Gou *et al. J Smart Environ Green Comput* 2021;1:131-45
DOI: 10.20517/jsegc.2021.07

Journal of Smart Environments and Green Computing

Review

Open Access

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

How to cite this article: Gou X, Xu Z. An overview of Big Data in Healthcare: multiple angle analyses. *J Smart Environ Green Comput* 2021;1:131-45. https://dx.doi.org/10.20517/jsegc.2021.07

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)[12] 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:

$$\overbrace{\underset{(\textit{sensor and networks})}{\text{Connection}}\quad \underset{(\textit{computing and data on demand})}{\text{Cloud}}\quad \underset{(\textit{model and memory})}{\text{Cyber}}\quad \underset{(\textit{meaning and correlation})}{\text{Content/context}}\quad \underset{(\textit{sharing and collaboration})}{\text{Community}}\quad \underset{(\textit{personalization and value})}{\text{Customization}}}^{\text{6C system}}$$

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

$$\overbrace{\text{Connection}\quad \text{Conversion}\quad \text{Cyber}\quad \text{Cognition}\quad \text{Configuration}}^{\text{5C architecture}}$$

In 2011, the McKinsey Global Institute[13] 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[15-26], medical image[27-40], Internet of Things (IoT) and healthcare informatics[41-45], remote medical monitoring[46-48], and biomedical big data[49-57].

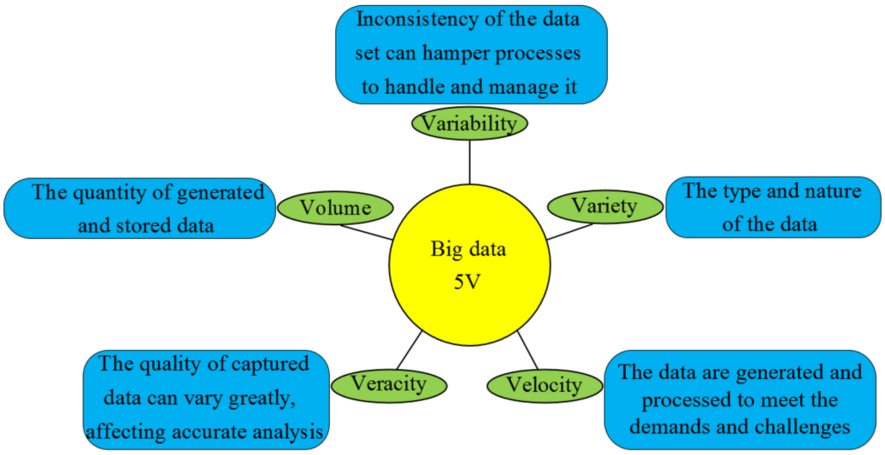

Figure 1. Five characteristics of big data.

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 the 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[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[16]. So far, several experts and scholars have researched the EHRs (or EMRs) from different angles, such as building model and developing algorithms[16-20], studying metadata and standards[15], natural language processing with EHRs[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[21]. Recently, Maheswaranathan *et al.*[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.*[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[24]. Furthermore, the rise of EMRs has led to the increasing of large-scale observational research on perioperative period[25]. However, the EHRs with patient treatments and outcomes are rich but underutilized information[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[27]. Additionally, medical image plays a pivotal role in surgery[28] and physicians' diagnostic decision-making[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[30]. Besides, a damage-free multi-component medical image compression approach was developed by Xin and Fan[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.*[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.*[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.*[11] analyzed all compression stages' effectiveness and the overall performance of the algorithm. Meanwhile, Ullah and Arslan[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.*[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[35], the eural-assisted image-dependent encryption scheme[36], and the novel framework for online medical image visualization based on shadow agent[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[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[39]. Furthermore, by the 5V features of medical imaging big data, Zhang[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[41]. The big data and IoT techniques can address the health care information processing challenges[42], have important impact and implications for health care delivery[43], and be applied in the evaluation process of sustainable smart city for real-time evaluation[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[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[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[47] and the remote real-time medical monitoring system architecture on the basis of IOT and cloud computing[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[49].

For the biomedicine with big data, some scholars developed methods to transform biomedicine such as the population approach (P5 Medicine)[50], the approach of adjusting covariables that affect features and/or goals in Tree-based Pipeline Optimization Tool[51]. Additionally, the application of biomedical image fusion in big data computing is developing strongly[52-54], such as the biomedical big data image fusion computing method based on spherical coordinates[52], and some big data technologies for the purpose of solving urgent problems of biomedical diagnostics[54]. Furthermore, some methods are also developed to improve the biomedical signal search results in big data, including the randomized Monte Carlo sampling method[55], the approach named bootstrapping for unified feature association measurement (BUFAM)[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[57].

Table 1. Some fields of health-IoT

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

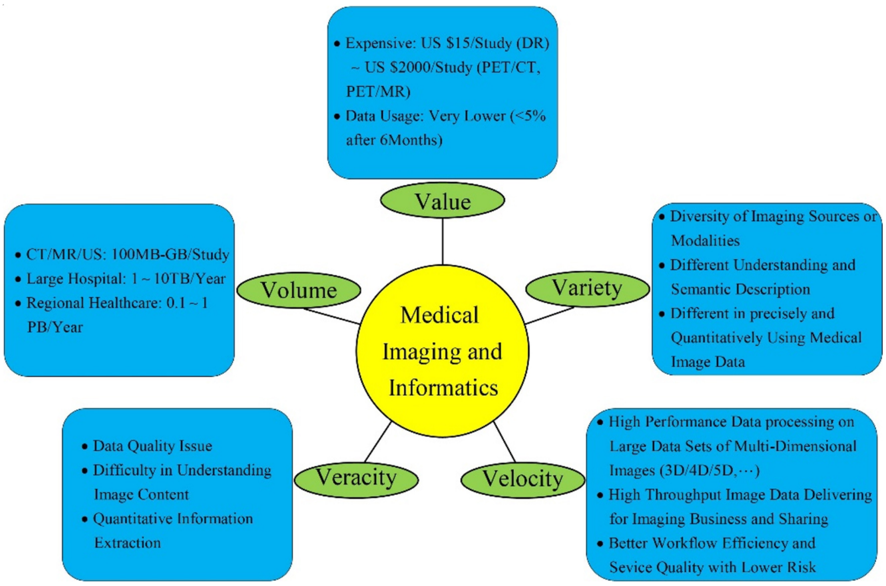

Figure 2. 5V features of medical imaging big data.

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[58]. In big data healthcare, data mining has important application potential in clinical medicine[59], 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[60], some data mining algorithms are developed to find the correlations between unilateral medical records[58]. 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[65,66] and Google[67], can be used to as data mining sources to deal with medical big data. For example, as a very popular information exchange platform, Twitter[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[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[68]. Specially, machine learning can serve as a useful tool to leverage complex clinical data and help guide critical clinical decisions[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[69].

Cloud computing

Cloud computing[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[70], and the home-diagnosis was proposed[71]. Moreover, cloud platforms[72], loud based healthcare systems[74], and similarity search-based clinical decision support systems[75] were structured.

Medical big data integration and clustering

In recent years, several methods[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[84], the hierarchical learning algorithm[85], and the feature selection algorithm[86].

Medical big data integration

Many integration methods have been developed, such as Big Linked Data[82] and G-DOC Plus[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[87]. Especially, Reina *et al.*[88] started the development of the integrated semantic framework for multidimensional data analysis. Based on the Hadoop platform, Lyu *et al.*[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.*[79] presented an integrated workflow which can be used to explore, analyze, and interpret translated medicine big data. Furthermore, Gligorijević *et al.*[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[81], and Merelli *et al.*[78] discussed three methods for data integration, including semantics, ontologies, and open format.

Medical big data clustering

Deng *et al.*[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[85-87,90], such as a hierarchical learning algorithm[85], the neural fuzzy Linguistic Classifier based on feature selection algorithm[86], the balanced clustering model for heterogeneous big data sets of intelligent healthcare[87], and some parts of the International Classification of Diseases Clinical Modification (ICD-CM)[90].

Personalized healthcare and precision medicine

Personalized healthcare[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[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[94]. Additionally, an ART framework was proposed for large-scale patient index in personalized healthcare[95]. More importantly, Poh *et al.*[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[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.*[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.*[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.*[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[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[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[105]. 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[116].

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 tele-cardiology, 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.

REFERENCES

1. Mashey JR. Big data and the next wave of infraS-tress. Slides from invited talk 1998. Available from: https://www.usenix.org/conference/1999-usenix-annual-technical-conference/big-data-and-next-wave-infrastress-problems. [Last accessed on 30 Aug 2021].
2. Lohr S. The origins of "big data": An etymological detective story. New York Times 2013. Available from: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/. [Last accessed on 30 Aug 2021].
3. Snijders C, Matzat U, Reips UD. "Big Data": Big gaps of knowledge in the field of Internet. *International Journal of Internet Science* 2012;7:1-5. DOI
4. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Ullah Khan S. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems* 2015;47:98-115. DOI
5. Wang H, Xu Z, Fujita H, Liu S. Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences* 2016;367-368:747-65. DOI
6. Xie W, Xu Z, Ren Z, Viedma EH. Restoring incomplete PUMLPRs for evaluating the management way of online public opinion. *Information Sciences* 2020;516:72-88. DOI
7. Wang H, Xu Z, Pedrycz W. An overview on the roles of fuzzy set techniques in big data processing: trends, challenges and opportunities. *Knowledge-Based Systems* 2017;118:15-30. DOI
8. Xu Z, Yu D. A Bibliometrics analysis on big data research (2009-2018). *J of Data, Inf and Manag* 2019;1:3-15. DOI
9. Cavanillas JM, Curry E, Wahlster W. New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe. Switzerland: Springer; 2016. DOI
10. Kong X, Feng M, Wang R. The current status and challenges of establishment and utilization of medical big data in China. *European Geriatric Medicine* 2015;6:515-7. DOI
11. Karimi N, Samavi S, Soroushmehr S, Shirani S, Najarian K. Toward practical guideline for design of image compression algorithms for biomedical applications. *Expert Systems with Applications* 2016;56:360-7. DOI
12. Hilbert M. Big data for development: a review of promises and challenges. *Dev Policy Rev* 2016;34:135-74. DOI
13. James M, Michael C, Jaques B, et al. Big Data: The next frontier for innovation, competition, and productivity. Washington:

McKinsey Global Institute; 2011. DOI

14. Huser V, Cimino JJ. Impending Challenges for the Use of Big Data. *Int J Radiat Oncol Biol Phys* 2016;95:890-4. DOI PubMed PMC
15. Sweet LE, Moulaison HL. Electronic Health Records Data and Metadata: Challenges for Big Data in the United States. *Big Data* 2013;1:245-51. DOI PubMed
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
17. Yang X, Zhang J, Chen S, Weissman S, Olatosi B, Li X. Comorbidity patterns among people living with HIV: a hierarchical clustering approach through integrated electronic health records data in South Carolina. *AIDS Care* 2021;33:594-606. DOI PubMed PMC
18. MacRae J, Darlow B, McBain L, et al. Accessing primary care Big Data: the development of a software algorithm to explore the rich content of consultation records. *BMJ Open* 2015;5:e008160. DOI PubMed PMC
19. Shen Y, Hsia T, Hsu C. Analysis of Electronic Health Records Based on Deep Learning with Natural Language Processing. *Arab J Sci Eng* 2021. DOI
20. Fan J, Chen M, Luo J, et al. The prediction of asymptomatic carotid atherosclerosis with electronic health records: a comparative study of six machine learning models. *BMC Med Inform Decis Mak* 2021;21:115. DOI PubMed PMC
21. Gai K, Qiu MK, Chen LC, Liu MQ. Electronic Health Record Error Prevention Approach Using Ontology in Big Data. 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems; 2015 Aug 24-26; New York, NY, USA. IEEE; 2015. p. 752-7. DOI
22. Maheswaranathan M, Chu P, Johannemann A, Criscione-Schreiber L, Clowse M, Leverenz DL. The impact of the COVID-19 pandemic and telemedicine implementation on practice patterns and electronic health record utilization in an academic rheumatology practice. *J Clin Rheumatol* 2021. DOI PubMed
23. Liu J, Ma J, Wang J, et al. Comorbidity analysis according to sex and age in hypertension patients in China. *Int J Med Sci* 2016;13:99-107. DOI PubMed PMC
24. Genta RM, Sonnenberg A. Big data in gastroenterology research. *Nat Rev Gastroenterol Hepatol* 2014;11:386-90. DOI PubMed
25. Ladha KS, Eikermann M. Codifying healthcare--big data and the issue of misclassification. *BMC Anesthesiol* 2015;15:179. DOI PubMed PMC
26. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. *IEEE J Biomed Health Inform* 2015;19:1193-208. DOI PubMed
27. Bairagi V, Sapkal 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
28. Plassard AJ, Kelly PD, Asman AJ, Kang H, Patel MB, Landman BA. Revealing latent value of clinically acquired CTs of traumatic brain injury through multi-atlas segmentation in a retrospective study of 1,003 with external cross-validation. *Proc SPIE Int Soc Opt Eng* 2015;9413:94130K. DOI PubMed PMC
29. Durand WM, Lafage R, Hamilton DK, et al; International Spine Study Group (ISSG). Artificial intelligence clustering of adult spinal deformity sagittal plane morphology predicts surgical characteristics, alignment, and outcomes. *Eur Spine J* 2021;30:2157-66. DOI PubMed
30. Zhang DG, Li WB, Liu S, Zhang XD. Novel fusion computing method for bio-medical image of WSN based on spherical coordinate. *J Vibroengineering* 2016;18:522-38. DOI
31. Xin G, Fan P. A lossless compression method for multi-component medical images based on big data mining. *Sci Rep* 2021;11:12372. DOI PubMed PMC
32. Kurc T, Qi X, Wang D, et al. Scalable analysis of Big pathology image data cohorts using efficient methods and high-performance computing strategies. *BMC Bioinformatics* 2015;16:399. DOI PubMed PMC
33. Ullah R, Arslan T. Parallel delay multiply and sum algorithm for microwave medical imaging using spark big data framework. *Algorithms* 2021;14:157. DOI
34. Zhang D, Wang X, Song X. New medical image fusion approach with coding based on SCD in wireless sensor network. *Journal of Electrical Engineering and Technology* 2015;10:2384-92. DOI
35. Huang X, Yi W, Wang J, Xu Z, Jiang Y. Hadoop-based medical image storage and access method for examination series. *Mathematical Problems in Engineering* 2021;2021:1-10. DOI
36. Lakshmi C, Thenmozhi K, Rayappan JBB, Rajagopalan S, Amirtharajan R, Chidambaram N. Neural-assisted image-dependent encryption scheme for medical image cloud storage. *Neural Comput & Applic* 2021;33:6671-84. DOI
37. Li W, Yu K, Feng C, Zhao D. SP-MIOV: A novel framework of shadow proxy based medical image online visualization in computing and storage resource restrained environments. *Future Generation Computer Systems* 2020;105:318-30. DOI
38. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. *Biomed Res Int* 2015;2015:370194. DOI PubMed PMC
39. Flechet M, Grandas FG, Meyfroidt G. Informatics in neurocritical care: new ideas for Big Data. *Curr Opin Crit Care* 2016;22:87-93. DOI PubMed
40. Zhang JG. Big data issues in medical imaging informatics. In: Conference on Medical Imaging-PACS and Imaging Informatics-Next Generation and Innovations; Orlando, FL, USA. 2015. p. 941803. DOI
41. Ma YJ, Zhang Y, Dung OM, Li R, Zhang DQ. Health internet of things: recent applications and outlook. *Journal of Internet*

Technology 2015;16:351-62. DOI

42. Almagrabi AO, Ali R, Alghazzawi D, AlBarakati A, Khurshaid T. A reinforcement learning-based framework for crowdsourcing in massive health care Internet of Things. *Big Data* 2021. DOI PubMed
43. Kelly JT, Campbell KL, Gong E, Scuffham P. The Internet of Things: impact and implications for health care delivery. *J Med Internet Res* 2020;22:e20135. DOI PubMed PMC
44. Nagarajan SM, Deverajan GG, Chatterjee P, Alnumay W, Ghosh U. Effective task scheduling algorithm with deep learning for Internet of Health Things (IoHT) in sustainable smart cities. *Sustainable Cities and Society* 2021;71:102945. DOI
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
46. Teng KA, Longworth DL. Personalized healthcare in the era of value-based healthcare. *Per Med* 2013;10:285-93. 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
48. Bouslama A, Laaziz Y, Tali A, Eddabbah M. AWS and IoT for real-time remote medical monitoring. *IJIE* 2019;6:369. DOI
49. Costa FF. Big data in biomedicine. *Drug Discov Today* 2014;19:433-40. DOI PubMed
50. Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW. Collaborative biomedicine in the age of big data: the case of cancer. *J Med Internet Res* 2014;16:e101. DOI PubMed PMC
51. Manduchi E, Fu W, Romano JD, Ruberto S, Moore JH. Embedding covariate adjustments in tree-based automated machine learning for biomedical big data analyses. *BMC Bioinformatics* 2020;21:430. DOI PubMed PMC
52. Erdman AG, Keefe DF, Schiestl R. Grand challenge: applying regulatory science and big data to improve medical device innovation. *IEEE Trans Biomed Eng* 2013;60:700-6. DOI
53. Bhuvaneshwar K, Belouali A, Singh V, et al. G-DOC Plus - an integrative bioinformatics platform for precision medicine. *BMC Bioinformatics* 2016;17:193. DOI PubMed PMC
54. Osipovich VS, Yashin KD, Dzik SK, Bykov AA. Fluorescent Nanoscale Structures for Selective Medical Diagnostics. In: Sontea V, Tiginyanu I, editors. 3rd International Conference on Nanotechnologies and Biomedical Engineering. Singapore: Springer; 2016. p. 414-8. DOI
55. Woodbridge J, Mortazavi B, Bui AA, Sarrafzadeh M. Improving biomedical signal search results in big data case-based reasoning environments. *Pervasive Mob Comput* 2016;28:69-80. DOI PubMed PMC
56. Chen H, Chen W, Liu C, Zhang L, Su J, Zhou X. Relational network for knowledge discovery through heterogeneous biomedical and clinical features. *Sci Rep* 2016;6:29915. DOI PubMed PMC
57. Hoffman S, Podgurski A. The use and misuse of biomedical data: is bigger really better? *Am J Law Med* 2013;39:497-538. DOI PubMed
58. Ramos-Miguel A, Perez-Zaballos T, Perez D, Falconb JC, Ramosb A. Use of data mining to predict significant factors and benefits of bilateral cochlear implantation. *Eur Arch Otorhinolaryngol* 2015;272:3157-62. DOI PubMed
59. Guo CH, Chen JF. Big data analytics in healthcare: data-driven methods for typical treatment pattern mining. *J Syst Sci Syst Eng* 2019;28:694-714. DOI
60. Behadada O, Trovati M, Chikh M, Bessis N. Big data-based extraction of fuzzy partition rules for heart arrhythmia detection: a semi-automated approach: A SEMI-AUTOMATED APPROACH. *Concurrency Computat : Pract Exper* 2016;28:360-73. DOI
61. Kitakaze M, Asakura M, Nakano A, Takashima S, Washio T. Data mining as a powerful tool for creating novel drugs in cardiovascular medicine: the importance of a “back-and-forth loop” between clinical data and basic research. *Cardiovasc Drugs Ther* 2015;29:309-15. DOI PubMed
62. Boytcheva S, Angelova G, Angelov Z, Tcharaktchiev D. Text mining and big data analytics for retrospective analysis of clinical texts from outpatient care. *Cybernetics and Information Technologies* 2015;15:58-77. DOI
63. Zhang HL, Zarei R, Pang C, Hu X. Discovering new analytical methods for large volume medical and online data processing. In: Zhang Y, Yao G, He J, Wang L, Smalheiser NR, Yin X, editors. Health information science. Cham: Springer International Publishing; 2014. p. 220-8. DOI
64. Choi JK, Jeon KH, Won Y, Kim JJ. Application of big data analysis with decision tree for the foot disorder. *Cluster Comput* 2015;18:1399-404. DOI
65. Beykikhoshk A, Arandjelović O, Phung D, Venkatesh S, Caelli T. Using Twitter to learn about the autism community. *Soc Netw Anal Min* 2015;5:22. DOI
66. Hays R, Daker-White G. The care.data consensus? *BMC Public Health* 2015;15:838. DOI PubMed PMC
67. Ramos-Casals M, Brito-Zerón P, Kostov B, et al. Google-driven search for big data in autoimmune geoepidemiology: analysis of 394,827 patients with systemic autoimmune diseases. *Autoimmun Rev* 2015;14:670-9. DOI PubMed
68. Bose I, Mahapatra RK. Business data mining - a machine learning perspective. *Information & Management* 2001;39:211-25. DOI
69. Soriano-Valdez D, Pelaez-Ballestas I, Manrique de Lara A, Gastelum-Strozzi A. The basics of data, big data, and machine learning in clinical practice. *Clin Rheumatol* 2021;40:11-23. DOI PubMed
70. Zeng LF, Meng CG, Li ZP, Huang XJ, 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
71. Lin W, Dou W, Zhou Z, Liu C. A cloud-based framework for Home-diagnosis service over big medical data. *Journal of Systems and Software* 2015;102:192-206. 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
73. Rajabion L, Shaltooki AA, Taghikhah M, Ghasemi A, Badfar A. Healthcare big data processing mechanisms: the role of cloud computing. *International Journal of Information Management* 2019;49:271-89. DOI
74. Sundharakumar K, Dhivya S, Mohanavalli S, Chander RV. Cloud based fuzzy healthcare system. *Procedia Computer Science* 2015;50:143-8. DOI
75. Tsymbal A, Meissner E, Kelm M, Kramer M. Towards cloud-based image-integrated similarity search in big data. IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); 2014 Jun1-4; Valencia, Spain. IEEE; 2014. p. 593-6. DOI
76. Ko KD, El-Ghazawi T, Kim D, Morizono H. Predicting the severity of motor neuron disease progression using electronic health record data with a cloud computing Big Data approach. 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology; 2014 May 21-24; Honolulu, HI, USA. IEEE; 2014. p. 1-6. DOI
77. Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China. *Quant Imaging Med Surg* 2014;4:426-9. DOI PubMed PMC
78. Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D. Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. *Biomed Res Int* 2014;2014:134023. DOI PubMed PMC
79. Satagopam V, Gu W, Eifes S, et al. Integration and visualization of translational medicine data for better understanding of human diseases. *Big Data* 2016;4:97-108. DOI PubMed PMC
80. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. *Proteomics* 2016;16:741-58. DOI PubMed
81. Mezghani E, Exposito E, Drira K, Da Silveira M, Pruski C. A semantic big data platform for integrating heterogeneous wearable data in healthcare. *J Med Syst* 2015;39:185. DOI PubMed
82. Saleem M, Kamdar MR, Iqbal A, Sampath S, Deus HF, Ngonga Ngomo A. Big linked cancer data: Integrating linked TCGA and PubMed. *Web Semant* 2014;27-28:34-41. DOI
83. Dhayne H, Haque R, Kilany R, Taher Y. In search of big medical data integration solutions - a comprehensive survey. *IEEE Access* 2019;7:91265-90. DOI
84. Deng Z, Zhu X, Cheng D, Zong M, Zhang S. Efficient k NN classification algorithm for big data. *Neurocomputing* 2016;195:143-8. DOI
85. Mei K, Peng J, Gao L, Zheng NN, Fan J. Hierarchical classification of large-scale patient records for automatic treatment stratification. *IEEE J Biomed Health Inform* 2015;19:1234-45. DOI PubMed
86. Azar AT, Hassanien AE. Dimensionality reduction of medical big data using neural-fuzzy classifier. *Soft Comput* 2015;19:1115-27. DOI
87. Li X, Jiao H, Li D. Intelligent medical heterogeneous big data set balanced clustering using deep learning. *Pattern Recognition Letters* 2020;138:548-55. DOI
88. Reina ST, Zamorano MR, Bjørnerud A. Towards an integrated semantic framework for neurological multidimensional data analysis. In: Ferrández Vicente JM, Álvarez-sánchez JR, de la Paz López F, Toledo-moreo FJ, Adeli H, editors. Artificial computation in biology and medicine. Cham: Springer International Publishing; 2015. p. 175-84. DOI
89. Lyu DM, Tian Y, Wang Y, Tong DY, Yin WW, Li JS. Design and Implementation of Clinical Data Integration and Management System Based on Hadoop Platform. 2015 7th International Conference on Information Technology in Medicine and Education (ITME); 2015 Nov 13-15; Huangshan, China. IEEE; 2015. p. 76-9. DOI
90. Pérez A, Gojenola K, Casillas A, Oronoz M, Díaz de Ilarraza A. Computer aided classification of diagnostic terms in spanish. *Expert Systems with Applications* 2015;42:2949-58. DOI
91. Gu J, Taylor CR. Practicing pathology in the era of big data and personalized medicine. *Appl Immunohistochem Mol Morphol* 2014;22:1-9. DOI PubMed PMC
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
93. Shin O, Han C, Pae CU, Patkar AA. Precision medicine for psychopharmacology: a general introduction. *Expert Rev Neurother* 2016;16:831-9. DOI PubMed
94. Chen Y, Elenee Argentinis JD, Weber G. IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther* 2016;38:688-701. DOI PubMed
95. Wang F. Adaptive semi-supervised recursive tree partitioning: The ART towards large scale patient indexing in personalized healthcare. *J Biomed Inform* 2015;55:41-54. DOI PubMed
96. Poh N, Tirunagari S, Windridge D. Challenges in designing an online healthcare platform for personalised patient analytics. 2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD); 2014 Dec 9-12; Orlando, FL, USA. IEEE; 2014. p. 1-6. DOI
97. Han H, Liu Y. Transcriptome marker diagnostics using big data. *IET Syst Biol* 2016;10:41-8. DOI PubMed
98. Schapranow MP, Kraus M, Perscheid C, Bock C, Liedke F, Plattner H. The Medical Knowledge Cockpit: Real-time analysis of big medical data enabling precision medicine. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2015 Nov 9-12; Washington, DC, USA. IEEE; 2015. p. 770-5. DOI
99. Noor AM, Holmberg L, Gillett C, Grigoriadis A. Big Data: the challenge for small research groups in the era of cancer genomics. *Br J Cancer* 2015;113:1405-12. DOI PubMed PMC
100. Issa AM, Marchant GE, Campos-Outcalt D. Big data in the era of precision medicine: big promise or big liability? *Per Med* 2016;13:283-5. DOI PubMed
101. Wimmer H, Yoon VY, Sugumaran V. A multi-agent system to support evidence based medicine and clinical decision making via data

sharing and data privacy. *Decision Support Systems* 2016;88:51-66. DOI
102. Mostert M, Bredenoord AL, Biesaart MC, van Delden JJ. Big data in medical research and EU data protection law: challenges to the consent or anonymise approach. *Eur J Hum Genet* 2016;24:956-60. DOI PubMed PMC
103. Saraladevi B, Pazhaniraja N, Paul PV, Basha MS, Dhavachelvan P. Big Data and Hadoop-a Study in Security Perspective. *Procedia Comput Sci* 2015;50:596-601. DOI
104. Gray EA, Thorpe JH. Comparative effectiveness research and big data: balancing potential with legal and ethical considerations. *J Comp Eff Res* 2015;4:61-74. DOI PubMed
105. Dou W, Zhang X, Liu J, Chen J. HireSome-II: towards privacy-aware cross-cloud service composition for big data applications. *IEEE Trans Parallel Distrib Syst* 2015;26:455-66. DOI
106. Taneja H, Kapil, Singh AK. Preserving privacy of patients based on re-identification risk. *Procedia Comput Sci* 2015;70:448-54. DOI
107. Mohammed N, Barouti S, Alhadidi D, Chen R. Secure and Private Management of Healthcare Databases for Data Mining. 2015 IEEE 28th International Symposium on Computer-Based Medical Systems; 2015 Jun 22-25; Sao Carlos, Brazil. IEEE; 2015. p. 191-6. DOI
108. Khan MFF, Sakamura K. Fine-grained access control to medical records in digital healthcare enterprises. 2015 International Symposium on Networks, Computers and Communications (ISNCC); 2015 May 13-15; Yasmine Hammamet, Tunisia. IEEE; 2015. p. 1-6. DOI
109. Srivathsan M, Arjun KY. Health Monitoring System by Prognotive Computing Using Big Data Analytics. *Procedia Comput Sci* 2015;50:602-9. DOI
110. Steinberg GB, Church BW, Mccall CJ, Scott AB, Kalis BP. Novel predictive models for metabolic syndrome risk: a "big data" analytic approach. *Am J Manag Care* 2014;20:221-8. PubMed
111. Mcgregor C. Wearable monitors on babies: Big data saving little people. IEEE International Symposium on Technology and Society (ISTAS) 2013; 203-203. Mcgregor C. Wearable monitors on babies: Big data saving little people. 2013 IEEE International Symposium on Technology and Society (ISTAS): Social Implications of Wearable Computing and Augmediated Reality in Everyday Life; 2013 Jun 27-29; Toronto, ON, Canada. IEEE; 2013. p. 203. DOI
112. Velsko S, Bates T. A conceptual architecture for national biosurveillance: moving beyond situational awareness to enable digital detection of emerging threats. *Health Secur* 2016;14:189-201. DOI PubMed
113. Milicchio F, Rose R, Bian J, Min J, Prosperi M. Visual programming for next-generation sequencing data analytics. *BioData Min* 2016;9:16. DOI PubMed PMC
114. Suresh S. Big data and predictive analytics applications in the care of children. *Pediatr Clin North Am* 2016;63:357-66. DOI PubMed
115. Waldman SA, Terzic A. Big data transforms discovery-utilization therapeutics continuum. *Clin Pharmacol Ther* 2016;99:250-4. DOI PubMed PMC
116. Lu J, Zhou J, Ruan H, Luo G. Establishing a university library-based health information literacy service model in the age of big data. *J Med Imaging Health Inform* 2016;6:260-3. DOI
117. Liang Y, Guo N, Xing C, Zhang Y, Guo C. Chronic knowledge retrieval and smart health services based on big data. In: Zheng X, Zeng DD, Chen H, Leischow SJ, editors. Smart health. Cham: Springer International Publishing; 2016. p. 231-40. DOI
118. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and Disease Prevention: From Quantified Self to Quantified Communities. *Big Data* 2013;1:168-75. DOI PubMed