Big data in healthcare: Privacy and Security issues

Professor Shahnaz N. Shahbazova[1] , Associate professor Amir Mosavi[2], Kalbiyeva Solmaz Rahim[3], Feyzullayeva Sevinc Yaqub[4]

[1] Department of Digital technologies and applied informatics,
Azerbaijan State University of Economics, UNEC, Baku, Azerbaijan
https://orcid.org/ [0000-0002-9898-6829]
shahbazova@gmail.com
[2]Obuda University, Department of Fon Neuman, Hungary
amir.mosavi@uni-obuda.hu

[3]Azerbaijan University of Architecture and Construction
solmaz.kalbiyeva@azmiu.edu.az
https://orcid.org/ 0000-0003-2470- 9818
[4]Azerbaijan University of Architecture and Construction
sevinj.feyzullayeva@azmiu.edu.az,
https://orcid.org/0000-0001-6155-7265

Abstract: The latest achievements in big data security and privacy in the healthcare industry are discussed in this study. Big data has the ability to dramatically revolutionize patient care and improve health outcomes. However, the collection and analysis of massive amounts of sensitive medical data raises serious privacy and security concerns. In order to identify relevant difficulties and potential solutions, this paper undertakes a comprehensive analysis of existing research publications on big data in healthcare, focusing on privacy and security challenges. The authors investigate various cryptographic algorithms, such as AES, RSA, and SHA-256, that can be used to protect healthcare datasets. They also handle issues like as patient privacy, data breaches, and legal and ethical concerns. Furthermore, the paper proposes a framework for storing and sharing data while preserving privacy. It underscores the importance of implementing strong security measures in healthcare organizations to prevent data breaches and safeguard confidential patient information, all while leveraging big data for research and analysis purposes.

Keywords: healthcare, privacy, security, big data

I. Introduction

The healthcare industry generates vast amounts of data, including electronic health records, medical images, and patient-generated data. The emergence of big data has revolutionized the way healthcare providers collect, store, and analyze this data. Big data encompasses the gathering, handling, and examination of extensive and intricate datasets that surpass the capabilities of conventional data processing methods. The three V's of big data, volume, velocity, and variety, make it challenging to manage and analyze. However, utilization of large-scale data in healthcare holds immense promise for enhancing patient outcomes, diminishing expenses, and elevating the standard of care.

To understand how Big Data can help improve healthcare, we first need to define what it is. Big data is a term used to describe datasets that are very large and complex. These types of datasets require advanced computer infrastructure and analysis skills to get any useful information from them. Techniques like ML and AI play crucial roles as essential tools in the field of Big Data analytics, and they are helpful because they can process large and complex data sets and extract it to valuable information that are difficult to do with traditional statistical methods. The process of analyzing Big Data starts with collecting high-quality data. Healthcare organizations need to make sure that the data they collect is accurate, reliable, and consistent from different sources. During data collection, it is important to clean, validate, and normalize the data to make sure it is free from any errors or inconsistencies. Data cleaning is a process of removing any duplicate or irrelevant data from a dataset. Data Validation process ensures collected data is valid and meets specific standards. As data collected from different sources with different formats, data normalization process is needed for bringing data to standard formats to compare and understand from different sources. Data mining plays an essential role in utilizing the power of big data analytics to enhance and optimize healthcare. It harnesses a combination of AI, ML, statistics, databases, and various other tools. Its primary objective is to extract valuable insights and discern patterns from vast volumes of data. The applications of data mining in the healthcare sector are profound, extending to clinical medicine, biological and biomedical research, the analysis of patients with bilateral implants, as well as the discovery of novel pharmaceutical substances. Big data mining uses several methods and technologies. Fuzzy theory, Bayesian network, next-generation sequencing (NGS) and rough set theory are examples of such methods.

Using, collecting and storing big data in healthcare presents significant privacy and security challenges, as the collection and analysis of sensitive patient data can lead to privacy breaches and cyber attacks. Patient privacy concerns, data breaches, and legal and ethical issues are significant privacy challenges in big data healthcare. Data protection and encryption, access control and authentication, and network security and infrastructure are significant security issues in medical big data. The potential of big data in healthcare is enormous, and it has the potential to help a variety of medical and healthcare tasks, such as population health management, clinical decision support, and disease surveillance. However, the challenges associated with big data in healthcare cannot be ignored. The use of privacy-enhancing technologies, risk assessment and management, and regulatory compliance and standards can help healthcare providers address privacy and security challenges in big data healthcare. Successful implementation of privacy and security measures in big data healthcare can provide valuable insights into best practices for securing patient data. The confidentiality of patient data is a fundamental principle in healthcare. However, putting big data analytics into practice has revealed a reactive, bottom-up, and technology-centric approach, which is insufficient to make sure of the security along with the privacy of health data. This has led to a heightened risk of breaches and other security incidents, which not only undermines patient trust but also threatens the very essence of healthcare delivery. Healthcare organizations are now at a crossroads. While big data analytics in healthcare holds substantial promise, the privacy and security concerns pose significant hurdles.

II. Research Methodology

For this research, we systematically reviewed 17 papers from various sources. We searched papers with “big data”, “healthcare”, “big data analytics”, “privacy” and “big data privacy” keywords. Selected papers written in English and published in recent years. Papers deals with Big Data in healthcare and privacy, security challenges. From that sources, we write discussion about following questions:
What is big data in healthcare and why is it important?

1. What are some of the privacy and security issues that big data in healthcare is bringing up?
2. How can healthcare organizations ensure the protection of patient data while still utilizing big data for research and analysis purposes?
3. What are some of the existing solutions for addressing privacy and security concerns related to big data in healthcare, and what are their strengths and limitations?
4. What are some current research issues and potential future strategies for enhancing big data security and privacy in the healthcare industry?

III. Related works

Yan Luo et al. [14] highlight the challenges associated with sharing medical data securely and privately. There is a rising demand for secure and privacy-preserving ways to exchange medical information because of the prevalence of big data, cloud computing, and the Internet of Things in healthcare. The authors point out that sharing medical information without going against security and privacy laws has for a long time been a difficult issue. The authors analyze and evaluate the latest strategies, aiming to shed light on the advancements made in exchange of medical data that is safe and preserves privacy within the past decade. They divide these methods into permissionless and permissioned blockchain-based methods and consider their benefits and drawbacks Additionally, they talk about prospective areas for study on blockchain-based medical data exchange. In addition to blockchain-based approaches, the authors also review cloud-based approaches and SDN-based methods to safe sharing of medical data. They highlight the importance of access control, data integrity, authenticity, and privacy protection are all incorporated into these methods.
The paper [3] provides a thorough analysis of the security and privacy concerns associated with big data. Akbar Khanan et al. Begin by discussing the various technologies that have led to the production of big data, including smartphones, computers, and social network sensors. Despite its potential benefits, big data is beset by security and privacy issues. The authors highlight some of the major security concerns associated with big data, including issues related to data breaches, unauthorized access, and misuse of personal information. They emphasize that although social media platforms allow users to post data in various formats, the increasing volume can eventually become difficult to manage and regulate. In addition, the authors explore that, to make big data meaningful, a variety of tools and approaches are being deployed, including analytics and data mining. However, these techniques also raise privacy concerns since they involve collecting large amounts of personal information about individuals. The authors discuss how researchers from different domains are attempting to overcome these issues in big data. Also, Salwani Abdullah et al. Provides valuable insights into how researchers are working towards overcoming these challenges through innovative solutions such as advanced encryption techniques and secure cloud

computing platforms. In paper [12], the authors Chauhan et al. Introduce an optimized integrated framework for regulating security and confidentiality in healthcare data through big data analytics. They examine the possibilities of big data analytics in healthcare and its transformative impact on global decision-making by uncovering knowledge trends. The authors emphasize the significance of using big data in healthcare, particularly in improving patient diagnostic systems and ensuring end-user privacy. Their proposed framework that aims to capture big data from diverse sources while prioritizing data privacy and security. The framework is specifically designed to identify hidden patterns that can contribute to clinical decision-making. By capturing data from different heterogeneous resources, the framework enables the discovery of valuable patterns for informed medical decisions. The authors also delve into the practical implementation of the proposed framework. They emphasize the presence of technical challenges resulting from obstacles related to data security and privacy. The authors specifically focus on applying big data analytics in the context of healthcare databases, with a particular emphasis on patients affected by HIV and TB.Liang Qiao et al. [9] provides a technical analysis of the privacy and security risks associated with the use of big data in China's healthcare industry. The authors examine how China's healthcare system is developing in light of big data, and note any privacy or security risks that may arise at different stages of the data life cycle. The paper presents a case study analysis based on data and case data analysis to identify these risk factors. They analyze the number of healthcare facilities, healthcare workers, and insured individuals in China, all of which exhibit an upward tendency year after year. But in recent years, there has also been a pattern of hospitalization bills rising yearly. One of the main concerns discussed in the paper revolves around note that successful analysis methods and technology are lacking. for medical data privacy and security. They also highlight the increasing trend of hospitalization costs in recent years as a potential risk factor. The paper provides further details on these risk factors and explores possible protection measures that can be implemented to mitigate them. Nagalakshmi et al. [10], explores the state-of-the-art security challenges in big data and discusses ways to address them in their paper. The authors focus primarily on the recently suggested anonymization and encryption techniques, their strengths and constraints, and envisaged future directions for studies. In introduction they provide an overview of big data in healthcare and its potential benefits. The authors then discuss the security challenges associated with big data in healthcare systems, citing several sources to support their claims. They highlight the importance of protecting sensitive patient information from unauthorized access or disclosure, as well as ensuring the integrity and availability of this data. One of the key issues discussed in this paper is the burden that inbound and outbound safety solutions can place on healthcare systems. These systems handle communications decryption, network traffic analysis, and information re-encryption. This could represent a major resource load and reduce the effectiveness of these safety solutions by more than 80%. To address these challenges, the authors explore various anonymization and encryption techniques that can be used to protect sensitive patient information while still allowing for meaningful analysis of this data.

- **Challenges**

The use of Big data in healthcare poses new challenges for protecting the privacy and security of patient information. While the implementation of big data analytics promises a myriad of potential benefits, it also reveals substantial hurdles that need to be overcome

to make sure its effective and ethical use. A lot of different types of data are used in healthcare, and this data is often very sensitive. This can make it really hard to keep this data secure and maintain patient privacy. At the core of these issues, we have to remember that keeping patient information confidential and using it in an ethical way is incredibly important. Keeping health data secure is a complex task. It's not just about keeping out people who shouldn't have access. It's also about making sure the data stays accurate, that it's available for those who need it, and that it's safe from other possible security risks. Each of these parts of the issue brings its own challenges and requires different strategies to handle them. Privacy issues often come down to making sure patients have knowledge of how their data is being used and having control over it. With big data becoming more common in healthcare, it's becoming harder to make sure patient information is used in a way that's ethical and responsible. This has led to a lot of discussion about who owns, can access, and can use health data. The Cloud Secure Alliance (CSA) has identified four primary challenges related to big data security, which include infrastructure security, data privacy, data management, and integrity and reactive security. Also, traditional objectives of data security includes ensuring data confidentiality, integrity, and availability [2]. Decentralized computing architecture is the main infrastructure security challenge. The security aspect of decentralized computing architecture involves using network connections for data transfer, communication, and coordination.

- **Methods and technologies**

Currently, a growing number of experts and researchers are looking into technologies or methods to address the issue of privacy security in big data. They are dedicated to finding solutions that can protect sensitive information while still allowing the effective use of big data. Their work is crucial in making sure big data can be used safely and effectively. When patients visit the doctor, the majority of them do not have concerns regarding their safety. They simply anticipate that their information and well-being will be adequately protected through trustworthy safety measures. It is crucial for organizations to implement data security solutions specifically designed for healthcare in order to safeguard valuable assets and comply with healthcare regulations [13]. Various technologies are utilized to ensure the safety and privacy of healthcare data. The following are the most frequently employed techniques:

- **Blockchain**

In 2018, Bhuiyan [15] et al. suggested a system solution that used blockchain technology to address concerns about privacy and security of data in the healthcare sector. A block in a blockchain is a set of information or a transaction, and many such blocks linked together form a blockchain, which grows with increasing transactions. In blockchain, every block is assigned a distinct hash, which functions as a digital fingerprint for the transaction. This hash serves as a unique identifier, allowing for easy identification and verification of the block's integrity within the blockchain network. This proposed system involved various stakeholders, such as healthcare professionals, clinics, hospitals, insurance firms, and patients. Their proposed system utilized a private blockchain solution which aims to ensure the security and privacy of data by using distributed ledgers that are difficult to tamper with. The intention behind this design was to create a more secure and transparent healthcare data management system that would uphold patient privacy while allowing for efficient data handling among the involved parties.

With its decentralized design, blockchain gets rid of the need for a central figure to control health data, which decreases the chance of data breaches and unauthorized access. The unchangeable nature of blockchain guarantees that once information is put on the blockchain, it can't be changed or removed without everyone involved agreeing. This makes it difficult for unauthorized actors to get access to confidential medical data. Blockchain's transparency lets everyone involved in a transaction check the data, making sure it's correct and reliable. Finally, consensus-based verification means that everyone involved in a transaction must agree it's valid before it's recorded on the blockchain. This makes the system even more secure and reliable. There are two main types of approach in blockchain: permissioned blockchain-based solutions and permissionless blockchain-based solutions. Permissionless blockchain-based approaches include some examples such as MedRec. MedRec is a blockchain-based system that decentralizes electronic medical record (EMR) management. It uses three types of Ethereum smart contracts to connect patients' medical data from various healthcare providers, and allows access to third-parties following successful authentication.

- **Authentication**

Authentication privacy tools are created to guarantee that individuals who access systems, applications, or data are securely verified and authenticated, while also safeguarding their privacy. An attack commonly happens when two communicating networks are compromised by a third party, allowing unauthorized access between the information channel and the attacker to gain complete access to the data flow in the communication protocol. In order to mitigate these threats, endpoint authentication procedures are established, incorporating cryptographic protocols.[16] These tools employ various techniques to balance the need for authentication with the need to protect sensitive personal information.

Authentication plays a vital role in organizational operations, including granting access to systems, safeguarding network access, verifying user identities, and ensuring the authenticity of users. [18] Authentication provides many advantages in data security, access control, audit and accountability. Authentication mechanisms, such as username/password combinations, biometric identification, or two-factor authentication, help ensure that only authorized individuals can access sensitive healthcare data. This enhances data security and protects patient privacy. Also, Authentication systems can track and log user access, providing an audit trail of who accessed what data and when. This promotes accountability and helps in investigating any potential security breaches or data misuse. Authentication enforcement in healthcare can face challenges such as identity theft, costs, and resource requirements. User-friendly interfaces and fast login methods can be offered so that authentication processes do not adversely affect the user experience. For example, it can prevent users from logging into multiple applications separately by using technologies such as single sign-on. Second factor authentication options may be offered against identity theft. Second-factor authentication options may be offered to provide additional security, instead of single-factor authentication such as username and password only. For example, methods such as SMS verification codes, verification applications or biometric verification can be used.Such steps help reduce the disadvantages of authentication in healthcare [17].

- ***Multi Factor-Authentication***

Multi-factor authentication (MFA) techniques offer an extra layer of identity protection by requiring users to provide a combination of elements, such as biometric fingerprints, mobile phones, usernames, and passwords, to gain access to a device and authenticate themselves. Biometric tools like fingerprint scanners or retina scanners are often utilized in these solutions to detect unique physical attributes. [11] For example, the U.S. The Department of Health and Human Services (HHS) encourages the use of multi-factor authentication, including biometrics, in healthcare systems to enhance security and protect sensitive patient data. It recognizes biometric factors, such as fingerprints, as strong authentication methods due to their uniqueness and difficulty to replicate. İn addition, many smartphones and wearable devices now incorporate biometric authentication, such as fingerprint scanners or facial recognition, as part of their MFA capabilities. These features enable users to secure their devices and access healthcare-related applications and data using biometric factors in addition to passwords or PINs. These cases highlight the growing adoption and effectiveness of biometric authentication as a factor in multi-factor authentication, particularly in healthcare settings.
Multi-factor authentication (MFA) in healthcare can present privacy disadvantages such as user burden, accessibility challenges, potential for misuse, and data privacy concerns. MFA can impose additional steps and requirements on healthcare professionals, increasing their workload and potentially causing frustration. This may impact workflow efficiency and user satisfaction, which can have repercussions on the quality of patient care. There is also the possibility of false rejections, where authorized users are denied access due to technical issues or authentication inaccuracies, resulting in delays in accessing critical patient information and systems. Moreover, MFA methods can be susceptible to abuse or unauthorized access, especially if biometric data is stolen or compromised, allowing malicious individuals to gain unauthorized entry into healthcare systems and patient records. It is worth noting, however, that these disadvantages can be addressed through careful implementation, user education, and continuous monitoring and improvement of MFA systems in healthcare environments. Healthcare Information and Management Systems Society (HIMSS) shows that MFA adoption in the healthcare industry is an effective way to increase security measures and protect sensitive health information. MFA not only enhances security but also reduces the risk of personal information exposure by combining multiple factors. It can also serve as a gateway to a password-free authentication realm [13].

- ***End-to end authentication***

End-to-End Authentication is a privacy tool that guarantees secure and confidential communication between two parties by verifying the identities of both ends of the communication channel. It offers protection against unauthorized access, tampering, and interception of sensitive data. Public Key Infrastructure (PKI) serves as the foundation for end-to-end authentication. PKI verifies users, systems, and gadgets that use digital certificates, eliminating the requirement for tokens, complex regulations governing passwords or user-initiated variables. An example of successful end-to-end authentication in the healthcare industry is the Electronic Health Records (EHR) System. Healthcare organizations can use end-to-end authentication to securely manage patient data and limit access. In this system, healthcare professionals log in with personal authentication information to access patients' electronic health records. Thus, a database that can only be accessed and protected by authorized personnel is created. This decentralized approach allows authentication to occur seamlessly between various

systems. In the field of healthcare, the successful implementation of end-to-end authentication relies on the collaboration of makers of devices, medical facilities, insurance providers, software vendors, and security providers, as they share the responsibility for its effectiveness. End-to-end authentication in healthcare can impose additional steps and authentication requirements on healthcare professionals, resulting in increased workload and potential user frustration. This can lead to reduced efficiency and a reluctance to consistently follow the authentication process. Also , there are privacy disadvantages associated with end-to-end authentication, such as accessibility challenges, complexity, and the risk of data exposure. During the transmission of authentication credentials and sensitive data across various systems and networks, there is a possibility of interception or unauthorized access, compromising patient privacy. To mitigate these privacy concerns, healthcare organizations should meticulously design and implement end-to-end authentication systems that prioritize privacy and security. This entails incorporating robust encryption mechanisms, conducting regular system audits, providing user training and awareness programs, and establishing ongoing monitoring procedures to identify and address vulnerabilities and risks. End-to-End Authentication ensures that only the intended recipient possesses the capability to decrypt and access the transmitted data, thus preserving its privacy and confidentiality [19].

- **Encryption**

Privacy tools based on encryption play a crucial role in safeguarding sensitive data by transforming it into a secure and incomprehensible format. These tools utilize cryptographic algorithms to scramble the data, rendering it indecipherable to unauthorized individuals. Efficiently encrypting data can effectively decrease the interception of packets and mitigate threats. Additionally, it is advisable to minimize the amount of keys each node holds to lessen the danger of privacy and security breaches [1]. Encrypting data can hinder data interoperability between different healthcare systems or organizations. If encryption methods and keys are not standardized or compatible, it can be difficult to safely exchange and access encrypted data. Encrypted data can pose challenges for data recovery and disaster recovery processes. For example, If encryption keys are lost or damaged, it can be difficult or impossible to recover encrypted data, resulting in potential data loss. To address these encryption privacy disadvantages, healthcare organizations need to carefully implement encryption. For the sake of the security of their IT infrastructure, healthcare-focused IT organizations must employ encryption techniques in two specific ways. Firstly, they need to protect the information transmitted across networks. Secondly, they must secure the network infrastructure that stores and disseminates sensitive data. Encryption is a method that involves converting information into an unreadable form, serving as a protective measure. It adheres to a methodical encoding process that can only be undone with the right decryption key [20].

- ***Symmetric Encryption***

Symmetric encryption calculations involve the use of identical cryptographic keys for both encrypting plaintext and decrypting ciphered text. These keys can be identical or can be easily transformed to switch between encryption and decryption.
One approach to symmetric encryption is *Attribute-Based Encryption*. In this approach, key generation for users is managed by a discrete key generation specialist, while user

attributes are handled by a property management expert. This proposed approach helps address the significant computational cost associated with using symmetric encryption algorithms to ensure the privacy of multimedia big data in the Internet of Things (IoT). Attribute-Based Encryption consists of four main components: a trusted authority, a cloud service provider, a data owner, and a data client. The trusted authority generates the necessary key materials for the system by establishing public constraints and the master private key. The cloud service provider serves as the data storage provider for users, storing the encrypted data content provided by the data owner. The data owner is a user who wants to anonymously upload their encrypted data content to the storage system. The data client is a third-party storage subscriber that sends queries to the service provider for encrypted data in the storage system using an alias representing the data owner. Another approach to symmetric encryption is *Attribute-Based Encryption Approach.* In the Attribute-Based Encryption Approach proposed by Aljawa, a reliable encryption method is designed for encrypting multimedia big data in the Internet of Things. It utilizes the Feistel Encryption Plan, Advanced Encryption Standard (AES), and genetic algorithms. The system framework comprises three main components: the Feistel network, AES with S-box, and genetic algorithm. The Feistel network divides the input into small blocks and performs a series of permutation and rotation operations on them. AES with S-box takes the plaintext and computed key as input, and through the Substitution-Permutation Network (SPN), it runs ten encryption cycles to produce the encrypted content. The genetic algorithm incorporates two stages of operations: crossover and mutation. In crossover, parts of the encrypted text and encryption key are switched, while in mutation, an arbitrary selected bit is reserved in both the encrypted text and encryption key. [4]

- ***Tokenization***

Tokenization is a privacy tool widely used in the healthcare industry to protect sensitive data and maintain patient privacy. In healthcare, tokenization is primarily employed to secure Personally Identifiable Information (PII), the names of the patients, social security numbers, and medical record numbers, and insurance details for example. The original data is securely stored in a separate location, while the tokens are used for various purposes within the healthcare ecosystem. These tokens can be utilized in electronic health records (EHRs), medical billing systems, research databases, and other healthcare applications where data is shared and processed. Many end users can use tokenization as a constant value. This eliminates the need for users to constantly decrypt and re-encrypt data in order to interact with the tokenized data [10]. By implementing tokenization in healthcare, several benefits can be achieved. Firstly, it helps protect patient data from unauthorized access and potential breaches. Even if a token is intercepted or accessed by an unauthorized party, it does not reveal any sensitive information or allow the reconstruction of the original data. Secondly, tokenization reduces the complexity of observing healthcare laws like the Health Insurance Portability and Accountability Act (HIPAA). By substituting sensitive data with tokens, the scope of HIPAA compliance audits can be reduced, as the protected health information (PHI) is not directly stored or transmitted. The use of data tokenization requires staff instruction on how to handle and utilize tokenized data. Through training and regular use of tokens, employees can develop a stronger understanding of the importance of data protection. This knowledge can help them identify control weaknesses that may exist in other areas of the organization [5]. Tokenization methods and formats can vary between different systems

or organizations, which can present interoperability challenges when exchanging tokenized data. Establishing common standards and protocols for tokenization is crucial for seamless data exchange. Against the challenges, common standards and protocols for tokenization methods in the health sector should be established. This ensures that data can be exchanged seamlessly between different systems and organizations. For example, using a compliant tokenization method that follows healthcare industry standards can facilitate data sharing. In addition, cooperation and regular communication between different systems or organizations is important. Tokenization methods and practices need to be continuously evaluated and updated. Collaboration should be made through regular meetings, working groups or standardization bodies to resolve cross-system compatibility issues. While tokenization provides an additional layer of security, it should be implemented in conjunction with other security measures and best practices to provide comprehensive data protection in healthcare environments.

- ***Data masking***

Data masking is a privacy tool widely used in the healthcare industry to protect sensitive information by replacing real data with fictional or modified data that retains the characteristics and format of the original information. The purpose of data masking is to ensure data privacy while still allowing organizations to use and analyze the data for various purposes, such as software development, testing, and research. By preserving total and average values within a data collection without changing all of the individual data pieces, masking is effective in maintaining aggregate values across a database as a whole [6].

Data masking techniques involve substituting original record values with meaningful values to conceal sensitive information when it is shared outside of a controlled environment, while still ensuring the integrity of the data relationships. The purpose of data masking is twofold: to protect the data by preventing the identification of individual records and to enable its use for research and analysis purposes. By modifying the actual values of specific records while maintaining the total integrity of the data, data masking serves as an effective measure in mitigating the risk of unauthorized data exposure by both internal and external parties. It is considered a best practice for enhancing the security of production databases. [7]

Data masking techniques alter the original data values, which can have implications for the usefulness and applicability of the data in analysis and research. The masked data may not accurately represent the true attributes or associations found in the original data, which can limit its suitability for certain types of analysis and decision-making processes. Moreover, data masking can create challenges when attempting to link or integrate data from various systems or databases. The alterations made during the masking process can hinder the ability to establish connections or perform data linkage, restricting the comprehensive analysis and understanding of patient health information. Additionally, data masking may not completely mitigate the risks associated with insider threats. Despite the protective measures in place, authorized users with access to both masked and original data may still have the potential to discern sensitive information or reverse-engineer it. Unauthorized data access or misuse by insiders can compromise the privacy and security objectives of data masking. To address these privacy concerns, healthcare organizations should thoroughly evaluate their data masking strategies and consider implementing complementary measures such as access controls, encryption, and data governance frameworks. Regular monitoring, auditing, and evaluation of data masking

Researches	Method	Advantages	Disadvantages
[13] Bhuiyan et al.	Suggested Three layered framework based on blockchain technology for managing medical data and secure sharing.	Increased security, efficiency, secure sharing, patient ownership over their data.	Complex implementation and limited scalability.
Yue et al. [14]	Propose a blockchain-based, access control policy-equipped healthcare data gateway with a focus on the user's needs.	Making sure patients have ownership and control over their medical data, encouraging safe medical data sharing, and protecting patient privacy.	Study does not include enough details about how a service makes sure that during computation it does not have access to the content of raw medical data.
[10] Nagalakshimi et al.	Using authentication for protection of sensitive patient information	Enhanced security, protection against unauthorized access, and safeguarding patient data.	Implementation challenges, costs and resource requirements and risk of credential theft
[11] Gaik-Ye et al.	Using Symmetric, asymmetric, hashing and attribute-based encryption methods to protect sensitive patient data from unauthorized access or breaches.	Ensures the protection of sensitive patient information, preventing unauthorized access and maintaining the confidentiality of patient data. Additionally, it mitigates the potential risks associated with data breaches and cyber-attacks.	Methods has the potential to decrease the speed of data processing, thereby impacting the overall efficiency of healthcare services. Additionally, encryption can pose challenges when it comes to sharing data among different healthcare providers.
[15] Luvai et al.	Using data masking for replacing sensitive data	Enhanced data protection, reduced risk of data exposure, and compliance with privacy regulations	Limited data utility, data linkage challenges and complexity
[19] Kan et al.	Encryption operation based on cyclic shift and XOR operation	Implementing RFID technology in the medical monitoring system is a relatively straightforward process that can be executed efficiently, ensuring a fundamental level of data transmission security. Furthermore, it can be effectively utilized on devices with limited resources, including IoT devices.	It is susceptible to known plaintext attacks, wherein an attacker can exploit the presence of known plaintext to infer the encryption key. Additionally, it lacks authentication or integrity checking mechanisms. Moreover, it does not offer the same level of security as more intricate encryption techniques like AES or RSA.

[21] Mohammad et al.	Anonymity and traceability blockchain based solution	The implementation of blockchain technology can enhance the efficiency of various processes, such as managing medical records, processing insurance claims, and monitoring the drug supply chain. It also empowers patients by granting them increased control over their personal health data, enabling them to determine who can access it.	The protocol utilizes sophisticated encryption and decryption methods, which can potentially result in interoperability challenges among various healthcare systems.
[20] Chinmay et al.	Homomorphic Encryption on IoT-based healthcare	The protocol aids in safeguarding sensitive healthcare information against unauthorized access and cyber threats, while also demonstrating efficiency and speed, making it advantageous in time-critical healthcare environments. Furthermore, a lightweight protocol is easier to implement and manage, leading to time and resource savings.	Less complex protocols may lack the same extent of functionality found in more intricate protocols, potentially restricting their suitability for specific healthcare applications. Additionally, they might exhibit increased vulnerability to certain types of attacks or exploits.
[10] Reddy et al.	FTP and RTT tokenization methods using for substitution of sensitive data with unique tokens	By mitigating the likelihood of data breaches and identity theft, it aids in minimizing the risk associated with such incidents. Additionally, it assists healthcare organizations in adhering to regulations such as HIPAA by safeguarding patient privacy.	If tokens are inadequately secured or managed, they may remain susceptible to attacks. Furthermore, the implementation of a tokenization system can be intricate and demanding, necessitating substantial resources and staff training.

techniques are crucial for ensuring ongoing compliance and effectiveness [8]. Data masking techniques can be applied to various types of sensitive data in healthcare, including patient demographics, medical records, billing information, and genetic data. The masked data is typically used for non-production environments, such as development, testing, and training, where the original data is not necessary but realistic data is required. Masking methods provide true shielding protection and protect against unauthorized access. It provides a secure environment for masking, data analysis and research work. Masking protections, protection analysis and data mining can be done while protecting. Also, Data masking ensures the secure sharing of confidential information. By masking sensitive data, access to real values is limited and security is

ensured in data transmission. In the table presented below, we will conduct a comparison of these methods and outline their respective advantages and disadvantages:

III. Conclusion

In this section, we will examine research questions based on papers that have been previously discussed. Our goal is to find answers and deepen our understanding of the subject matter.

Q1: What is big data in healthcare and why is it important?

In the context of healthcare, big data is an extensive and complicated datasets that require advanced computer infrastructure and analysis skills to extract valuable information from them. It is important because it has the potential to improve healthcare outcomes by providing insights into disease prevention, diagnosis, treatment, and public health management. Analytics of big data can be used to spot patterns and trends that may not be apparent with traditional statistical methods, leading to more personalized and effective healthcare interventions.

Q2: What are some of the privacy and security issues that big data in healthcare is bringing up?

Our research identifies several issues with privacy and security that big data in healthcare raises. These include:

1. Patient privacy concerns: The collection and analysis of sensitive patient data can lead to privacy breaches and cyber attacks.
2. Data breaches: Healthcare organizations are vulnerable to data breaches, which might result in the exposure of confidential patient data.
3. Legal and ethical issues: Big data use in healthcare raises moral and legal concerns about patient consent, ownership of data, and the potential for discrimination.
4. Data protection and encryption: Healthcare organizations must ensure that patient data is protected through encryption and other security measures.
5. Access control and authentication: Access controls must be implemented by healthcare institutions to guarantee that only authorized individuals have access to patient data.
6. Network security and infrastructure: Healthcare organizations must ensure that their networks are secure from external threats such as hacking or malware attacks.

Q3: How can healthcare organizations ensure the protection of patient data while still utilizing big data for research and analysis purposes?

By implementing a comprehensive range of security measures and protocols, healthcare organizations can strike a delicate balance between safeguarding patient data and harnessing the immense promise of big data analytics in the healthcare industry.

Q4: What are some of the existing solutions for addressing privacy and security concerns related to big data in healthcare, and what are their strengths and limitations?

Research materials discuss several existing solutions for addressing privacy and security concerns related to big data in healthcare. These include:

1. Anonymization: Anonymization techniques can be used to remove identifying information from patient data, while still allowing for analysis. The strength of this approach is that it can protect patient privacy, but the limitation is that it may not always be effective in preventing re-identification.
2. Encryption: Encryption can be used to protect patient data both in transit and at rest. The strength of this approach is that it can provide strong protection for patient data, but the limitation is that it can be difficult to implement and manage.
3. Access controls: Access controls may be put in place to guarantee that only individuals with the proper authorization can access patient data. The strength of this approach is that it can help prevent unauthorized access, but the limitation is that it may not always be effective in preventing insider threats.
4. Data governance: Policies and procedures for the gathering, keeping, and utilization of healthcare data can support ensuring the security of patient data. The strength of this approach is that it provides a framework for managing patient data, but the limitation is that it may not always be followed or enforced.
5. Risk assessments: Regular risk assessments can help identify potential vulnerabilities in systems and processes related to big data in healthcare. The strength of this approach is that it helps organizations proactively identify and address security risks, but the limitation is that it may not always capture all potential risks.

Overall, each solution has its own strengths and limitations, and healthcare organizations must carefully consider which solutions are most appropriate for their specific needs and circumstances.

References

[1] Gou, X., & Xu, Z. (2022). An overview of Big Data in Healthcare: multiple angle analyses. Journal of Smart Environments and Green Computing.

[2] Wu, S. M., Guo, D., Wu, Y. J., & Wu, Y. C. (2018, November 30). Future Development of Taiwan's Smart Cities from an Information Security Perspective. MDPI.

[3] Khanan, A., Abdullah, S., Mohamed, A. H. H. M., Mehmood, A., & Ariffin, K. A. Z. (2019). Big Data Security and Privacy Concerns: A Review. Smart Technologies and Innovation for a Sustainable Future, 55–61.

[4] Vijayan, V., Connolly, J. P., Condell, J., McKelvey, N., & Gardiner, P. (2021, August 19). Review of Wearable Devices and Data Collection Considerations for Connected Health. Sensors, 21(16), 5589.

[5] Upadhyay, S., & Hu, H. F. (2022, January). A Qualitative Analysis of the Impact of Electronic Health Records (EHR) on Healthcare Quality and Safety: Clinicians' Lived Experiences. Health Services Insights, 15, 117863292110707.

[6] Cruz Rivera, S., McMullan, C., Jones, L., Kyte, D., Slade, A., & Calvert, M. (2020, July 2). The impact of patient-reported outcome data from clinical trials: perspectives from international stakeholders. Journal of Patient-Reported Outcomes, 4(1).

[7] Deshpande, P., Sudeepthi, B., Rajan, S., & Abdul Nazir, C. (2011). Patient-reported outcomes: A new era in clinical research. Perspectives in Clinical Research, 2(4), 137.

[9] Lv, Z., & Qiao, L. (2020, August). Analysis of healthcare big data. Future Generation Computer Systems, 109, 103–110.

[10] Nagalakshmi, N., Babu, G. A., Reddy, D. K., & Ashalatha, T. (2019, October 30). Security Challenges Associated with Big Data in Health Care System. International Journal of Engineering and Advanced Technology, 9(1), 4057–4060.
[11] Enhancing Privacy for Big Data in Healthcare Domain based on Cryptographic and Decentralized Technology Methods. (2019, October 22). International Journal of Recent Technology and Engineering, 8(3S), 129–135.
[12] Chauhan, R., Kaur, H., & Chang, V. (2020, February 19). An Optimized Integrated Framework of Big Data Analytics Managing Security and Privacy in Healthcare Data. Wireless Personal Communications, 117(1), 87–108.
[13] Bhuiyan, M. Z. A., Zaman, A., Wang, T., Wang, G., Tao, H., & Hassan, M. M. (2018, May 12). Blockchain and Big Data to Transform the Healthcare. Proceedings of the International Conference on Data Processing and Applications.
[14] Jin, H., Luo, Y., Li, P., & Mathew, J. (2019). A Review of Secure and Privacy-Preserving Medical Data Sharing. IEEE Access, 7, 61656–61669.
[15] Motiwalla, L., & Li, X. B. (2013). Developing privacy solutions for sharing and analysing healthcare data. International Journal of Business Information Systems, 13(2), 199.
[16] Zhao, H., Zhang, Y., Peng, Y., & Xu, R. (2017, March). Lightweight Backup and Efficient Recovery Scheme for Health Blockchain Keys. 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS).
[17] Hardjono, T., & Pentland, A. (2019). Verifiable Anonymous Identities and Access Control in Permissioned Blockchains. ResearchGate. https://www.researchgate.net/publication/331701045_Verifiable_Anonymous_Identities_and_Access_Control_in_Permissioned_Blockchains
[18] Soni, M., Barot, Y., & Gomathi, S. (2021, July).Privacy in Preprocessing of Healthcare Data. Journal of Cybersecurity and Information Management (JCIM), 7(1), 13–25.
[19] Wang, X., Fan, K., Yang, K., Cheng, X., Dong, Q., Li, H., & Yang, Y. (2022, March). A new RFID ultra-lightweight authentication protocol for medical privacy protection in smart living. Computer Communications, 186, 121–132.
[20] Othman, S. B., Almalki, F. A., Chakraborty, C., & Sakli, H. (2022, July). Privacy-preserving aware data aggregation for IoT-based healthcare with green computing technologies. Computers and Electrical Engineering, 101, 108025.
[21] Garg, N., Wazid, M., Das, A. K., Singh, D. P., Rodrigues, J. J. P. C., & Park, Y. (2020). BAKMP-IoMT: Design of Blockchain Enabled Authenticated Key Management Protocol for Internet of Medical Things Deployment. IEEE Access, 8, 95956–95977.