An overview of Big Data in Healthcare: multiple angle analyses

Xunjie Gou, Zeshui Xu

Business School, Sichuan University, Chengdu 610064, China.

Correspondence to: Prof. Zeshui Xu, Business School, Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu 610064, Sichuan, China. E-mail: xuzeshui@263.net


Received: 8 May 2021  First Decision: 16 Jun 2021  Revised: 1 Aug 2021  Accepted: 26 Aug 2021  Available online: 31 Aug 2021

Academic Editors: Jie Lu  Copy Editor: Xi-Jun Chen  Production Editor: Xi-Jun Chen

Abstract

Big data have been in use since the 1990s, which usually include some complex data sets whose sizes are beyond the ability of commonly used software to handle within a reasonable period of time. In recent years, big data analytics by providing personalized medicine and regulation analysis, providing clinical risk intervention and forecast analysis, reducing waste and nursing patients with external and internal variability, standardization of medical terminology and patient registration, and fragmentation of the solution, help to improve health care. This paper provides an overview of the contents of big data healthcare. We summarize some kinds of medical big data, including the electronic health records, the medical image data, the healthcare system big data, the health Internet of Things and healthcare informatics, the remote medical monitoring big data, the biomedical big data, and other sources of big data. Furthermore, we discuss some methods for handling different kinds of medical big data. Additionally, we analyze the privacy of medical big data and summarize some methods and technologies to protect privacy. Aiming at some special cases, we list some other analyses and methods for them. Most importantly, we discuss the potential challenges and future research directions related to big data healthcare.

Keywords: Big data healthcare, medical big data, big data methods and technologies, privacy, challenges
INTRODUCTION

Big data have been in use since the 1990s, with some giving credit to Mashey[1] and Lohr[2] for inventing or at least popularizing it. Big data generally include some complex data sets whose sizes are beyond the capabilities of common software to handle within a reasonable period of time[3]. Big data “size” is rapidly expanding, ranging in size from tens of terabytes to many petabytes since 2012. Also, big data require a range of technologies with new integration patterns to reveal insights into large numbers of different and complex data sets[4-8].

Meanwhile, big data are the large or complex sets of data that traditional data processing applications cannot handle. The term “big data” usually refers only to the use of forecast analyses, user behavior analyses, or some other high-grade data analytic approaches to extract value from data, and rarely involves data sets of a particular size[9]. Moreover, big data can mean different things to different people[10]. Under normal circumstances, people like to divide big data into two main categories, structured and non-structured data[11]. Up to now, big data mainly consist of the five characteristics (5V) including Variety, Volume, Velocity, Variability, and Veracity, described in Figure 1.

Also, there exists a 6C system[4,9] in factory work and physical information systems:

<table>
<thead>
<tr>
<th>Connection</th>
<th>Cloud</th>
<th>Cyber</th>
<th>Content/context</th>
<th>Community</th>
<th>Customization</th>
</tr>
</thead>
<tbody>
<tr>
<td>(sense and network)</td>
<td>(computing and data on demand)</td>
<td>(sensory and memory)</td>
<td>(meaning and correlation)</td>
<td>(sharing and collaboration)</td>
<td>(personalization and value)</td>
</tr>
</tbody>
</table>

Besides, the big data analyses in production applications are also known as the following 5C structure:

<table>
<thead>
<tr>
<th>Connection</th>
<th>Conversion</th>
<th>Cyber</th>
<th>Cognition</th>
<th>Configuration</th>
</tr>
</thead>
</table>

In 2011, the McKinsey Global Institute[12] reported the characteristics about the major components and ecosystem of big data, which involve data analysis techniques, big data technologies, visualization, and more.

Furthermore, big data have improved the need for information management experts. For example, a great many technology giants such as IBM and Microsoft have spent more than $15 billion on software companies. Simultaneously, big data can be used to amounts of fields such as international development, manufacturing, cyber-physical modelling, healthcare, education, media, sports, and sampling big data.

In recent years and in medical domain, big data analytics by providing personalized medicine and regulation analysis, providing clinical risk intervention and forecast analysis, reducing waste and nursing patients with external and internal variability, standardization of medical terminology and patient registration, and fragmentation of the solution, help to improve health care[14]. With the added adoption of mobile health (mHealth) and eHealth, the scale of data will expand further. As we know, except the above 5V characteristics, big data healthcare has its own characteristics including large base size, fast growth rate, low value density, polymorphic data structure, incompleteness, and privacy. Therefore, more and more scholars have spread the research of big data healthcare from different directions including electronic health records[15-24], medical image[27-40], Internet of Things (IoT) and healthcare informatics[41-45], remote medical monitoring[46-48], and biomedical big data[49-57].
Aiming at different data types and practical necessities, lots of methods, models, algorithms, and technologies have been developed. For example, data mining and machine learning can be utilized to find patterns and knowledge based on large amounts of data. Medical big data integration and clustering are also important for solving various medical problems such as distinguishing the kinds of patients, etiological analysis, therapy selection, and so on. By tailor treatment and prevention plans, personalized healthcare, and precision medicine, the best outcomes can be achieved for everyone\textsuperscript{[46,57]}.

Considering that medical big data are more and more important for the whole society, in this paper, we make an overview about big data healthcare, and the contributions are summarized as follows: (1) we will comb and summarize different forms of medical big data; (2) lots of methods and technologies about medical big data are reviewed; and (3) the challenges and further research directions are discussed. Therefore, this paper could be taken as guidance for understanding big data healthcare.

The remaining contents of this paper are arranged as follows: Section 2 mainly discusses several different kinds of medical big data. In Section 3, some methods of different big data healthcare are reviewed including data mining, machine learning, medical big data integration and clustering technologies, and personalized healthcare and precision medicine. Section 4 discusses the current situation of medical big data privacy and the methods and techniques for protecting that privacy. Section 5 mainly discusses some other analyses and methods. The challenges and further research directions are analyzed in Section 6. Finally, we conclude the paper by summarizing the main conclusions in Section 7.

**DIFFERENT FORMS OF MEDICAL BIG DATA**

As we know, medical big data appear in different kinds of modalities. Here we mainly enumerate five medical big data forms, including electronic medical records (EMRs) (or electronic health records, EHRs), medical images, health IoT and healthcare informatics, remote medical monitoring, and biomedical big data.

**Electronic medical records**

In the United States, EMR is also the digitized version of medical record, and it is a part of the larger EHR. Additionally, the EMR can rapidly access to a wide range of clinical and demographic data and avoid the latency needed to get administrative data\textsuperscript{[16]}. So far, several experts and scholars have researched the EHRs (or EMRs) from different angles, such as building model and developing algorithms\textsuperscript{[16-20]}, studying metadata...
and standards\cite{15}, natural language processing with EHRs\cite{19}, and so on.

Additionally, in some actual processes, the EHR systems also occupy important positions. For example, EHR system plays an important role in the field of telemedicine\cite{21}. Recently, Maheswaranathan et al.\cite{22} studied the impact of COVID-19 and telemedicine implementation on EHR utilization and practice patterns. By big data techniques on EMRs, Liu et al.\cite{23} examined gender and age detection rates for some important hypertension comorbidities and outlined their relationships to reveal the risk of hypertension in patients. With the universal use of EMRs, biomedical research provides access to a wealth of health-related information\cite{24}. Furthermore, the rise of EMRs has led to the increasing of large-scale observational research on perioperative period\cite{25}. However, the EHRs with patient treatments and outcomes are rich but underutilized information\cite{26}.

**Medical image data**

*The concept of medical image*

Medical image is a visual representation of a human body or the part of body, which is usually applied to detect, diagnose, or monitor disease by electromagnetic radiation in medical procedures. It is also the primary method of the present medical diagnostic processes. As we know, medical image comes from a wide spectrum of imaging techniques including plain X-ray, ultrasound, computed tomography, and others\cite{27}. Additionally, medical image plays a pivotal role in surgery\cite{28} and physicians’ diagnostic decision-making\cite{29}.

*Some methods and techniques about medical images*

Firstly, numerous methods, models, and algorithms have been developed to handle medical image information. Since the last century, X-ray imaging has become one of the most widely used tools in the field of medical diagnosis. However, the main challenges always remain in medical diagnosis. For example, considering that there are many limitations in existing methods involving big data biomedical image fusion, a fusion approach was proposed on the basis of the spherical coordinate for biomedical images big data\cite{30}. Besides, a damage-free multi-component medical image compression approach was developed by Xin and Fan\cite{31} depending on big data mining.

Furthermore, some graphical models were also constructed to describe medical images. Aiming at brain CT images, Durand et al.\cite{29} first constructed a graph involving the topological relationships between lesions and ventricles, and then developed an approach denoted by Frequent Approximate Subgraph Mining based on Graph Edit Distance (FASMGED). Besides, Kurc et al.\cite{32} defined three functions to search images and image regions, compute quantitative features on images, and store and index computed quantitative features.

Many damage-free methods have been developed for medical images. To compare and assess these multistage compression techniques and design some more effective big data compression methods, Karimi et al.\cite{11} analyzed all compression stages’ effectiveness and the overall performance of the algorithm. Meanwhile, Ullah and Arslan\cite{33} proposed a parallel time-delay multiplier algorithm for microwave medical imaging on the basis of spark big data framework. As big data medical image fusion is a key problem, Zhang et al.\cite{34} gave a big data medical image wireless sensor network fusion method based on spherical coordinate domain (SCD) coding. In view of the image storage problem, some methods are also developed such as the medical image storage and access method based on Hadoop\cite{35}, the eural-assisted image-dependent encryption scheme\cite{36}, and the novel framework for online medical image visualization based on shadow agent\cite{37}. 
Moreover, medical image can also be used for organ depiction, lung tumor identification, spinal malformation diagnosis, arterial stenosis detection, aneurysm detection, and so on\cite{38}. In addition to machine learning methods, image processing techniques such as enhancement, segmentation and denoising are used in these applications. Additionally, medical imaging includes a wide range of different image acquisition methods that are commonly used in a variety of clinical applications\cite{39}. Furthermore, by the 5V features of medical imaging big data, Zhang\cite{40} discussed the feasible and long-range perspective solutions to big data problems in medical imaging informatics, which is drawn in Figure 2.

Health IOT

The health-IOT is a milestone in the development of health information systems\cite{41}. The big data and IoT techniques can address the health care information processing challenges\cite{42}, have important impact and implications for health care delivery\cite{43}, and be applied in the evaluation process of sustainable smart city for real-time evaluation\cite{44}. In recent years, the health-IoT has been applied into lots of fields including medical industry, exercise promotion, mental support, physical health condition analysis, and so on\cite{41,45}. Table 1 shows these applications.

Remote medical monitoring big data

Remote medical monitoring has expanded the use of telemedicine to treat patients with chronic diseases and diseases by monitoring their daily health conditions so that preventive and emergency care can be provided as needed\cite{46}. As the technique gets better and better, it will become the standard procedure that can be used to manage some conditions. In the era of big data, one of the important areas of research is storing, monitoring, and analyzing signs of the body using the wireless body area network. For this area of research, some medical monitoring system are constructed such as medical service middleware system\cite{47} and the remote real-time medical monitoring system architecture on the basis of IOT and cloud computing\cite{48}.

Biomedical big data

Biomedicine is a frontier interdisciplinary subject developed by integrating the theories and methods of medicine, life science, and biology. Its basic task is to apply biological and engineering techniques to study and solve the problems related to life science, especially medicine. As an important part of medical “big data”, biomedical information is closely related to the formation and development of biotechnology in the 21st century and is an important engineering field related to improving the level of medical diagnosis and human health\cite{49}.

For the biomedicine with big data, some scholars developed methods to transform biomedicine such as the population approach (P5 Medicine)\cite{50}, the approach of adjusting covariables that affect features and/or goals in Tree-based Pipeline Optimization Tool\cite{51}. Additionally, the application of biomedical image fusion in big data computing is developing strongly\cite{52-54}, such as the biomedical big data image fusion computing method based on spherical coordinates\cite{52}, and some big data technologies for the purpose of solving urgent problems of biomedical diagnostics\cite{54}. Furthermore, some methods are also developed to improve the biomedical signal search results in big data, including the randomized Monte Carlo sampling method\cite{55}, the approach named bootstrapping for unified feature association measurement (BUFAM)\cite{56}.

In fact, so far, very large biomedical research databases have recently been identified as having the potential to accelerate scientific discovery and significantly improve medical treatment. The study of these databases could also result in deep changes in laws, policies, and litigation strategies\cite{57}.
Table 1. Some fields of health-IoT

<table>
<thead>
<tr>
<th>Applications</th>
<th>Services contents</th>
<th>Major applicable objects</th>
<th>Device type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical industry</td>
<td>The positioning of medical staff and patients, patients’ detection, wireless mobile ward-round system, etc.</td>
<td>Doctors, patients, management personnel, and hospital employees</td>
<td>Dedicated devices</td>
</tr>
<tr>
<td>Health monitoring</td>
<td>Monitor physiological indexes of patients; Providing reference for disease treatment</td>
<td>Patients of chronic diseases</td>
<td>Dedicated devices</td>
</tr>
<tr>
<td>Exercise promotion</td>
<td>Monitor physiological indexes during the exercise; Providing guidance for physical exercise</td>
<td>Common people and athletes</td>
<td>Wearable devices</td>
</tr>
<tr>
<td>Mental support</td>
<td>Relieve psychological stress and treat psychological diseases</td>
<td>Patients of psychological stress and psychological diseases</td>
<td>Dedicated devices or wearable devices</td>
</tr>
</tbody>
</table>

Figure 2. 5V features of medical imaging big data.

SOME METHODS FOR HANDLING MEDICAL BIG DATA

Data mining, machine learning, and cloud computing

Data mining

Data mining is a technology that uses artificial intelligence, automatic learning, statistics, databases, and other tools to find patterns and knowledge from a large amount of data\(^{[57]}\). In big data healthcare, data mining has important application potential in clinical medicine\(^{[58]}\), biological and biomedical research\(^{[60]}\), bilateral implanted patients\(^{[58]}\), and identifying novel drugs in the cardiovascular field\(^{[61]}\), assessing the compensation of diabetes and arterial blood pressure control\(^{[62]}\), and coping with the problem of information overload in healthcare\(^{[48]}\). Big data mining depends on several basic theories and technologies such as fuzzy theory, Bayesian network, and rough set theory. Some concrete methods and algorithms based on data mining are developed including the next-generation sequencing (NGS) technologies\(^{[52]}\), a text mining method\(^{[49]}\), some data mining algorithms are developed to find the correlations between unilateral medical records\(^{[59]}\). Additionally, some scholars applied data mining methods in cardiovascular field, such as finding hidden relations between many arguments and clinical outcomes\(^{[61]}\), a solution which combines both technologies in a single analytical system\(^{[63]}\). Another application of data mining methods was in a decision tree prediction model that was established to analyze the patterns of foot diseases\(^{[64]}\).
Besides, some highly popular platforms, such as Twitter\cite{65,66} and Google\cite{67}, can be used as data mining sources to deal with medical big data. For example, as a very popular information exchange platform, Twitter\cite{65} can be utilized as a source of data mining to understand the people affected by autism spectrum disorder (ASD)-their behavior, worries, demands, etc. Hays and Daker-White\cite{66} identified and described the scope of views represented about healthcare data, posted data on Twitter during the project’s delay and provided insight into the project’s strengths and weaknesses.

**Machine learning**

Machine learning is a multi-domain interdisciplinary subject, and it is specialized in the study of how computers simulate or realize human learning behavior to acquire new knowledge or skills and reorganize existing knowledge structure to continuously improve its own performance\cite{68}. Specially, machine learning can serve as a useful tool to leverage complex clinical data and help guide critical clinical decisions\cite{16}, as well as describing the concepts and techniques of machine learning for processing and analyzing health data, especially those that are most widely used in rheumatology\cite{69}.

**Cloud computing**

Cloud computing\cite{53,63,70-76} is a kind of distributed computing. It refers to the decomposition of huge data processing programs into countless small programs through the network “cloud”. Then, the results of these small programs are processed and analyzed by a system composed of multiple servers and returned to users. Based on the concept, gordian technologies, kernel problems, and theories of cloud computing, the key problems involving cloud computing of medical big data and health informatization were discussed\cite{70}, and the home-diagnosis was proposed\cite{71}. Moreover, cloud platforms\cite{72}, cloud based healthcare systems\cite{74}, and similarity search-based clinical decision support systems\cite{75} were structured.

**Medical big data integration and clustering**

In recent years, several methods\cite{53,77-83} have been established for achieving big data integration. Similarly, big data clustering methods are also important in solving various medical problems such as the K nearest neighbors (kNN) classification algorithm\cite{84}, the hierarchical learning algorithm\cite{85}, and the feature selection algorithm\cite{86}.

**Medical big data integration**

Many integration methods have been developed, such as Big Linked Data\cite{82} and G-DOC Plus\cite{53}. For example, the G-DOC Plus was used to process a series of biomedical big data on the basis of cloud computing and other tools\cite{87}. Especially, Reina et al.\cite{88} started the development of the integrated semantic framework for multidimensional data analysis. Based on the Hadoop platform, Lyu et al.\cite{89} proposed a system involving dealing with clinical data to copy with the issues in the integration of big data, and it can unite a variety of multi-source heterogeneous data, including EMR, Laboratory Information Management System (LIS), and others.

Translational medicine is the field of translating the basic life science research achievements into new instruments and approaches in the clinical setting. Satagopam et al.\cite{79} presented an integrated workflow which can be used to explore, analyze, and interpret translated medicine big data. Furthermore, Gligorijević et al.\cite{80} presented the latest developments involving big data integration approaches that can discover personalized messages from big data generated by all kinds of multi-omics studies. Additionally, a universal semantic big data framework for data integration was developed\cite{81}, and Merelli et al.\cite{78} discussed three methods for data integration, including semantics, ontologies, and open format.
Medical big data clustering

Deng et al.\cite{84} divided the whole data set into many parts through a constructed K-means clustering, in which each part is an efficient kNN clustering algorithm for big data. Additionally, some algorithms and models have also been developed\cite{85-87,90}, such as a hierarchical learning algorithm\cite{85}, the neural fuzzy Linguistic Classifier based on feature selection algorithm\cite{86}, the balanced clustering model for heterogeneous big data sets of intelligent healthcare\cite{87}, and some parts of the International Classification of Diseases Clinical Modification (ICD-CM)\cite{90}.

Personalized healthcare and precision medicine

Personalized healthcare\cite{46,91-93} also known as precision medicine, refers to a customized medical model based on personal genome information, combined with proteome, metabolome, and other relevant internal environmental information, to design the best treatment plan for patients, in order to maximize the treatment effect and minimize the side effects. Precision medicine considered as trend of future medical model, is a rising healthcare model that can offer precise diagnoses\cite{93}.

Personalized healthcare

To promote the development of personalized healthcare under the big data environment, some research was carried out. For example, cognitive computing is one of the new technologies for integrating and analyzing big data sets, which was allocated to sustain life science study\cite{94}. Additionally, an ART framework was proposed for large-scale patient index in personalized healthcare\cite{95}. More importantly, Poh et al.\cite{96} presented strategies for secure healthcare platforms for personalized analyses which focus on three parts: big data expression, big data affair and safety, and personalized analyses through machine learning algorithms. Furthermore, Han and Liu\cite{97} proposed a novel transcriptome marker diagnosis based on big RNA-seq data by systematically treating the entire transcriptome as a contour marker.

Precision medicine

Precision medicine, benefited from the development of technologies, is more and more important in people’s real life. Shin et al.\cite{93} provided the basic information of precision medicine and some fire-new proposed concepts, such as sophisticated healthcare ecological system, big data handling, and omics technology. Schapranow et al.\cite{98} shared the details about their Medical Knowledge Cockpit, which is an instantaneous analysis of medical big data achieving precision medicine. Additionally, Noor et al.\cite{99} highlighted the big data challenges faced by converting study teams in the era of precision medicine. Specially, some problems of precision medicine were noticed including administrative claims, regulatory concerns and legal issues, the standard of care concept, the safeguard of privacy, and incidentally identified risks\cite{100}.

MEDICAL BIG DATA PRIVACY

The current situation of medical big data privacy

In this connected world of social networks, privacy is a word that matters to everyone. Certainly, health related data are the very sensitive data that people do not want to make public, and there are concerns about a lack of respect for confidentiality and privacy in hospitals. Thus, the hospitals should do more work to preserve their clients’ privacy. As we know, data sharing still is an issue in healthcare because of the wide range of privacy problems. Although there is a lot of research on data mining to protect privacy, medical institutions are reluctant to release their data due to the stipulation from some related laws and regulations\cite{101}.
Now, our society can gain much value from big data. For example, the census data can be used to study the allocation of public resources and fight diseases based on the medical records from hospitals. Additionally, taking account of all legitimate interests, it is necessary to make appropriate research exemptions and agree to not use sensitive personal data in medical research\[102\]. In some existing research, more on security issue arisen from Hadoop Architecture base layer called Hadoop Distributed File System\[103\] were focused and analyzed. Considering that there may exist significant legal and ethical issues when using healthcare big data, Gray and Thorpe\[104\] addressed the scope of them and discussed how to effectively manage these issues to achieve the full potential of big data.

Some methods and techniques for medical big data privacy
Aiming at protecting the big data medical privacy, large amounts of research and analyses have been done and several approaches have been researched such as the multi-agent architecture\[101\] and the HireSome-II\[101\]. Additionally, some algorithms are also established including the approach for preserving the privacy of the EMRs\[106\], a secure and private data management framework\[107\], and a context sensitive approach based on DAC and RBAC models\[108\].

SOME ANALYSES FOR MEDICAL BIG DATA
In addition to the above categories, some analyses for big data healthcare are summarized.

Although the technological advancement in medical sphere has reached saturation, a break-through can be achieved by Prognotive computing\[109\], and it has to do with big data analyses. In addition, Reverse Engineering and Forward Simulation (REFS)\[110\], as a proprietary “big data” analytic platform, can be applied to dimensions of metabolic syndrome.

Additionally, there exist amounts of big data healthcare methods for different aspects, such as wearable monitor\[111\], security domain\[112\], NGS technologies\[113\], predictive analytics\[114\], big data service platform\[115\], health information services\[116\], Smart Health Service methods\[117\], and chronic disease\[118\]. Firstly, a conceptual architecture for biosurveillance is established and it is focused on the long-term goal owing the real early warning capabilities\[112\]; Liang et al.\[117\] achieved the health knowledge which came from big data on chronic diseases and then the knowledge was applied to Smart Health Service methods. Secondly, some technologies about NGS and graphical tools for NGS analytics were also researched\[113\], and the fusion of predictive parsing and big data also has large potential in healthcare\[114\]. Furthermore, the potential for sensors use in healthcare data acquisition was explored\[115\], and the importance and urgency of promoting health information literacy also was discussed\[114\].

CHALLENGES
This section mainly discusses some challenges of healthcare big data in different areas, and some future research directions about the development of healthcare big data.

Some challenges
Firstly, the opportunities and challenges of the healthcare IoT can be embodied in a lot of forms, such as innovative business models, nonfunctional requirements or system qualities described system attributes or constraints, application context and physical environment, diagnosis and monitoring health sensors, node physical characteristics, market analysis and positioning, portfolio of wireless networks, the fifth generation of communication networks, medical body area networks, personal health device communication standards, wearable technology and cloud platforms, and big data analytics technologies.
Secondly, there were several opportunities and challenges of biomedical informatics \(^{[50]}\). The era of big data, with new technologies that bring massive measurements, has arrived such as high-performance computing technology, artificial intelligence and machine learning, visualization and visual analytics, biomedical informaticians, and recent advances in information technology.

Thirdly, in the future, the aggregation of electrocardiograms and images from hospitals around the world will be the focus of big data medicine. While promoting the application of information technology in telecardiology, we still need to solve several problems including big data confidentiality in the cloud, data interoperability among hospitals, and network latency and accessibility.

Big data has ushered in a major transformation of the era. The challenges of big data in healthcare can be represented by the following several applications: (1) analysis of EMR: at present, most electronic medical records cannot be shared, largely for safety and compliance reasons, but finding a safe way to mine data from patients can improve the quality of care and reduce costs; (2) analyzing the hospital system: by using big data, we can analyze the benefits of patient admission trends and so on; (3) managing data for use in public health research: big data analysis enables standardized integration of raw patient data; (4) protecting the patient’s identity: with big data analytics, healthcare fraudsters and identity thieves can be exposed; and (5) more efficient clinics: using big data can simplify workflows, transfer certain clinical tasks from doctors to nurses, reduce unnecessary tests, and improve patient satisfaction.

**Future research directions**

This section discusses some future research topics for the development of big data healthcare, including privacy preservation, data integration, and medical image processing.

With the increasing capacity of data sets in the cloud environment, privacy protection in big data analysis, sharing, and mining is a challenging research topic. Therefore, it is necessary to study the scalability of privacy protection in big data applications under cloud service access \(^{[105]}\). Additionally, Medical images in the cloud are often merged with images from other customers in a shared environment. The cloud-based medical image exchange has unique properties, which poses many security and privacy challenges in data design, image security, and so on. It also involves some legal problems, that is, regulatory compliance and auditing.

In the future, some research topics for different industries can be summarized as follows: Firstly, healthcare big data has two main exports of value, data connectivity and products that are integrated with new technologies. However, the construction of value closed loop also needs to consolidate the foundation of each link. Secondly, the analysis of medical big data requires response speed, responsiveness and accuracy of results, and enterprises still need to improve their technical capabilities. Thirdly, compliance problems exist in any link of medical big data collection, management and analysis, and relevant subjects need to pay attention to corresponding compliance obligations according to their business fields. Fourthly, in terms of investment, state capital plays a leading role and encourages the participation of social capital. From the enterprise side, the threshold for starting a medical big data business is high, and it needs to meet the four requirements of channel opening, strong data collection ability, excellent technical ability, and compliance.

**CONCLUSIONS**

This paper has provided an overview of some contents of healthcare big data. Firstly, we have summarized some kinds of medical big data, including the electronic health records, the medical image data, the health IoTs and healthcare informatics, the remote medical monitoring big data, the biomedical big data, and so
on. Afterward, some methods for handling different kinds of medical big data have been discussed. Additionally, we have analyzed the privacy of medical big data and summarized some methods and technologies to protect the privacy. Most importantly, we have discussed certain challenges and future directions for big data healthcare.

**DECLARATIONS**

**Authors’ contributions**
Made substantial contributions to conception and design of the study: Xu Z
Performed data acquisition, as well as completed the paper writing and modification work: Gou X

**Availability of data and materials**
Not applicable.

**Conflicts of interest**
Both authors declared that there are no conflicts of interest.

**Financial support and sponsorship**
This work was supported by National Natural Science Foundation of China (No. 71532007), China and the Postdoctoral Science Foundation (No. 2020M680151).

**Ethical approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Copyright**
© The Author(s) 2021.

**REFERENCES**

McKinsey Global Institute; 2011. DOI
16. Lissovoy G. Big data meets the electronic medical record: a commentary on "identifying patients at increased risk for unplanned readmission". Med Care 2013;51:759-60. DOI PubMed
27. Bairagi V, Sapkale A. Automated region-based hybrid compression for digital imaging and communications in medicine magnetic resonance imaging images for telemedicine applications. IET Sci Meas Technol 2012;6:247. DOI
41. Ma YJ, Zhang Y, Dung OM, Li R, Zhang DQ. Health internet of things: recent applications and outlook. Journal of Internet


45. Li F, Shankar A, Santhosh Kumar B. Fog-Internet of things-assisted multi-sensor intelligent monitoring model to analyse the physical health condition. Technol Health Care 2021. DOI PubMed


47. Lu RJ, Zeng B. A medical services middleware system for BAN. AMEII 2015: International Conference on Advances in Mechanical Engineering and Industrial Informatics. 2015. p. 550-5. DOI


70. Zeng LF, Meng CG, Li ZP, Huang XI, Liang ZH. Cloud computing and its decision-making for medical and health informatization in the context of big data. 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2014 Nov 2-5; Belfast, UK. IEEE; 2014. p. 201-6. DOI


72. Fan WW, Zhao DS, Wang SJ. A fast statistics and analysis solution of medical service big data. 2015 7th International Conference on
Information Technology in Medicine and Education (ITME); 2015 Nov 13-15; Huangshan, China. IEEE; 2015. p. 9-12. DOI


82. Saleem M, Kamdar MR, Iqbal A, Sampath S, Deus HF, Ngonza Ngomo A. Big linked cancer data: Integrating linked TCGA and PBMC. *Web Semant* 2014;27-28:34-41. DOI


92. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Curr Cardiol Rep* 2014;16:441. DOI PubMed


101. Wimmer H, Yoon VY, Sugumaravan V. A multi-agent system to support evidence based medicine and clinical decision making via data


106. Taneja H, Kapil, Singh AK. Preserving privacy of patients based on re-identification risk. Procedia Comput Sci 2015;70:448-54. DOI


