The International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7048/6/20220952

Development and Application Status of Healthcare Big Data

Hanyu Deng[1,a,*]

[1]The High School Affiliated to Renmin University of China, Beijing, China,
a. denghanyu2016@icloud.com
**corresponding author*

Abstract: As informatization progresses in society, the data in the health and medical fields is soaring quickly, which promotes the healthcare big data generated. Healthcare is closely related to each individual's life and health, and healthcare big data has both brought great opportunities and challenges to the health and medical field. Therefore, the development of big data in healthcare is being focused on in the medical and health fields. At present, healthcare big data are being applied in clinical diagnosis, public health, infectious disease prevention, medical research, and drug development. However, these applications are not being used very efficiently nowadays. A large amount of big data is stored, but most of it still cannot be used because of the hazards of storing uniform and different structure data types. This paper uses literature searching to summarize the status quo of healthcare big data at present. Based on the background of big data, this paper describes the sources and characteristics of healthcare big data and figures out the promising applications in the health and medical areas. Moreover, this research also focuses on the data governance of healthcare big data and represents the problems now, such as the barrier between information due to non-uniform data collection methods, and provides development strategies.

Keywords: Big data, healthcare, healthcare big data, data governance

1. Introduction

Under the background of big data era, all walks of life are constantly innovating, big data has produced remarkable influence in various areas and has played a significant role in the development of this field in these areas. Among them, healthcare generates a huge amount of information data, which is growing exponentially, and the collection of these data is almost the same size and nature as the real big data, hence the name is called healthcare big data. Although the application of big data in medical and health care is lagging behind compared with that in other fields, many other researches mentioned that the value of big data application in this field is inestimable. In medicine, clinical, insurance, health monitoring and other aspects, big data in health care can create great benefits and enable stakeholders to obtain a lot of profits[1]. According to a McKinsey report, if big data were used effectively in America's healthcare sector, the industry could be worth an additional $300 billion a year[2].

The amount of these data is huge, the structure is diverse, the storage mode is diverse, the data quality is uneven, and the data types are scattered, which makes the utilization and analysis of these medical and health data greatly difficult, and also greatly reduces the availability of medical and

© 2023 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/6/20220952

health data[3]. Therefore, efficient data governance is very important, only after sorting out the data can it really have research value, otherwise, even if a large amount of data is collected, the data is still useless and cannot be analyzed. But there is a relative lack of research on data governance of healthcare big data at present. This paper adopts the method of literature reference through searching and consulting relevant literature to summarize the status quo of healthcare big data application and development and expound on the data governance problem.

2. Big Data in Healthcare

Healthcare big data is complicated and comprehensive, this chapter outlines healthcare data from three aspects: source, characteristics and application.

2.1. Source

Healthcare big data covers a wide range of information, including residents' electronic health records, clinical reports, physician notes, health insurance, electronic medical records, and nearly all health-related or medical-related information[4]. Moreover, the sources of these information are also diverse and complex, the sources of this information are also diverse and complex, sensor systems, mobile devices such as mobile phones, some of the online platforms related medical or health, and even some types of photos and videos are all sources of huge amounts healthcare information data. As a result of these complex sources of information, big data for health care included multisource data.

Big data in healthcare are similar with biomedical data, so it can be broadly divided into two categories. One type is the data, which are generated from biological sources such as human tissue and body fluids. The other type named observation data such as patient diagnoses include the notes or documents of medical information, which are provided by clinicians and other relative professionals[5]. According to the different classify standard based on Chinese society, seven basic sources of healthcare data are provided:

2.1.1. Hospital Information System (HIS)

The hospital information system refers to the use of network communication technology, computer software and hardware technology, and other means to manage the operation of various hospital departments and in the various stages of medical activities to collect, store, process, extract, transmit, summarize, process, and generate various information in order to provide the hospital with a comprehensive and automated management information system. Moreover, HIS includes five parts: electronic medical record database, medical image database, hospital management database, public resource database and knowledge database.

2.1.2. Regional Population Health Information Platform

Health and family planning department hospitals, public health institutions, family planning institutions, primary medical and health institutions and health and family planning administrative institutions provide most of information source of regional population health information platform. Furthermore, lots of social departments for instance bank, insurance, education, public security are also linked to this platform. Thus, regional population health information has the characteristics of a wide range of sources, complex types, a large amount of information, and scattered storage.

DOI: 10.54254/2753-7048/6/20220952

2.1.3. Public Health System

Public health information system refers to the application system that provides business operation and management services for disease prevention and control institutions, health supervision and administration institutions, maternal and child health care institutions, mental health and health management institutions, 120 emergency centers, blood stations, etc. Generally, it includes disease prevention and control data, health supervision data, health emergency command data, medical rescue data, maternal and child health care data, mental health management information system, blood management data, etc.

2.1.4. Internet Data

With the rapid development of mobile devices and mobile Internet technology, mobile devices have been integrated into social life, and a large number of medical and health-related data are generated on the Internet. To be specific, network registration on platform, online trade of medical equipment or drugs.

2.1.5. Biological Information

Biological information mainly refers to the relevant information obtained through gene technology, such as gene identifier, name, species origin, etc.

2.1.6. Industry Related Data

Government, education and commercial are the industries related to healthcare. So government insurance, medical literature, the pharmaceutical industry and pharmaceutical sales are grouped into this type of data.

2.1.7. Subject Related Data

Disciplines related to medical research include life sciences, demography and environmental sciences, so relevant data are also covered by healthcare big data [6].

2.2. Characteristics

The characteristic of big data are commonly represented by a few of "V", namely volume, velocity, variety, value and veracity. Volume of big data is measured in terabytes (TB = 1012 bytes), petabytes (PB = 1015 bytes), and zettabytes (ZB = 1021 bytes), and can be used to determined the size of data. Velocity both indicates the rate of generation of data and the speed of data processing; it can be used to determine whether the collection and processing of data are efficient. Variety indicates the basic nature of big data, which includes structured, semi-structured, and unstructured data. Value expresses the worth of collected data, which can be used to determine whether the data is useful or useless. The veracity of the data indicates its trustworthiness; the accuracy and authenticity of the data may differ because of various data sources or for other reasons.

5"V" characteristics of healthcare big data:

Volume, the information content of healthcare big data is very huge, including biological data such as gene sequence, data generated by medical and healthcare departments such as electronic medical records, and a large amount of related information on the Internet. As information technology in healthcare area matures, the amount of healthcare data grows exponentially day to day.

DOI: 10.54254/2753-7048/6/20220952

Velocity. Huge amounts of healthcare data are being generated from different sources all the time, the data is being generated very quickly. However, the huge amount of data, which contains various structures and types, makes the processing of the data inefficient.

Variety. Different sources of healthcare big data diversify data types.

Value. In the large amount of generated healthcare data, only the valuable part of the data can have real effect.

Veracity. Healthcare data comes from different sources. Therefore, ensure the reliability and authenticity of the data.

Besides, in the relevant literature, it is also mentioned that healthcare big data has the characteristics of timeliness, incompleteness, privacy and redundancy[7]. The update speed of healthcare big data is fast, the storage is easy to be incomplete, and similar information is easy to be recorded repeatedly. Also, in healthcare area, privacy is extremely important.

2.3. Application of Healthcare Big Data

The application prospect of healthcare big data is extensive, it will make huge contribution to healthcare area, and even take the healthcare field to a new stage. The application of healthcare big data including clinical decision-making, medical evaluation, scientific research, health management, personalized treatment, disease prevention and so on. At present, more in-depth research and widely considered to be more promising application areas are as follows:

2.3.1. Clinical Decision-making

Clinical diagnosis is very complex, and any small factor can affect a doctor's judgment of a patient's condition. In traditional diagnostic methods, doctors rely on experience to deal with these complex situations. Healthcare big data can provide a lot of auxiliary decision-making help for clinical medical diagnosis. Neural networks are built through machine learning algorithms to more accurately judge the patient's condition through the past big data. Making clinical decisions more efficient, more accurate and more intelligent.

2.3.2. Personalized Treatment

The medical big data system can establish medical records for each patient and record their healthcare related data, so it can better provide personalized services for patients.

2.3.3. Scientific Research

Healthcare data has always been an important support for scientific research. A large number of medical data provide research basis and evidence for scientific research.

2.3.4. Disease Prevention

Healthcare big data from various health departments can monitor diseases with infectious risks, track relevant personnel in real time, analyze and deal with them, provide timely information to health departments, and provide strong support for epidemic prevention and control. The use of healthcare big data in epidemic prevention and control can greatly improve the efficiency of this aspect, and reduce the related medical expenses, which will be a large amount of money.

DOI: 10.54254/2753-7048/6/20220952

2.3.5. Medical Evaluation

Healthcare big data contains a lot of biological data, such as gene sequence and gene structure, which are very useful for medicine development.

3. Data Governance

Data governance refers to the way to manage data. Due to the particularity of medical and health big data, a large number of medical and health data are generated and collected all the time. These data come from various sources, and they cannot be directly processed or analyzed without data processing. Data governance will be for a large number of data and information organization structures, set up standard systems, and standardization of data. The data after processing has structural organization, more specifications, which can greatly improve the efficiency of analysis, so as to improve the efficiency of the whole medical or scientific research process and greatly reduce the cost.

3.1. The Problems of Data Governance

There are many problems in data governance, which means that big data can be effectively applied to only a small part of healthcare for the time being. If most of the data can be effectively governed, the overall medical efficiency will be greatly improved, and the government's expenditure on medical treatment will be reduced by hundreds of billions of dollars, which will make the medical and health field greatly develop. There are a few problems with healthcare big data governance, including the following:

(1) Different data standards

At present, the storage methods for healthcare big data are not unified, the information systems used by different institutions are from different manufacturers, and the data storage standards are different, which makes it difficult to form a connection between relevant data. So the integration is difficult, which means the processing and analysis efficiency is greatly reduced.

(2) Difficulties in information sharing

The sources of healthcare big data are complex, there are barriers between information, it is difficult to establish a connection, so it has become an information island. For example, an electronic medical record of the same person treated in a hospital cannot be linked to his health record registered in a community health facility. Therefore, it is important to break down the barriers between different kinds of information.

(3) Data security

Healthcare big data includes a large amount of medical information from various people, therefore, data privacy and confidentiality have become essential issues.There are security risks in the transmission and acquisition of personal health information. If a person's health information is obtained or tampered with by outsiders, the consequences will be unbearable. However, China's domestic regulations on such data are incomplete, and there are great risks if the information is leaked.

(4) Complicated data content

The current healthcare big data system is not mature, and the data collection and storage are not standardized, resulting in a lot of invalid data. The incomplete data collection, the mixed data structure, the isolated information and the redundant of data are all called invalid data. These data may occupy space and even interference data application, it is very important to improve the quality of data content.[8]

DOI: 10.54254/2753-7048/6/20220952

3.2. Measures of Data Governance

A big problem in data governance is the lack of relevant talent that is good at using information systems. Relevant medical personnel have a shallow understanding of healthcare big data. Strengthening the popularization of healthcare big data enables medical staff to make more efficient use of it, which can greatly help some data governance issues —improve the accuracy of data information, make it more complete, and reduce the generation of low-quality data. The construction of a more complete healthcare big data system, unified data standards, unified operation mechanism, can break the barriers between information, truly connect the data, will greatly improve the application efficiency. Also, laws and regulations on big data need to be improved to ensure data security while sharing data.

4. Conclusion

In the information age, most information is transformed into data storage, among which the medical and health fields are also becoming informationized. The explosive growth of big data in healthcare has brought great opportunities and challenges to the medical and health fields. At present, the application of healthcare big data has a good development prospect, and its correct application will bring huge benefits to this field. But at the same time, there are still considerable problems in current data governance that prevent making the most of healthcare big data. Medical health directly affects everyone's life and health, so the development of this field is extremely important, and medical big data will bring great value. Furthermore, because of the lack of primary data, it is impossible to conduct a practical analysis. The future research may mainly focus on how to break down data barriers and ensure data security so that healthcare big data can play a greater role, which will promote the medical and health fields' development.

Reference

[1] *Pramanik, P. K., Pal, S., & Mukhopadhyay, M. (2022) Healthcare Big Data: A Comprehensive Overview. In I. Management Association (Ed.), Research Anthology on Big Data Analytics, Architectures, and Applications pp. 119-147. IGI Global.*

[2] *Meng, Q., Bi, D., Zhang, Y. M. & Yin, X. (2016) Research on the development status and application pattern of healthcare big data. Chinese Journal of Health information Management(06),547-552.*

[3] *Raja, R., Mukherjee, I., Sarkar, B. K. (2020) A Systematic Review of Healthcare Big Data. Scientific Programming, 15.*

[4] *Lina Ai (2020) Design and Implementation of Multi-source Heterogeneous Healthcare Big Data Governance Platform. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202002&filename=1020063020.nh*

[5] *Richterich, A. (2018) Big Data in Biomedical Research. In The Big Data Agenda: Data Ethics and Critical Data Studies. pp. 53–70.*

[6] *Bo Tao (2019) Research on the Analysis and Application Mode of Medical and Health Data Based on Big Data Platform. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202001&filename=1019917829.nh*

[7] *Dai, M. F. & Meng, Q. (2017).Opportunities and challenges in healthcare big data mining and analysis. Chinese Journal of Health Information Management(02),126-130.*

[8] *Zhang, Z., Yang, C. M., Xu, J., Li, L. & Zhou, Y. (2022). Development status and data governance of healthcare big data application. Journal of Medical Informatics(07),2-8.*