A Systematic Review of Big Data Potential to Make Synergies between Sciences for Achieving Sustainable Health: Challenges and Solutions

Shirin AYANI[1], Khadijeh MOULAEI[2,*], Sarah DARWISH KHANEHSARI[1], Maryam JAHANBAKHSH[3], Faezeh SADEGHI[4]

[1] Rayavaran Medical Informatics Company, Smart Hospital and Telemedicine Research Center, Tehran, Iran.
[2] Iran University of Medical Sciences, School of Management and Medical Information, Khadijeh Moulaei, No.2, Corner of Hamsayegan Street, Valiasr Ave, Tehran, Iran.
[3] Isfahan University of Medical Sciences, Isfahan, Iran.
[4] Faran (Mehr Danesh) Non-governmental Institute of Virtual Higher Education, Department of English Language, Tehran, Iran.
E-mail: Moulaei.kh@tak.iums.ac.ir

* Author to whom correspondence should be addressed; Tel: +982188783115; Fax: +982188661654

Received: January 9, 2019 /Accepted: May 16, 2019/ Published online: June 30, 2019

Abstract
The importance of the healthcare industry, benefiting from the synergies between sciences, adds to the necessity of discovering knowledge, which is achievable with big data analytics tools. The purpose of this article is to examine the challenges and provide solutions for using big data in the healthcare industry. The methods of this article are derived from PRISMA guidelines and its models. A variety of databases and search engines including PubMed, Scopus, Elsevier, IEEE, Springer, Web of Science, Proquest, and Google Scholar were searched according to credible keywords. The results of the present study showed that the problems associated with the use of big data in the healthcare industry could be classified in four groups including "data gathering, storage and integration", "data analysis", "knowledge discovery and information interpretation", and "infrastructure". Although the results point to a high frequency of challenges in the "data gathering, storage and integration" group, the greatest weight of problems, due to their importance, appears to be visible in the "infrastructure" group. Considering the numerous benefits of using big data, it is imperative to identify the challenges and resolve them accurately. It is expected that all the barriers can be removed soon. Big data analytics tools will be able to offer the best possible strategies based on human individual and social conditions in the context of artificial intelligence methods.

Keywords: Big data; Data analysis; Data integration; Internet of things; Medical informatics; Biological informatics

Introduction

At the end of the 1990s, in order to make the right decisions and gain a better understanding of market behaviors, the role of gathering data, integrating and interpreting business information was emphasized by the researchers. For this purpose, the term "Big Data" was introduced by Michael Cox and David Ellsworth in 1997 [1]. Big data is referred to as a set of data whose volume is beyond the capabilities of current databases and technologies. Therefore, in order to analyze these data, databases with volume capabilities higher than terabyte and Exabyte are needed [2].

Big data separating factors from other data include volume (scale and size of data in storage), velocity (the speed in which this data is generated, produced, created, refreshed, and streamed), variety (multiple different forms of the data), veracity (uncertainty of the data that leads to confidence or trust in the data), and value (deriving business value and insights from the data) [3-6].

Additionally, the outcomes of big data analyses contribute to the identification of unknown patterns that show the causal relationships between different events in a wide range of information in the real world and ends in knowledge production [7].

Before the introduction of the concept of big data, with the emergence of information revolutionary age, it was possible to collect the data associated with healthcare activities in related centers [8], and healthcare providers who exploited health information systems like hospital information systems (HISs) started generating massive data [9]. While the information systems were used, specialists' level of expectation went up, and another need was formed: how to understand the multidimensional causal relationships associated with individual and social health. Simultaneously, big data was introduced to the world and health researchers showed interest in this field [10-12]. Big data in the healthcare industry refers to a set of data related to diagnosis and treatment of diseases, contagious diseases, nutritional status, climate, political status, security of a country (especially war conditions), cultural status, social system, regional and vernacular status, metabolism and micronutrients (ions), genetics and cells, the economic status, the insurance companies' bills and other things [1, 10, 13-15]. To collect these data, equipment, and tools such as the Internet, smartphones, social media, sensors and databases - which are related to the scientific societies and hospitals- are used as well as clinical and hospital information systems [14, 16]. After gathering data, it was possible to discover unknown patterns associated with some features carried out by the use of advanced analytics tools. These features that are used by advanced analytics tools include individual and social disease management, changes in habits and pathogenic conditions, prevention, diagnosis and treatment of diseases especially rare diseases, forecasting, individualization of health services, support and supervision of health social services [14]. One of the valuable advantages of analyzing data in the healthcare industry is the knowledge discovery beyond researchers' imagination, which ends in the successful medical decision making of healthcare providers and producing clinical decision support systems [3, 17, 18]. Analysis of this data is done by the use of particular computing technologies, which requires specific hardware structures and operating platforms. At this time, operating platforms, hardware structures, and advanced technologies for using big data are acceptably reachable. Extensible Markup Language (XML), Web Services, Database Management Systems, Hadoop, SAP HANA and analytical software, are their examples [1, 14, 19, 20]. On the other hand, because of the considerable data volume, it is impossible to store and transfer data using traditional methods and technologies. Today, SQL, MySQL, and Oracle databases are widely used for implementing information systems, while for storing large data, Apache and Nosql databases are required [1].

The big data formation is performed in three stages of collecting, processing and visualizing data [15, 21]. Each step is accompanied by significant challenges that prevent successful implementation of big data operationally. Therefore, the purpose of this study was to survey the applications of big data in the healthcare industry in order to increase the synergy, as well as achieve sustainable health to survey challenges and provide suitable solutions.

Material and Method

The methods of this article are derived from PRISMA guidelines and its models (for further information see www.prisma-statement.org).

Information Resources

A variety of databases and search engines including Pubmed, Scopus, Elsevier, IEEE, Springer, Web of Science, Proquest and Google Scholar were searched according to credible keywords and the pre-specified search strategy mentioned below. The databases were searched from May 24, 2018, to July 30, 2018.

Keywords and Search Strategy

The keywords of this research used in the search strategy are as follows: Big Data, Data Sets, Big Data Analytics, Big Data Analytics Tools, Administrative Data, Structured Data, Unstructured Data, Business Process Analytics, Real-time Analytics, Information Technology Management, Health Care, Health Care Industries, E-Health Solutions, Social Health, Clinical Registries, Bioinformatics, Health Informatics, Medical Informatics, Sensor Informatics, Challenge, Solution, Problems.

The applied search strategy was [Big data* AND (Business Process OR Data sets OR Real-time OR Administrative Data OR Unstructured Data OR Structured Data OR Information Technology Management OR Resource-based Theory) AND (Solutions* OR Challenges* OR Problems*) AND (Healthcare OR E-Health OR Social Health OR Public Health OR Clinical Registries OR Medical OR *Informatics OR Bioinformatics)].

Inclusion, Exclusion and Data Extraction

The inclusion criteria were as follows:

- Full-text resources were available.
- The articles were published in the last 10 years.
- The articles were published in scientific and high-ranking journals.
- Big data challenges and their potential solutions in the healthcare industry were suggested.

The exclusion criteria were as follows:

- The challenges were unrelated to the big data in the healthcare industry.
- The definitions were not clear and related to the challenges and their solutions.
- Concerning challenges and their solutions, the articles were not comprehensive.

First, all the challenges introduced in the selected articles were extracted. The challenges were categorized into four groups: "data gathering, storage and integration", "data analysis", "knowledge discovery and information interpretation" and "infrastructure".

Then, to find or propose solutions for each challenge, the necessary examinations were carried out and solutions were put in different groups along with their related challenges.

Results

The search results are shown in Figure 1, and the results of these studies are illustrated in Table 1. In this table, the problems were grouped, and solutions to each problem were specified.

Data Gathering, Storage and Integration

Over the years, the volume of generated data has been increased significantly by healthcare organizations. These data are collected from various sources as well as by various tools and technologies such as information systems, cell phones, wireless sensors, RFID (radio frequency identification) and so on [22, 25]. Therefore, in order to create big data in the healthcare industry, heterogeneous sources and different formats are used [28]. For this reason, it is normal to face some problems such as noise, confounding factors, and inconsistencies in the gathered data collection [25]. Also, during data gathering, for some reasons such as inadequate storage space and gathering data from various sources, some valuable data can be possibly ignored or removed [22].

Ladha and his colleagues argued that lost data might cause the creation of invalid patterns. Therefore, three potential constraints in gathering data should be taken into account. First, some data may be missing or artificial. Second, in some cases, some found data could lead to the definition of ambiguous or contradictory variables. Third, some variables can act as confounding factors during the analysis. The measure of the error rate in each one of these three stages indicates the ineffectiveness of big data and presents the risk of its use by conducting scientific researches [11].

It is noteworthy that in order to avoid gathering data redundancy, it is necessary to identify and provide methods to prevent data abundance and redundant data storage [35].

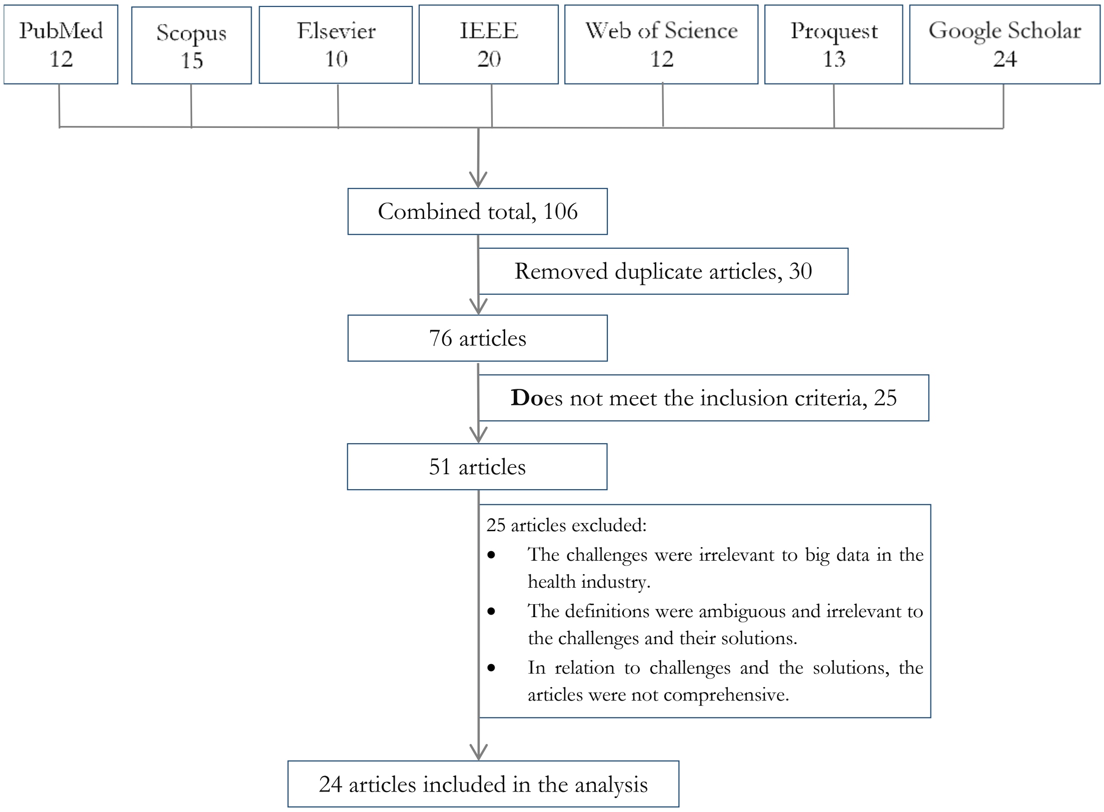

Figure 1. Literature search criteria with inclusion and exclusion criteria

3.2. Data Analysis

Analyzing big data, some errors occurring during the data gathering and those errors that were hidden in the databases were identified and corrected as much as possible. Current information systems in the healthcare industry are not integrated, so by gathering data in a big data repository, these errors can be identified [39]. Because of this and due to the lack of practical methods for accurate and rapid processing of massive volume of data, big data analysis brings up a critical issue [32]. However, since analyzing these health data leads to significant worthwhile outcomes, many researchers are trying to overcome these challenges [40, 41]. For example, Mathew and Pillai described the SAP HANA platform in data analysis as highly useful. SAP HANA utilizes some data mining methods for analyzing complex and large volumes of data[3]. It is also imperative to use some techniques such as networks, graphs, and charts for data analysis [42-44]. Note that if no pattern is extracted, it is necessary to re-formulate and repeat the analysis step[14]. The system analyst's skill in recognizing the patterns, the definition of the rules, process modeling, error detection, and setting error threshold are important [45]. Evaluating the results of processing big data in order to validate the acquired patterns is also a significant challenge that must be done by professional multidisciplinary teams [1].

Another critical point is the need for a remarkable space of temporary memory that is accessible on unique hardware platforms while analyzing the data [13].

Table 1. Big data challenges and their solutions in the healthcare industry

ID	Challenge Group	Challenge	Solution	Reference
1	Data gathering, storage and integration	Difficulty in gathering and integrating data	Creating distributed databases which are interrelated and choosing a scientific manner to gather data from health centers nationally	[2,22,23]
		Lack of time priority in gathering data	Having data gathering patterns so that data interconnections are considered	[11,18]
		Data ambiguity	Using various scientific techniques to clarify data and code them and defining flexible formats for high-clarity data gathering	[13]
		Heterogeneity of different sources of gathered data	Using semantic networks and ontological interpretations	[10,24,25]
		Artificial nature of some data elements due to the loss of other relevant data	Using Loss data analysis and providing clear definitions of variables	[11,26,27]
		Existence of noisy data	Using preprocessing techniques (like PCA)	[26]
		Massive data	Centralized data monitoring, preventing redundancy and filtering unnecessary data	[17,28,29]
2	Data analysis	Difficulty in analyzing big data	Using advanced computing technologies and processing parallelism	[30,31]
		Difficulty in extracting patterns and models	Using advanced data mining techniques to obtain valuable models and patterns	[13,23]
		Uncertainty about the accuracy of extracted information, patterns, and models	Evaluating is done to confirm the accuracy and value of the model	[13,23]
3	Knowledge discovery and information interpretation	Interpretation of patterns	Information interpretation after getting help from experts in multidisciplinary or interdisciplinary fields	[1]
		Difficulty in representing knowledge	Defining flexible formats to represent the interpreted information. The multidisciplinary or interdisciplinary specialists should document extracted knowledge	[23]
		Studying the generalizability of knowledge and the accuracy of explicit knowledge	The validity of explicit knowledge should be confirmed statistically and epidemiologically	[26]
4	Infrastructure	Absence of specific rules and standards	Defining a set of rules in the form of specific frameworks to achieve standards and to apply the rules in a correct way	[13,32,33]
		Immaturity of required infrastructure	Identifying and implementing modern technologies and then complying the existing infrastructure with them	[25,27,36]
		Lack of a stakeholder	Making collective business policies for applicability of big data	[32, 34]
		Data security (availability, confidentiality, and integrity]	Using advanced security standards and related technologies and then continuing the monitoring over data accuracy and quality	[23,35,36]
		Lack of proper bandwidth for data transfer	Using high bandwidth or special data transfer protocols	[26, 37]
		Lack of some important information systems for storing data digitally as in Electronic Medical Records	Applying information systems like transaction processing systems, registration systems, and decision support systems	[38]

Knowledge Discovery and Information Interpretation

One of the major challenges of big data is the interpretation of information and patterns after the analysis. At this point, the leading question is how new knowledge can be obtained from the aggregation of information; then how it can be documented, displayed, and verified [46].

Knowledge discovery during data analysis based on the Internet of Things is one of the biggest challenges that professionals encounter. Devices and software based on the Internet of Things generate substantial data streams, which result in the mass production of data. Researchers can use Artificial Intelligence and Machine Learning to interpret this information. Machine Learning algorithms and Intelligent Computing are considered achievable solutions for analyzing big data based on the Internet of Things [22, 47]. On the other hand, the simple representation of the knowledge extracted from big data is a serious issue. If it is not possible to demonstrate the novel knowledge, it is impossible to develop and apply it. Therefore, to eliminate the complexity of the discovered patterns and create relations between them, experts' opinions in different subjects areas and from other perspectives are needed [35]. Identifying invalid patterns and accrediting extracted knowledge also needs multidisciplinary expert teams. Since a pattern may be taken from a specialized field to the others, it is not easy to perceive the relationships between them and to evaluate the results [35, 48].

Infrastructure

Infrastructure refers to the use of a combination of hardware, software, and services that should be robust, supportive, and scalable [49]. The infrastructure of big data refers to all cases that can support its lifecycle on a large scale over time. In the infrastructure of big data, it is essential to be ensured of data security (including confidentiality, integrity, and availability) [50, 51). There are many challenges in healthcare that make it impossible to create a secure infrastructure; for example, high bandwidth is effective to the speed of data stream [32]. Communication channels based on high-bandwidth help gathering and managing data and protecting its security [52].

On the other hand, storing and retrieving data from clinical and hospital information systems is a very complex, time-consuming, expensive endeavour, which requires a robust infrastructure. Databases and computing systems such as Mongo, Hadoop, and MapReduce can provide proper infrastructure for applying big data[53]. Concerning the characteristics of health data, the Mongo database can provide high performance; accessibility and scalability for big data. This can successfully create data repositories [54, 55].

The conceptual framework of big data analysis in the healthcare industry varies from the traditional data analysis frameworks; so, processing should be distributed and performed all across the nodes of the network. Therefore, instead of using a machine, the processing is broken down and carried out by variant machines, and their analytics can be performed in parallel with the help of MapReduce [56, 57]. Besides, open source platforms such as Hadoop, which operates in cloud space, support the use of big data analysis in the healthcare industry [13, 19].

One of the other challenges is the absence of national and international laws and standards, which if available, could guarantee the success of the work [33]. On the other hand, the absence of information systems that record events causes the unsuccessful gathering of required data. Therefore, it seems necessary to develop these systems, especially Electronic Medical Records [58].

Ultimately, what was learned from the whole study reflects the fact that although the results of studying the four groups of big data challenges represent high frequency in the "data gathering, storage and integration" group, the greatest weight of problems, due to their importance, appears to be visible in the "infrastructure".

The study found that many researchers are trying to discover health-related causal relationships from big data by the use of artificial intelligence techniques. The discovery of these relationships is expected to significantly increase the human ability to control the risk factors associated with individual and social health.

Discussion

Different studies have pointed to a variety of challenges of big data, and they have been classified from different perspectives. Acharjya and Kauser stated that for surveying these challenges, it is essential to know different types of computational complexity, information security, and computational methods of data analysis [22]. In a study, Andreu Perez and his colleagues also emphasized the challenges of privacy, security, data ownership, stewardship and data governance [2]. Huffman declared that vendors, healthcare providers, and government officials must carefully consider the big data challenges and design appropriate strategies. This will cause progress in all science areas, especially in healthcare [59]. Nasser and Tariq divided big data challenges based on their "life cycles" into three distinct categories: "data", "process" and "management.

They associated the challenges of the "data" category with data volume, variety, velocity, veracity, volatility, quality, discovery, and dogmatism. They also ascribed the challenges of the "process" category as: to how to capture data, how to integrate data, how to transform data, how to select the right model for analysis and how to provide the results. Eventually, privacy, security, governance, and ethical aspects were grouped into the "management" category [29]. Philip and his colleagues argued that challenges and opportunities come together and the challenges associated with big data will bring many attractive opportunities in the future. They categorized big data challenges into data capture, storing, searching, sharing, analyzing and visualizing categories [35].

According to this study, researchers found that the categorizations in other studies are too general. Moreover, there was no study exactly addressing the challenges and proposing related solutions. Therefore, in order to achieve a thorough understanding of the big data implementation problems, at first, the challenges have been identified. Then, after complementary studies about each of the challenges, a suitable solution was found or proposed. Note that in order to resolve many of these challenges, no precise executive solution has been suggested [1, 2, 11, 14, 17, 32]. Thus, most of the solutions represent some ideas and approaches that are expected to be achievable with current or future technologies and tools.

The results of the present study showed that the problems associated with the use of big data in the healthcare industry could be classified in four groups including "data gathering, storage and integration", "data analysis", "knowledge discovery and information interpretation", and "infrastructure". Although the results point a high frequency of challenges in the "data gathering, storage and integration" group, the greatest weight of problems, due to their importance, appears to be visible in the "infrastructure" group. Considering the numerous benefits of using big data, it is imperative to identify the challenges and resolve them accurately.

In connection with the importance of fixing the problems of using Big Data and benefiting from their advantages, the following items can be mentioned:

Human knowledge improvement has identified the interconnection of sciences and resulted in the creation of multidisciplinary sciences. It is anticipated that in the future, multidisciplinary sciences will revolutionize the world, and they can lead to the emergence of the knowledge age [60]. Multidisciplinary sciences focus on the data stream from different scientific areas to each other. These data are originated from effective factors that have come up from a scientific field and affect others [61]. In this regard, the following examples can be mentioned: differences in lifestyle and social conditions and their effects on diseases or their treatments, differences in geographical conditions and their effects on treatment and care. Some other items include wars and their long-term effects on the health, psychological conditions in raising children and their outcomes in adolescence and middle age as well as controlling chronic diseases in particular conditions, activation of a defective gene and its role in the development of the disease with regard to environmental and nutritional status and so on [62-67]. In such situations, knowledge discovery and understanding the effects of the factors are very complicated. Hence, achieving the great human ideal, which is understanding unknowns and discovering facts, can be fulfilled by the use of big data [6, 10]. Big data analytics tools are valuable to identify the relationship between sciences through knowledge discovery. It represents the actual movements of a scientific field at one point in the world and its effects on other areas of science in other parts of the world [6, 35, 68].

The relation between the sciences can be achieved by technological convergence that results in the synergy of sciences like NBIC containing four different sciences (Nanotechnology, Biotechnology, Information Technology, and Cognitive Science) [69, 70]. The technological convergence is important for some reasons including developing a person's perception of reality, creating robust technology platforms for health improvement and diseases diagnosis and treatment [10, 71, 72]. In order to achieve a technological convergence, the potentiality of big data potential that depicts the connection between different sciences should not be ignored.

On the other hand, big data is capable of detecting and displaying multilevel molecular and genetic interatomic. Epigenetic knowledge can open another perspective for the researchers and help them explain epigenetic rules and the effects on gene expression. Furthermore, big data can help to perceive the functions of body organs "as one of the best extraordinary biological systems in the world" [73, 74].

Another important point is that although clinical trials are the golden standards in medical science to determine the causality, clinical trial challenges provide an opportunity to use big data. Clinical trials are morally questionable, expensive, and time-consuming. Bias and subjective attitudes can accompany them, and in many cases are not technically supportable, while the use of big data provides researchers with the opportunity to design and develop research hypotheses after considering disease patterns and their commonalities and to identify causal relationships, despite the challenges of clinical trials [11].

From another perspective, the big data analytics tools can be considered as an appropriate simulator for some health information systems that monitors and controls the influential factors. For example, the Pharmacovigilance information system has been implemented to monitor drug effects and determine adverse reactions [75, 76]. In order to run this information system, governments, some organizations and specialized users are busy taking actions in some areas of the world. If big data fed by required distributed subsystems exist, it is potential to simulate Pharmacovigilance results.

At last, the existence of big data is essential for pure and applied researches. Most studies have sufficed to examine the existing big data challenges, some solutions have only been suggested theoretically [3, 11, 17, 32, 38]. Considering the huge benefits of big data, it is imperative to identify the challenges and resolve them accurately. It is expected that with modern technologies, all the barriers can be removed in near future and big data analytics tools, utilizing artificial intelligence, will be able to offer the best possible strategies based on social and individual conditions.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The researchers of this study would like to thank Dr. Rafat Bayat and the educational and research staff of Faran (Mehr Danesh) Non-governmental Institute of Virtual Higher Education for the financial and spiritual support of this research.

References

1. Cox M, Ellsworth D (Eds.). Application-controlled demand paging for out-of-core visualization. VIS '97 Proceedings of the 8th conference on Visualization '97 (Cat No 97CB36155); 1997: IEEE. 97-010, July 1997.
2. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ. Big Data for Health. IEEE Journal of Biomedical and Health Informatics 2015;19(4):1193-1208.

3. Mathew PS, Pillai AS (Eds.). Big Data solutions in Healthcare: Problems and perspectives. 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS); 2015, 19-20 March.
4. Lee I. Big data: Dimensions, evolution, impacts, and challenges. Business Horizons 2017;60(3):293-303.
5. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015;35(2):137-144.
6. Sagiroglu S, Sinanc D, editors. Big data: A review. Collaboration Technologies and Systems (CTS), 2013 International Conference; 2013: IEEE. 20-24 May 2013.
7. Chen C, Ma J, Susilo Y, Liu Y, Wang M. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation Research Part C: Emerging Technologies 2016;68:285-299.
8. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. British Dental Journal 2008;204(6):291-295.
9. Li J-S, Zhang Y-F, Tian Y. Medical Big Data Analysis in Hospital Information System. Big Data on Real-World Applications. InTech; 2016. Available at: https://www.intechopen.com/books/big-data-on-real-world-applications/medical-big-data-analysis-in-hospital-information-system (accessed December 10, 2018)
10. Wang Y, Hajli N. Exploring the path to big data analytics success in healthcare. Journal of Business Research 2017;70:287-299.
11. Ladha KS, Arora VS, Dutton RP, Hyder JA. Potential and Pitfalls for Big Data in Health Research. Advances in Anesthesia 2015;33(1):97-111.
12. Riabacke M, Danielson M, Ekenberg L. State-of-the-art prescriptive criteria weight elicitation. Advances in Decision Sciences 2012;2(3):1-24.
13. Wassan JT. Big Data Paradigm for Healthcare Sector. Maitreyi College, University of Delhi, India: Maitreyi College; 2016.
14. Huang T, Lan L, Fang X, An P, Min J, Wang F. Promises and Challenges of Big Data Computing in Health Sciences. Big Data Research 2015;2(1):2-11.
15. Koufi V, Malamateniou F, Vassilacopoulos G. A Big Data-driven Model for the Optimization of Healthcare Processes. Stud Health Technol Inform. 2015;2(10):697-701.
16. Bhatt CM, Dey N, Ashour A. Internet of Things and Big Data Technologies for Next Generation Healthcare. Springer; 2017.
17. Liu W, Park EK, editors. Big Data as an e-Health Service. 2014 International Conference on Computing, Networking and Communications (ICNC); 2014 3-6 Feb 2014.
18. Belle A, Thiagarajan R, Soroushmehr S, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. BioMed Research International. 2015;4(12):20-30.
19. Zikopoulos P, Eaton C. Understanding big data: Analytics for enterprise class hadoop and streaming data: McGraw-Hill Osborne Media; 2011.
20. Färber F, May N, Lehner W, Große P, Müller I, Rauhe H, et al. The SAP HANA Database--An Architecture Overview. IEEE Data Eng Bull.2012;35(1):28-33.
21. Janke AT, Overbeek DL, Kocher KE, Levy PD. Exploring the potential of predictive analytics and big data in emergency care. Annals of Emergency Medicine 2016;67(2):227-236.
22. Acharjya DP, Kauser Ahmed P. A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. (IJACSA) International Journal of Advanced Computer Science and Applications 2016;7(2):512-8.
23. Cortés R, Bonnaire X, Marin O, Sens P. Stream Processing of Healthcare Sensor Data: Studying User Traces to Identify Challenges from a Big Data Perspective. Procedia Computer Science 2015;52(1):1004-1009.
24. Khan N, Yaqoob I, Hashem IAT, Inayat Z, Mahmoud Ali WK, Alam M, et al. Big Data: Survey, Technologies, Opportunities, and Challenges. The Scientific World Journal 2014;2014(2014):18-24.
25. Fan J, Han F, Liu H. Challenges of Big Data analysis. National Science Review 2014;1(2):293-314.

26. 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 Research International 2014;8(6):1-13.
27. Groenwold RH, Donders AR, Roes KC, Harrell FE, Jr., Moons KG. Dealing with missing outcome data in randomized trials and observational studies. Am J Epidemiol. 2012;175(3):210-7.
28. Tole AA. Big data challenges. Database Systems Journal 2013;4(3):31-40.
29. Nasser T, Tariq R. Big data challenges. Computer Engineering & Information Technology 2015;4(3):1-3.
30. Marx V. Biology: The big challenges of big data. Nature 2013;498(7453):255-260.
31. Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big data analytics: a survey. Journal of Big Data 2015;2(1):21-30.
32. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2(3):1-10.
33. Schultz T. Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle. Bulletin of the Association for Information Science and Technology 2013;39(5):34-40.
34. Chaudhari N, Srivastava S (Eds.). Big data security issues and challenges. 2016 International Conference on Computing, Communication and Automation (ICCCA); 29-30 April 2016.
35. Philip Chen CL, Zhang C-Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 2014;275:314-47.
36. Terzi DS, Terzi R, Sagiroglu S (Eds.). A survey on security and privacy issues in big data. 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST);14-16 Dec 2015.
37. Javadi B, Zhang B, Taufer M, editors. Bandwidth Modeling in Large Distributed Systems for Big Data Applications. 2014 15th International Conference on Parallel and Distributed Computing, Applications and Technologies 2014, pp. 9-11.
38. Kuziemsky CE, Monkman H, Petersen C, Weber J, Borycki EM, Adams S, et al. Big Data in Healthcare – Defining the Digital Persona through User Contexts from the Micro to the Macro: Contribution of the IMIA Organizational and Social Issues WG. Yearbook of Medical Informatics 2014;9(1):82-9.
39. Bologa A-R, Bologa R, Florea A. Big data and specific analysis methods for insurance fraud detection. Database Systems Journal 2013;4(4):30-39.
40. Kam HJ, Kim JA, Cho I, Kim Y, Park RW. Integration of heterogeneous clinical decision support systems and their knowledge sets: feasibility study with drug-drug interaction alerts. AMIA Annual Symposium Proceedings 2011;2011(1):664-73.
41. Zhang Z, Sarcevic A, An Y. A prototype system for heterogeneous data management and medical devices integration in trauma resuscitation. iConference Proceedings 2013;3(6):1-5.
42. Lee J, Kwon YS, Färber F, Muehle M, Lee C, Bensberg C, et al., editors. SAP HANA distributed in-memory database system: Transaction, session, and metadata management. Conference on Data Engineering (ICDE), 29th International Conference, 8-12 April 2013
43. Morgen C, editor Empowering SAS. Users on the SAP HANA Platform. Proceedings of the SAS Global Forum 2014. International Conference, 8-12 April 2014
44. Rudolf M, Paradies M, Bornhövd C, Lehner W, editors. The Graph Story of the SAP HANA Database. BTW 2013;4:403-20.
45. Abdullah N, Ismail SA, Sophiayati S, Sam SM. Data quality in big data: a review. International Journal of Advances in Soft Computing & Its Applications. 2015;7(3):16-2.
46. Das T, Kumar PM. Big data analytics: A framework for unstructured data analysis. International Journal of Engineering Science & Technology 2013;5(1):153.
47. Chen X-Y, Jin Z-G. Research on Key Technology and Applications for Internet of Things. Physics Procedia 2012;33:561-566.
48. Subasinghe K, Kodithuwakku S (Eds.). A big data analytic identity management expert system for social media networks. Conference on Electrical and Computer Engineering (WIECON-ECE), 2015 International WIE, 19-20 Dec 2015.

49. Schoenborn B. Big Data Analytics Infrastructure for DUMMIES. United States of America: John Wiley & Sons, Inc; 2014.
50. Demchenko Y, Ngo C, Membrey P. Architecture framework and components for the big data ecosystem. Journal of System and Network Engineering 2013;4(7):1-31.
51. Moreno J, Serrano MA, Fernández-Medina E. Main Issues in Big Data Security. Future Internet 2016;8(3):44.
52. Wasan SK, Bhatnagar V, Kaur H. The impact of data mining techniques on medical diagnostics. Data Science Journal 2006;5(2):119-26.
53. Nandimath J, Banerjee E, Patil A, Kakade P, Vaidya S, Chaturvedi D, editors. Big data analysis using Apache Hadoop. Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on; 20 July 2013.
54. Prasad S, Sha MSN (Eds.). NextGen data persistence pattern in healthcare: Polyglot persistence. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 4-6 July 2013.
55. Yu WD, Kollipara M, Penmetsa R, Elliadka S, editors. A distributed storage solution for cloud based e-Healthcare Information System. 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013), 9-12 Oct 2013.
56. Chen Y, Alspaugh S, Katz R. Interactive analytical processing in big data systems: A cross-industry study of mapreduce workloads. Proceedings of the VLDB Endowment. 2012;5(12):1802-1813.
57. Wu X, Zhu X, Wu G-Q, Ding W. Data mining with big data. IEEE Transactions on Knowledge and Data Engineering 2014;26(1):97-107.
58. Heiss H-U, Wagner R. Adaptive load control in transaction processing systems: Universität Karlsruhe, Fakultät für Informatik; 1991.
59. Hoffman S. Medical Big Data and Big Data Quality Problems. 2014 ;289 (2014): 2015-2018.
60. Guba EG, Lincoln YS. Fourth generation evaluation. Sage; 1989.
61. Dierker S, Bergmann U, Corlett J, Falcone R, Galayda J, Gibson M, et al. Science and Technology of Future Light Sources. Brookhaven National Laboratory; 2008.
62. McGinnis JM, Foege WH. Actual causes of death in the United States. Jama. 1993;270(18):2207-2212.
63. Paffenbarger Jr RS, Hyde R, Wing AL, Hsieh C-c. Physical activity, all-cause mortality, and longevity of college alumni. New England Journal of Medicine 1986;314(10):605-613.
64. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388(10053):1459-1544.
65. Short A. War and disease: War epidemics in the nineteenth and twentieth centuries. ADF Health 1949;11(1):15-18.
66. Briggs D. Environmental pollution and the global burden of disease. British Medical Bulletin. 2003;68(1):1-24.
67. Genes and human disease: World Health Organization(WHO); 2019 [cited 2017 SEP2017]. Available from: http://www.who.int/genomics/public/geneticdiseases/en/index2.html.
68. Agarwal R, Dhar V. Big data, data science, and analytics: The opportunity and challenge for IS research. INFORMS; 201;25(3):443-448.
69. Porter AL, Youtie J. Where does nanotechnology belong in the map of science? Nature Nanotechnology 2009;4(9):534-6.
70. Grunwald A, Orwat C. Technology Assessment of Information and Communication Technologies. Encyclopedia of Information Science and Technology, Fourth Edition: IGI Global 2018;2(1) 4267-4277.
71. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Quart. 2012; 36 (4):1165-1188.
72. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013;309(13):1351-1352.
73. Qin Y, Yalamanchili HK, Qin J, Yan B, Wang J. The Current Status and Challenges in Computational Analysis of Genomic Big Data. Big Data Research 2015;2(1):12-18.

74. Chui CK, Mhaskar H, Zhuang X. Representation of functions on big data associated with directed graphs. Applied and Computational Harmonic Analysis 2018;44(1):165-188.
75. Härmark L, Van Grootheest A. Pharmacovigilance: methods, recent developments and future perspectives. European journal of clinical pharmacology. 2008;64(8):743-752.
76. Organization WH. The importance of pharmacovigilance: Geneva: World Health Organization; 2002 [accessed July 20, 2017]. Available from: http://apps.who.int/medicinedocs/en/d/Js4893e/