Journal of Digital Media & Culture Technology Vol.2, No.1, June (2022), pp.1-9
ISSN 2983-0664(p) 2983-0850(e) *http://dx.doi.org/10.29056/jdmct.2022.06.01*

Big Data in Healthcare: An Investigation into Medical Data Management and Privacy Implications in China

P Sai Bhaskar[1], B. Dinesh Reddy[2]*

Abstract

Amidst the rapid technological advancements, big data has emerged as a crucial component in healthcare transformation, influencing clinical operations, pharmaceutical R&D and personalized treatments. The Chinese government has proactively responded to this trend by promoting and standardizing big data applications in healthcare. In this regard, the purpose of this study is to explore the evolution and implications of big data in healthcare with a specific focus on the medical data management in China. For this, this study uses a survey-based approach and proposing a risk assessment index system to underline the significance of big data and cloud resources. Findings indicates that areas such as data analysis, diagnosis processes and lack of protection are the most at risk. Side from legislation, healthcare providers, insurers and other entities handling medical data need to be educated on the importance of data privacy and the procedures to maintain it.

Keyword : Big data, security measures, health care, privacy

1. Introduction

The rapid advancement in technology has ushered in an era where big data plays a pivotal role in various sectors including healthcare. This evolution has transformed the functionality of medical organizations with new systems such as electronic medical records, hospital information systems and imaging playing a significant part in this transformation [1][2]. The utilization of big data today has revolutionized healthcare by improving clinical operations, fueled drug research and development, and tailored personalized treatment methods. In recognizing this trend, China's Council took a proactive approach and in 2016, issued the 'Guidance on Promoting and Standardizing the Application and Development of Big Data in Healthcare' [3]. This step indicates a strategic effort on part of China to integrate data production in the healthcare sector into its broader data expansion plans.

As big data continues to grow in healthcare, there are several key areas of development that hold potential for the future. These include the creation of a collaborative integrated data platform to facilitate

1 Department of Computer Science and Engineering, Vignan's Institute of Information Technology, India [Professor]
e-mail: saibaaskar24091999@gmail.com,
2 Department of Computer Science and Engineering, Vignan's Institute of Information Technology, India [Professor]
e-mail: dinesh4net@gmail.com (Corresponding author)

Received(April 9, 2022), Review Result(1st: May 3, 2022), Accepted(June 3, 2022), Published(June 30, 2022)

data sharing and analysis, further expansion of data implementation in the medical field and the increasing convergence of internet technologies and healthcare methods [4]. It is widely anticipated that with the continued refinement and enhancement of these areas, the application of big data in healthcare will not only protect but also yield effective results for the public. This will be achieved through improved service delivery and enhanced patient outcomes and further advancements in medical research and development. In this regard, the purpose of this paper is to investigate the privacy and security aspects of medical data management in China with a focus on the role of big data and cloud resources. For this, the survey was conduced and the risk assessment index system was suggested.

2. Related Work

2.1 Data Mining in medical sector

Big data science has found its application in a multitude of sectors such as healthcare, manufacturing, energy, environment and transportation. The process of accessing, distributing, processing, storing and analyzing big data plays a crucial role in extracting valuable insights from the vast sea of information [5]. In the context of healthcare, recent research emphasizes the potential benefits of big data in the early detection and diagnosis of diseases. By identifying viral symptoms in patients, healthcare professionals can predict the course of the illness more accurately [6]. Furthermore, precision in data identification can catalyze not only early disease detection but also the development of preventive measures, personalized care plans, community health services and predictive models for future disease outbreaks [7]. Moreover, as Juddoo et al. observed the role of big data extends to health management as well. A systematic approach to data management that ensures consistency and quality can help medical institutions take control of their data and use it to improve patient care [8]. In contemporary healthcare scenarios, big data analysis has become an indispensable tool. It aids in the identification of diseases and more importantly, it equips doctors with accurate predictive capabilities. This allows for more informed decision-making processes regarding patient treatment plans, contributing to better health outcomes overall.

2.2 Privacy issues of big data in medical sector

The advent of big data in the medical sector, while beneficial, has given rise to a number of privacy issues. As healthcare technologies evolve, so does the risk and complexity associated with data

management, especially in cloud services [9]. The examination of anonymized medical data has shown that sensitive information is particularly vulnerable to security threats. The security concern over diminishing data privacy is therefore a matter that deserves serious attention [10]. Furthermore, the distribution of user data introduces uncertainties regarding security. Intermediaries with access to these datasets can potentially mine confidential information. This process could inadvertently expose sensitive patient data in the information that is made publicly available. To ensure the privacy and security of patient information, it is crucial that healthcare organizations strictly follow guidelines to mitigate these risk factors. By leveraging the most up-to-date technologies, healthcare units can create robust defenses to protect patient data, thereby enhancing the privacy and security of information in the medical sector. The ultimate aim should be not just to harness the potential of big data in healthcare but also to ensure the ethical and secure handling of patients' personal health information.

2.3 Privacy problem solutions in medical sector

As the utilization of big data increases, researchers and scholars are investigating strategies and structures to address the concerns of privacy and data integrity. Initially, solutions like blockchain technology were proposed to connect different health entities such as healthcare units, insurance companies, clinics and patients. The suggestion to develop individual and separate chains for the preservation of data privacy and accuracy also emerged [11]. Galletta et al. tackled the issue of storing large files like MRI images through cloud technology but raised concerns about accuracy, reliability and data protection [12]. The importance of maintaining data integrity from the outset of clinical trials has been highlighted. It is suggested this can be achieved without transferring the data to the primary database and by verifying the data's accuracy [13]. During the transmission of data between locations, strict protocols must be followed. This includes implementing submission procedures, data administration protocols and encryption rules to safeguard the security of data transmission [14].

Despite its advantages, the use of big data in healthcare is not without challenges, primarily around privacy and security due to limitations in available tools and technologies. Many scholars are currently researching ways to enhance the security and privacy of medical data but there is a lack of robust theories and studies on these critical issues which affect every architectural stage of medical data. In the same line, this study examines the healthcare unit management system in China with a focus on privacy and security risk factors at four key stages: data gathering, data compilation, data analysis and data storage.

3. Methodology

3.1 The Healthcare Sector in China

Evaluation of China's medical and healthcare fields was done using recent data from the National Bureau of Statistics China and the China Health Statistics Yearbook in 2018. The variables considered include the number of medical institutions, healthcare personnel, mortality rates of various diseases in rural and urban areas, per capita costs for inpatient and outpatient departments, the number of public hospitals, insurance participants, expenditure and income as well as the top ten typical inpatient medical expenses.

3.2 Survey about Privacy and security of healthcare big data

A total of 100 questionnaires were distributed among healthcare professionals including nurses (30), doctors (25), technicians (30), and administrators (15). 92 responses were received, of which 89 were valid. The survey was conducted to identify the risk factors for privacy and security of the data. The questions were divided into four categories: data storage, privacy and security risks during data collection and destruction and data application. These addressed topics such as data collection by medical institutions, the medical treatment process, potential internal and external threats, inadequate authentication and diagnosis, poor data protection measures, faulty access systems, data analysis, insufficient personal privacy awareness, data encryption, lack of regulatory measures, data backup and key loss.

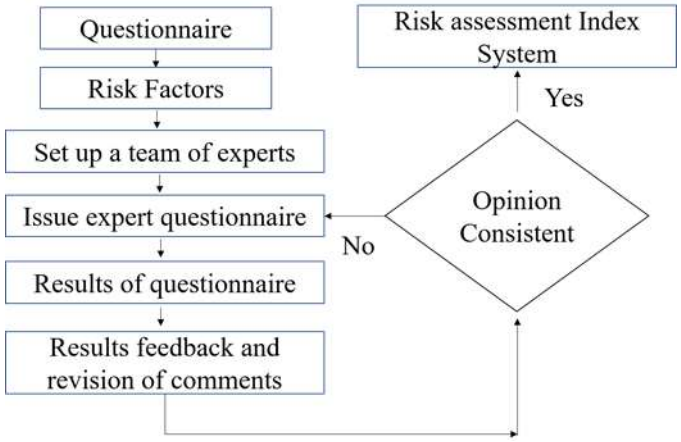
3.3 Privacy assessment model

According to the information security risk factors, the risk evaluation of data privacy and security in medical care big data was conducted from four aspects: utilization vulnerability (V), asset importance (Z), risk frequency (T), and vulnerability intensity (E). Then, the threat magnitude (R) of healthcare big data is calculated as follows.

$$R=f\ (A,B)=f\ [A\ (F,V),C\ (S,D)] \quad (1)$$

Where A defines the degree of risk, B defines risk losses, F defines frequency, V defines vulnerability severity, S describes assets and D defines the degree of vulnerability.

3.4 Risk Assessment Index System



[Fig. 1] Risk assessment index system for healthcare sector

As illustrated in [Fig. 1], the risk assessment index system is applied for the healthcare sector in terms of big data privacy. The evaluation process begins with the analysis of the completed questionnaires which allows us to establish a set of risk factors. These risk factors, in turn, are used to assemble a panel of experts. These experts are issued questionnaires designed to analyze the implications of the established risk factors. The responses from these questionnaires are used to garner feedback and suggestions for revision. To quantify the risk factors, we assign scores to each of them and define them as α1, α2, α3, ⋯.., α16 respectively. Following this, the weights of the risk coefficients are calculated. Here, 'βj' signifies the risk associated with each stage, while 'αji' represents the ith risk factor of the corresponding stage 'j'.

$$p(\beta j,\delta i,\alpha ji)=\frac{p(\delta i,\alpha ji)}{\sum_{i=1}^{n}p(\beta j,\delta i,\alpha ji)} \tag{2}$$

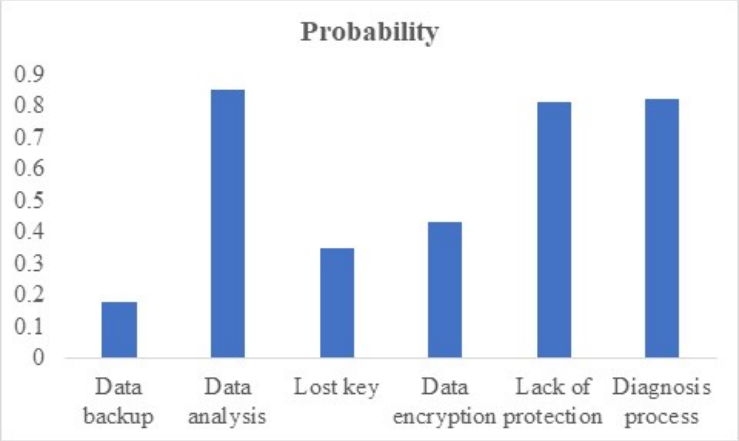
4. Results

The results based on the survey are shown in [Table 1] and [Fig. 2]. It lists the estimated probabilities associated with various risk factors pertinent to data management in a healthcare context. Each row in the table represents a different risk factor and its corresponding probability. A value of

0.18 in Data Backup suggests that there is an 18% probability that problems related to data backup will occur.

[Table 1] Risk factor probability

Serial Number	Risk factors	Probability
1	Data backup	0.18
2	Data analysis	0.85
3	Lost key	0.35
4	Data encryption	0.43
5	Lack of protection	0.81
6	Diagnosis process	0.82



[Fig. 2] Graphical representation of risk factors

The areas with the highest risk (highest probability of issues occurring) are in Data analysis (0.85), Diagnosis process (0.82), and Lack of protection (0.81). This suggests that these areas should be prioritized when it comes to risk mitigation and enhancement of data handling processes. Data encryption and the risk of lost keys have moderate probabilities (0.43 and 0.35, respectively), which suggests a moderate risk level. However, both factors relate to the protection and access of data, and problems in these areas could lead to significant data breaches or loss of access to crucial data. Therefore, despite the moderate probability values, these areas should not be neglected. The probability related to issues in data backup is relatively low (0.18). While this suggests a lower risk level, regular and secure backups are essential to ensure data recoverability in case of a data loss incident, and thus, sufficient attention should still be paid to this process. Overall it can be inferred that strategies should be developed to improve data protection measures, enhance data analysis procedures and refine the diagnosis process to reduce the associated risks. The encryption processes should be strengthened and a

secure key management process should be established to prevent key loss. Lastly, despite its lower risk, data backup procedures should be regularly reviewed and tested to ensure data recoverability.

5. Discussion

With the Chinese economy's transformation in 2012, there was a significant focus on the development of the medical and healthcare sectors to enhance the public safety. Technologies like the Internet, automation and cloud storage have simplified data handling to offer real-time solutions, enable easy transfer of medical data and help identify life-threatening diseases. However, data aggregation from various sources presents privacy and security challenges. Despite existing systems that predict and mitigate security risks, the privacy level maintained for patient data during testing and hospitalization is compromised with security levels only at 81%. This indicates a vulnerability to breaches in treatment plans and diagnosis processes.

Besides strict legislation that could enact more stringent laws and regulations, healthcare providers, insurers and other entities that handle medical data should be educated on the importance of data privacy and the proper procedures for maintaining it. Most importantly, patients should have more control over their own data. They should be able to easily view who has accessed their data and should have the right to revoke access if they choose. Informed consent should be mandatory for any data sharing. Thus, raising public awareness about the importance of data privacy and their rights can empower individuals to demand better protection from healthcare providers.

6. Conclusion

In conclusion, the integration of big data into healthcare has brought about revolutionary changes in disease detection, patient treatment and medical research. China's strategic efforts to incorporate data production into its broader healthcare sector have shown significant promise. However, the advent of big data also introduces notable privacy and security concerns. The study indicates that while strides have been made in data collection and processing, there are pressing issues regarding data privacy, security and protection that require immediate attention. Based on the survey and the proposed risk assessment index system, the results in this study illustrated that areas such as data analysis, diagnosis processes, and lack of protection are the most at risk. Therefore, it is essential to prioritize these areas for risk mitigation and process enhancement. While data encryption and the risk of lost keys show moderate probabilities, it is crucial not to underestimate their importance due to the potential severe consequences

of data breaches or loss of access. Aside from legislation, healthcare providers, insurers and other entities handling medical data need to be educated on the importance of data privacy and the procedures to maintain it. Most importantly, patients should be empowered with more control over their data, with transparent data access logs and the ability to revoke access if desired.

This study has a limitation that a relatively small sample size of healthcare professionals was surveyed. A more extensive survey involving a larger and more diverse pool of healthcare professionals, IT experts, policy-makers and patients could provide a more holistic understanding of the privacy and security issues at hand. As the fast-paced nature of technological advancements may introduce new vulnerabilities and risks that were not addressed, further research will reassess the privacy and security landscape in healthcare data technologies.

References

[1] Y. Liang and A. Kelemen, “Big Data Science and its Applications in Health and Medical Research: Challenges and Opportunities”, Journal of Biometrics & Biostatistics, January 2016, doi: 10.4172/2155-6180.1000307.

[2] S. Nazir, S. Kahn, H. U. Kahn, S. Ali, I. García-Magariño, R. B. Atan and M. Nawaz, “A Comprehensive Analysis of Healthcare Big Data Management, Analytics and Scientific Programming”, IEEE Access, vol. 8, 2020, pp. 95714-95733, doi: 10.1109/ACCESS.2020.2995572.

[3] Z. Lv and L. Qiao, “Analysis of Healthcare big data,” Future Generation Computer Systems, vol. 109, August 2020, pp. 103-110, doi: 10.1016/j.future.2020.03.039.

[4] Y. Liang and A. Kelemen, “Big Data Science and its Applications in Health and Medical Research: Challenges and Opportunities”, Austin Journal of Biometrics & Biostatistics, vol. 7, no. 3, 2016, doi: 10.4172/2155-6180.1000307.

[5] A. Enayet, M. A. Razzaque and M. M. Hassan, “A mobility-aware optimal resource allocation architecture for big data task execution on mobile cloud in smart cities”, IEEE Commun. Mag. vol. 56, no. 2, 2018, pp. 110-117, doi: 10.1109/MCOM.2018.1700293.

[6] E. Villeneuve, W. Harwin and W. Holderbaum, “Reconstruction of angular kinematics from wrist-worn inertial sensor data for smart home healthcare”, IEEE Access, vol. 5, 2017, pp. 2351-2363, doi: 10.1109/ACCESS.2016.2640559.

[7] M. Chen, Y. Hao and K. Hwang, “Disease prediction by machine learning over big data from healthcare communities”, IEEE Access, vol. 5, 2017, pp. 8869-8879, doi: 10.1109/ACCESS.2017.2694446.

[8] S. Juddoo, C. George and P. Duquenoy, “Data governance in the health industry: investigating data quality dimensions within a big data context”, ACM Appl. Syst. Innov. vol. 1, no. 4, 2018, pp. 43, doi: 10.3390/asi1040043.

[9] F.A. Kraemer, A.E. Braten and N. Tamkittikhun, “Fog computing in healthcare-a review and discussion”,

IEEE Access, vol. 5, 2017, pp. 9206-9222, doi: 10.1109/ACCESS.2017.2704100.

[10] A. Iyengar, A. Kundu and G. Pallis, "Healthcare informatics and privacy", IEEE Internet Comput. vol. 22, no. 2, 2018, pp. 29–31, doi: 10.1109/MIC.2018.022021660.

[11] X. Wang, C. Williams, Z. H. Liu and J. Croghan, "Big data management challenges in health research—a literature review", Briefings in bioinformatics, vol. 20, no. 1, January 2019, pp. 156-167, doi: 10.1093/bib/bbx086.

[12] A. Galletta, L. Carnevale, A. Bramanti and M. Fazio, "An Innovative Methodology for Big Data Visualization for Telemedicine," IEEE Transactions on Industrial Informatics, vol. 15, no. 1, January 2019, pp. 490-497, doi: 10.1109/TII.2018.2842234.

[13] K. Fritchman, K. Saminathan, R. Dowsley, T. Hughes, M. D. Cock, A. Nascimento and A. Teredesai, "Privacy-Preserving Scoring of Three Ensembles: A Novel Framework for AI in Healthcare", 2018 IEEE International Conference on Big Data, Seatlle, WA, USA, 10-13 December 2018, doi: 10.1109/BigData.2018.8622627.

[14] F. AI-Turjman and S. Alturjman, "Context-Sensitive Access in Industrial Internet of Things (IIOT) Healthcare Applications," IEEE Transactions on Industrial Informatics, vol. 14, no. 6, June 2018, pp. 2736-2744, doi: 10.1109/TII.2018.2808190.