determine a person's health, bimolecular and scientific datasets must be merged. The Internet of Things is a unique source of this data for medical treatment (IoT). Monitoring systems with sensors and computer chips enable data selection and transfer through the internet, thereby opening up new possibilities. The two types of device technologies are Near Field Communication (NFC) devices and RFID tags and readers. It may be possible to validate critical information that will aid in the advancement of healthcare, transportation, and lifestyle by analysing data collected from these chips or sensors. The Internet of Things (IoT) has emerged as a critical component in the healthcare industry. Because they generate a continuous stream of data, IoT devices are perfect for monitoring people's (or patients') health.

A doctor can use IoT devices to diagnose and track their patients from different places, such as their home or office. The expense of healthcare is decreased since a patient does not require therapy or even a visit to the doctor as a result of early diagnosis and treatment. Furthermore, the healthcare industry employs a variety of clinical tools and technology, such as biosensors and clinical IoT devices. These Internet of Things devices generate a large amount of health-related data. If we integrate this data with additional healthcare information that is now publicly available, such as EMRs or PHRs, we may anticipate a patient's health condition and how it will progress from a subclinical to a pathological state. Better forecasting and inspection are two ways that big data from IoT has proven quite valuable. Information from IoT devices on a broader scale can assist in health monitoring, as well as and also controlling the spreading of diseases.

4. Technologies in Big data

Hadoop and Apache Spark are the two most popular tools for working with enormous amounts of data.

4.1 Hadoop

Large amounts of sophisticated big data cannot be managed by a single memory. It is not a practical method. The most reasonable method for analysis is to spread and process it over multiple nodes at the same time. Thousands of computer units are frequently required to categorise and complete processing in a given amount of time. When working with nodes, one must deal with issues such as clustering algorithms, computing correlation, and breakdown handling. For this purpose, Hadoop is the most appealing distributed open-source application. To handle and output large datasets, Hadoop requires the Map Reduce method. Using the map and reduce operations in the Map Reduces architecture, each logical record in the input is turned into a collection of intermediate key/value pairs. The reduce procedure then collects all values with a common key. It is similar to a multiple master-slave procedure. This can manage system, failure, and data loss issues. Hadoop Distributed File System is the file system segment that requires a model-based, extendable, and active data repository across several cluster nodes (HDFS). Many large corporations, including Yahoo, Facebook, and others, use Hadoop because of the additional capabilities it provides to boost the repository and cleaning segments.

4.2 Apache Shark

Apache Spark is another free and open-source Hadoop alternative since it requires less code and can be seamlessly integrated to generate more sorts of sophisticated analyses. The programming interface of libraries such as spark SQL, MLib, and others contributes to increased developer efficiency. Apache Spark necessitates a substantial quantity of RAM. Because RAM is more expensive than hard drives, Map Reduce is predicted to be more cost effective than Apache Spark for large datasets. Apache Storm was created in a similar fashion to produce a rapid technique for data stream processing. This platform supports the majority of programming languages. It includes fault tolerance and scalability for large-scale data analysis.
5. Issues related to Big data

Big data presents numerous issues because it contains both structured and unstructured data. Key issues to be addressed are privacy, security, standards, and governance.

5.1. Security

Considering big data contains a client’s personal information and medical history, it is critical that the information be protected against hackers, cybercrime, and phishing, as stolen data can fetch a high price on the black market. Data encryption, firewalls, layered anti-virus verification, and communication security are all critical for safeguarding sensitive health information. As data becomes more regional and global, the impact on security and standards gets more convoluted and severe.

5.2. Segmentation of Data

Big data is huge, unstructured, and diverse. Data must be identified and classified before it can be used effectively. It was also necessary to review or consolidate the massive data in order to make it more useful to various people or groups.

5.3. Modeling Data

Data is a valuable resource for modeling and simulation; however, it is critical to identify structures and pool relevant data. It is difficult to see and extract the necessary information without a solid organisational framework.

5.4. Cloud storage

The repository will require cloud storage capacity, as well as an acceptable data transfer speed. Graphic data types such as X-ray, CT, and MRI must be supported by the storage system. If the technology can deliver graphic presentations based on readily available data, doctors should be able to swiftly understand, learn from, and make decisions.

5.5. Miscommunications Gap

One of the most serious problems with big data is the miscommunication or disconnect between users and data scientists. Health data from all laboratories and hospitals must be collected and stored in a single location (big data). Currently, each piece of data is saved separately. It is difficult to construct a more full picture of the patients due to insufficient information gathered. A significant amount of time is lost as a result of the doctor having to start the patient’s history from scratch. Big data can be useful in predicting future medical difficulties, but it can also be dangerous and undercut medical specialists. Patients rely on technology more than they do on consultations.

5.6. Lack of adoption of new technologies

One issue with large data is a lack of data to aid in strategy, policy planning, or judgment. One challenge with big data is the insufficient information to assist strategy, policy planning, or decision-making. Big data cannot create and transmit information without technology.

6. Block chain in Big data

Because of its efficiency in handling massive amounts of data, block chain provides structured data for prediction. It is a type of database that is intended to be accessed only once. As a result, block chain databases are created once and never modified or erased. In the big data process, block chain provides additional data layer, which meets two important Big Data analysis requirements:

1. Because the network architecture prohibits forgery, big data is secure.

2. Block chain-based big data is helpful since it is well-structured, plentiful, and complete, making it a good source for more research.

A block chain is an immutable, public record of transactions stored in a digital ledger that is made up of blocks of connected transactions. Transaction records in the block chain are shared by all network users; there is no centralized authority. It protects data integrity by using a consensus algorithm, which reduces data threats. Every network participant must be validated. The transactional type of data saved in a block chain distributed database (computer file asset) requires 1 KB or less of storage space, and no one else can access it as long as the proprietor has the private keys. The decentralized and immutable structure of block chain allows for secure big data transmissions. Furthermore, it enables secure data interchange from data sources to data sources. Block chain works when five conditions are satisfied:
1. Transactions with several parties lead to change
2. Data kept in a common repository.
3. Parties' confidence in the legal transactions
4. Unreliable truth-finders or ineffective mediators
5. Protection to guarantee total integrity.

7. Healthcare management with block chain apps

The healthcare sector use block chain technology to run a healthcare information system. Because block chain technology is transparent and decentralised, it is ideal for security applications. Block chain safeguards codes that can ensure medical data secrecy while simultaneously being transparent and private. Improved data exchange among healthcare professionals translates in a higher possibility of proper diagnosis, more successful treatments, and an overall improvement in healthcare organisations' ability to offer cost-effective care. Sharing information between patients, doctors, and other healthcare professionals is straightforward and secure because of the technology's decentralised design. Another alternative for maintaining individual EHRs is block chain technology. It's often referred to as "smart contracts" since it allows patients to get token payments in exchange for sharing their health information with service providers and research collaborators. In cases where openness and cooperation are critical, block chain solutions are designed to capture specific transactional data events that are transferred among a network of providers. Because it is based on the patient's medical history, block chain technology makes it simple to track a medicine. This improves drug traceability as it moves through the supply chain and helps to prevent drug simulation.

Implementation also necessitates a block chain protocol and a framework that governs the structure and development of apps. The Ethereum platform, for example, allows users to construct decentralised applications on top of the block chain architecture. It's often referred to as a protocol for both permitted and illegitimate block chain construction. Hyper Ledger, an open source project started by the Linux Foundation, is another platform. It is influenced by the quantity of network users and the range of potential applications.

According to the US Department of Health and Human Services (HHS), block chain technology has the potential to totally alter the healthcare sector. The consortium block chain enables the exchange of national health data. It is a transaction layer that all businesses employ. Through its transaction layer, a collection of identifying information can be accessible. Transparency and automation can also lead to increased productivity and lower administrative costs.

The existing state of medical records is fragmented due to common architectures and safe information flow inside the system. Healthcare stakeholders keep track of and update the patient's common clinical data after each medical operation. This data includes both common information such as the patient's gender and unique information such as their date of birth. Historically, this data has been tracked using a database. Instead, healthcare providers might take it a step further by sending a common set of data from each patient interaction to a worldwide block chain transaction layer.

To encrypt diverse private, public, and sensitive data, sensors and block chain technologies are employed. These files are distributed rather than centralised in storage. When a health care practitioner requests access to a patient's data, the patient must be notified in writing for the request to be processed quickly and the patient to receive the information. Every detail, including an encryption-decryption system, Internet of Things devices, and electronic health records. This block chain technology's secure communication technique remains linked to a wireless sensor network. A private block chain is established using the Ethereum protocol to interface with intelligent objects known as smart contracts. The continued integration of these smart contracts to the healthcare business also includes real-time patient monitoring. Because of the connectivity and authentication provided by the block chain, all medical records are secure.

Additionally, each block chained health record can keep an up-to-date copy of the historical data that makes up the healthcare ledger. Therefore, in order to update a single copy of the blockchain, 51% of network users would have to approve the modification. This increases security and can reduce the likelihood of hostile activities.
8. Management of Personal Health Record (PHR) Data Using Block chain

Wearable sensors and medical IoT devices have recently begun to capture personal life-log data in order to construct personal health records (PHR). Real-time healthcare analytics enabled by artificial intelligence connect patients, doctors, pharmaceutical researchers, and payers (AI).

9. Block chain in insurance for health

Currently, most insurance firms process and store their data in centralised systems. Every phase of the health insurance contracting process involves the involvement of a number of middlemen or third parties. Furthermore, data is shared around various insurance sector participants, making the process time-consuming, arduous, and unproductive. Another issue confronting the insurance industry that interoperability has the potential to solve is block chain. Smart contracts allow for the automated collection of transactional information, contract documents, and other data that may aid administrative operations. Because of blockchain technology, doctors now have access to data on their patients’ health insurance plans. Block chain technology has the ability to speed up the health insurance application process and improve the accuracy of provider directories. As a result, blockchain is the ideal alternative for health insurance.

10. Challenges

Block chain technology still faces various industry-specific difficulties despite its many advantages and prospects

10.1 Storage Cost

Massive amounts of data are generated by medical records and the healthcare business. The block chain design, however, only allows for a limited amount of on-chain data storage. The hashing, decentralised design of a block chain has an unreasonably high storage cost. Similarly, accessing, maintaining, and utilising block chain data can be costly if the amount of data is large. As a result, when creating block chain services, this element must be considered.

10.2 Scalability

Scalability is the best way to boost the adoption of block chains in the healthcare sector. Millions of individuals utilize private clinics, hospitals, public hospitals, health scientific institutions, insurance companies, and individual patients. Block chain technology also requires additional computer power, which increases the energy consumption of network device. The sustainability problem for big data in healthcare must be tackled for block chain to be more generally adopted.

10.3 Aspects of the law:

Another challenge is ensuring that blockchain technology complies with national privacy laws and regulations. Laws governing block chain technology differ by country. It is critical to comprehend data ownership, access privileges, and the distributed storage method employed by block chains. Healthcare groups should make suggestions to authorities and expressly state regulatory requirements.
10.4 Alteration of data:

While the long-term stability of block chain data offers structure, it also makes data updates and revisions unavoidable by eliminating the option of data deletion and tampering. Any alteration requires a new chain, or consensus among all nodes on a new block. Both of these two strategies are extremely expensive and unworkable. In order to reduce the necessity for data change, block chain application development must be done carefully.

10.5 Integrating with existing healthcare systems:

One of the biggest obstacles is integrating blockchain technology with existing systems. To move data to the block chain system, businesses in the healthcare sector must modify their present practices. The smooth integration of block chain technology into the healthcare sector is a significant problem since it necessitates significant changes to the way things are currently done.

11. Conclusion

In a variety of ways, block chain technology can play a key role in the health care sector. Processing times are considerably reduced due to the distributed nodes’ immediate availability when a patient enrolls in the system. Because health providers have access to original, correct data, inaccuracies in medical histories are less likely. Because the data is transparent, patients can get a better second opinion on their diagnosis from a doctor. Stakeholders can collaborate with people all over the world who have similar medical problems because the patients’ information is stored on a block chain network. This will not only enhance their health, but it will also boost their confidence and sense of belonging, giving them even more determination to combat the sickness. Patients will also have complete control over their information and the ability to choose who receives it. This article discusses the use of big data in healthcare as well as the security benefits of block chain. Because block chain ensures the immutability and security of stored health data, healthcare organisations’ big data management may be trusted and dependable. The most important applications of this technology in the realm of healthcare are explored. The problems in integrating block chain with big data in healthcare create issues that must be addressed.

Reference


