

Medical Data Security in the Context of Big Data

Jingyi Li *

School of Big Data, Zhuhai College of Science and Technology, Zhuhai, China

* Corresponding author Email: jingyi.li3@icloud.com

Abstract: With the rapid development of information technology and big data technology, the medical industry is undergoing a profound transformation. Massive medical data shows great potential in clinical diagnosis, personalized treatment, disease prediction, and more. However, medical data contains a large amount of sensitive information, such as patient identities, medical histories, diagnoses, and treatment records. The leakage of this information could pose serious threats to personal privacy. Therefore, this paper uses literature review and comparative analysis to examine research results on medical data security technologies and risk assessments both domestically and internationally, and proposes suggestions and strategies to enhance medical data security. This study shows that although medical data security faces many challenges, advanced technologies and effective measures can significantly improve data security and privacy protection.

Keywords: Data security; Privacy protection; Medical privacy; Risk assessment; Privacy protection.

1. Research background

With the rapid development of information technology and the internet, the medical industry is experiencing a profound digital transformation. The generation and widespread application of large amounts of medical data make medical information an important component of big data. Medical data includes sensitive information such as patients' personal information, medical records, diagnoses, treatment plans, and medication usage records. The aggregation and analysis of this data provide significant support for medical research, disease prevention, and personalized treatment, enhancing the quality of medical services and modernizing the healthcare system.

However, the sensitivity and privacy of medical data also make it a focal and challenging area for data security and privacy protection. Frequent data breaches not only threaten personal privacy but also could lead to medical disputes and social trust crises. As the demand for cross-institutional and cross-platform sharing of medical data increases, the security issues during data transmission and storage become more prominent.

In this context, how to effectively protect the security of medical data and ensure its privacy during generation, storage, transmission, and usage has become a hot topic of common concern in academia and industry. Scholars domestically and abroad have conducted extensive research on medical data security, proposing various technologies and methods such as data encryption, anonymization, access control, blockchain technology, artificial intelligence, and hybrid encryption methods. These technologies have improved the security of medical data to some extent, but further exploration and improvement are still needed. Moreover, domestic scholars have made significant achievements in medical data risk assessment. This study aims to systematically review existing research results in the field of medical data security, analyze the advantages and disadvantages of different technologies, and explore how to improve the security and privacy protection of medical data through the improvement of existing technologies and the application of emerging technologies. This research not only has important academic significance but also provides theoretical support and

practical guidance for the actual management of medical data, thereby promoting the healthy development of the medical industry.

2. Healthcare data security implications

2.1. Protection of patient privacy

Medical data contains highly sensitive personal information such as patient identities, medical histories, diagnoses, and treatment records. Ensuring data security is crucial to protecting patient privacy, preventing unauthorized access, and avoiding potential misuse of this information.

2.1.1. Trust in healthcare systems

Secure medical data is essential for maintaining trust between patients and healthcare providers. Patients are more likely to share accurate and complete information when they trust that their data will be handled securely. This trust is vital for effective diagnosis and treatment.

2.1.2. Enhancement of healthcare services

By protecting medical data, healthcare organizations can improve the quality of their services. Secure data enables better decision-making, efficient patient management, and the implementation of advanced technologies like AI and machine learning in healthcare.

2.1.3. Support of medical research

Medical data security is also crucial for research purposes. Researchers need access to large datasets to identify trends, develop new treatments, and improve healthcare outcomes. Secure data handling ensures that research can proceed without compromising patient privacy.

3. Current research status

3.1. Domestic research status

Zhang Gen et al. (2021) proposed a method for measuring the risk tolerance of medical privacy big data leakage through an improved BDM mechanism and constructed a PNN medical privacy big data leakage behavior model. The study found that the risk of leakage is higher in the storage and application modules of medical privacy big data, with a

higher tolerance for data leakage used for scientific research purposes, but a generally low tolerance for patient data privacy leakage. This method better evaluates the leakage risk of medical privacy big data and helps formulate more effective risk management strategies.

Shang Jingwei (2020) established a medical data privacy leakage behavior model based on the privacy Petri net model, quantitatively analyzing the process and risk of data privacy leakage. The study shows that data privacy leakage is most likely to occur during the application phase and has the most serious impact after leakage. This research provides theoretical support for understanding the mechanism of medical data privacy leakage and methods for preventing and controlling data leakage risks.

Research on medical data security technologies, Kang Haiyan et al. (2021) proposed an enhanced hybrid encryption method based on IBDES and EPNRSA for the secure storage of medical data. The study shows that this method performs well in improving data storage security and effectively prevents data leakage and tampering during storage.

Liao Yijie et al. (2020) suggested strengthening data anonymization, data encryption, data mining protection, and data access control technologies to enhance medical data management capabilities. The study shows that these technologies have significant effects on protecting medical data privacy. A comprehensive security protection system can be formed by combining different technologies.

3.2. Foreign research status

Shari et al. proposed a blockchain-powered decentralized data dissemination scheme combining certificate-based signcryption and proxy re-encryption (Sign-Proxy) with an improved delegated proof-of-stake (DPoS) consensus protocol, enhancing medical data security.

M et al. proposed a new method based on the Coati optimization algorithm (COA), demonstrating through experiments that this method has a better effect on data security protection.

Sworna et al. proposed the DKS-CWH algorithm based on dual kernel support vector (DKS) and crossbreed wildebeest optimization algorithm, with experimental results showing an accuracy of up to 97%.

Liu et al. proposed an efficient DNA homomorphic encryption (DNA-HE) algorithm for secure cloud storage of medical data, finding that this technology requires less encryption time compared to other techniques.

3.3. Existing key technologies

3.3.1. Anonymization

Data anonymization involves modifying personal data to prevent the identification of individuals. This is crucial for sharing medical data for research purposes without compromising patient privacy. Techniques include data masking, pseudonymization, and differential privacy. Anonymization allows researchers to analyze medical data while adhering to privacy regulations.

3.3.2. Secure Multi-Party Computation (SMPC)

SMPC enables multiple parties to jointly compute a function over their inputs while keeping those inputs private. This technology is beneficial for collaborative medical research where data from different sources need to be combined and analyzed without sharing raw data. SMPC ensures data privacy and security in multi-institutional research projects.

3.3.3. Intrusion Detection Systems (IDS)

IDS monitor network traffic and system activities to detect suspicious behavior and potential security breaches. In medical data systems, IDS can provide real-time alerts and responses to unauthorized access attempts, malware infections, and other security incidents. They are essential for maintaining the integrity and availability of medical data.

3.3.4. Hybrid Encryption

Homomorphic encryption allows computation on encrypted data without decrypting it, ensuring data privacy during processing. This is particularly useful for cloud-based medical data analysis, where sensitive data can be analyzed without exposing it to the cloud service provider. Homomorphic encryption supports secure outsourcing of computation tasks while maintaining data confidentiality.

4. Evaluation

Summarizing the above domestic and international research literature, current domestic scholars mainly focus on risk assessment and security technologies of medical data, achieving rich research results in these areas and providing a strong theoretical foundation and practical experience for the development of the medical data security field. Most domestic scholars believe that the application phase of medical data is the most vulnerable to leakage. Foreign scholars primarily focus on technologies for medical data, providing important references and strong support for enhancing the security and privacy protection of medical data through the application of various emerging technologies.

Despite the significant achievements in improving medical data security, there are still some limitations and deficiencies. First, many existing studies mainly focus on the application of a single technology, lacking systematic research on the comprehensive application of multiple technologies. In practical applications, multiple technical means are often needed to address complex security threats. Therefore, how to effectively integrate various technologies to form comprehensive security solutions is a key direction for future research.

Second, existing studies often concentrate on technical discussions, neglecting considerations of legal, ethical, and social aspects. With the continuous introduction of data privacy protection regulations, how to ensure compliance and ethicality while implementing technology has become an urgent problem. Moreover, cross-border data flows and multi-party sharing bring new privacy protection and legal regulatory challenges that need further exploration in future research.

5. Advice

5.1. Comprehensive Application of Multiple Security Technologies

Single technology applications are often insufficient to address complex security threats, necessitating the comprehensive application of multiple technologies to form a complete security protection system. For example, the decentralization and immutability of blockchain technology can ensure the security of data storage and transmission, while artificial intelligence and machine learning technologies can be used for real-time monitoring and predicting security threats. Homomorphic encryption technology allows computation on encrypted data, effectively protecting data

privacy. Combining these technologies organically can provide multi-level and multi-dimensional security protection in different application scenarios.

5.2. Strengthening Relevant Laws and Regulations

As the demand for cross-border data flows and multi-party sharing of medical data increases, clear data protection laws and regulations are particularly important. These regulations should specify regulatory bodies and responsibilities, and compliance requirements during data collection, storage, transmission, and usage to ensure the legal and compliant use of data. Referencing advanced legal regulatory systems abroad, such as the EU's General Data Protection Regulation (GDPR) and the US Health Insurance Portability and Accountability Act (HIPAA), and combining them with China's actual situation to develop a suitable legal framework.

5.3. Developing Intelligent Data Management Tools

With the continuous increase in medical data, intelligent data management tools can significantly improve the efficiency and security of data management. By automating data classification, labeling, and encryption processing, it reduces human errors and risks. Utilizing big data analysis technology to optimize data storage and access strategies, improving the efficiency and security of data usage.

5.4. Enhancing Education and Training

By promoting and educating, public awareness and emphasis on medical data privacy protection can be improved. Educate patients on the basic knowledge and rights of data privacy protection, guiding the public to actively participate in data privacy protection and jointly maintain the security of medical data.

5.5. Establishment of a data compliance review mechanism

Establish a data compliance review mechanism in medical institutions and data processing organisations to ensure that the data processing process complies with laws, regulations

and technical standards. Through regular review and appraisal, problems in data management will be identified and corrected, and the compliance and security of data management will be continuously improved.

References

- [1] Zhang Gen, Tang Zhong, Li Shaoyang, et al. Simulation of risk tolerance measurement of medical privacy big data leakage [J]. Computer Simulation, 2021, 38 (12): 480-484.
- [2] Shang Jingwei. Analysis and modeling of medical big data privacy leakage behavior [D]. Yunnan University of Finance and Economics, 2020.
DOI: 10.27455/d.cnki.gycmc.2020.000051.
- [3] Kang Haiyan, Deng Jie. Enhanced hybrid encryption method for secure storage of medical data [J]. Journal of Beijing Institute of Technology, 2021, 41 (10): 1058-1068.
DOI: 10.15918/j.tbit1001-0645.2020.198.
- [4] Wang Ye, He Juan. Research on foreign experience and domestic countermeasures of medical data regulation [J]. Medicine and Law, 2021, 13 (01): 33-36.
- [5] Liao Yijie, Zhang Jing, Li Pinghui, et al. Analysis of security risks and privacy protection ideas for medical big data [J]. Journal of Chinese Health Information Management, 2020, 17 (05): 656-660+665.
- [6] Shari M F N ,Malip A . Enhancing privacy and security in smart healthcare: A blockchain-powered decentralized data dissemination scheme [J]. Internet of Things, 2024, 27 101256- 101256.
- [7] M. M ,T. S ,Kranthi G K . Securing Sensitive Medical Information with Basic and Pre-large Coati Optimization Algorithm for E-Health System Data Sanitation [J]. Wireless Personal Communications, 2024, 136 (2): 1261-1281.
- [8] Sworna L M K ,E. F ,P. N P , et al. Securing cloud-based medical data: an optimal dual kernel support vector approach for enhanced EHR management [J]. International Journal of System Assurance Engineering and Management, 2024, 15 (7): 3495-3507.
- [9] Liu Q ,Zhou F ,Chen H . Secure medical data on cloud storage via DNA homomorphic encryption technique [J]. Physical Communication, 2024, 64 102295.