

Review Article

A Systematic Review of Healthcare Big Data

Rakesh Raja , **Indrajit Mukherjee**, and **Bikash Kanti Sarkar**

Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, India

Correspondence should be addressed to Rakesh Raja; rajarakeshchauhan@gmail.com

Received 24 December 2019; Revised 14 March 2020; Accepted 20 June 2020; Published 13 July 2020

Academic Editor: Shaukat Ali

Copyright © 2020 Rakesh Raja et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Over the past decade, data recorded (due to digitization) in healthcare sectors have continued to increase, intriguing the thought about big data in healthcare. There already exists plenty of information, ready for analysis. Researchers are always putting their best effort to find valuable insight from the healthcare big data for quality medical services. This article provides a systematic review study on healthcare big data based on the systematic literature review (SLR) protocol. In particular, the present study highlights some valuable research aspects on healthcare big data, evaluating 34 journal articles (between 2015 and 2019) according to the defined inclusion-exclusion criteria. More specifically, the present study focuses to determine the extent of healthcare big data analytics together with its applications and challenges in healthcare adoption. Besides, the article discusses big data produced by these healthcare systems, big data characteristics, and various issues in dealing with big data, as well as how big data analytics contributes to achieve a meaningful insight on these data set. In short, the article summarizes the existing literature based on healthcare big data, and it also helps the researchers with a foundation for future study in healthcare contexts.

1. Introduction

The era of big data has opened the door in the healthcare industry as a response to the digitization of healthcare data. Over the past decade, the exponential growth in data [1] has introduced a new domain called big data within the field of information technology (IT) and data science. The term big data is commonly used to describe a large amount of data which are too big and not easy to handle using traditional techniques of the database management system. The idea of big data is not very new, but the manner in which it is characterized is continuously changing. In 1997, Michael Cox and David Ellsworth introduced the term “big data” for the first time in the world during a paper conferred at an IEEE conference to explain the visual representation of data and the difficulties it exhibits to computer systems [2]. The data that go beyond the processing capacity of traditional database management systems are termed as big data. These data are so large that they do not fit the structure of typical database management systems.

The notion of big data given by Doug Laney was characterized by volume, velocity, and variety known as 3Vs [3]. Generally, big data can be defined as a collection of very

large amount of data with a wide range of types, making it very hard to process using conventional database management systems. As per the author in [4], big data is a data set with large volume, high speed, and high diversity that requires a new style of processing to facilitate decision-making and exploring knowledge and optimization of techniques. Typically, a massive volume of data may be referred to as big data when capturing, analysing, and visualizing of data with current technologies are overwhelming. Big data plays an important role in the current digital era due to the significant advancement of healthcare technologies [5]. As the sources of big data concerned in healthcare industries and various sectors are well known for their volume and diversity, hence, the healthcare domain gained its effect through the impact of big data. The healthcare industries have generated enormous amount of healthcare data over the past couple of years. These healthcare data are similar to the big data in terms of their characteristics, therefore named as healthcare big data. Healthcare data generally incorporate electronic medical records (EMRs) such as patient’s medical history, physician notes, clinical reports, biometric data, and other medical data related to health. All these data together result in healthcare big data. The evolution of healthcare big data is

advance and cost-effective for both public and private healthcare. The success of healthcare applications with regard to big data entirely relies upon the underlying architecture and use of suitable tools as proven in pioneering research efforts. It also gives an idea of the analytics of big data in healthcare systems. More specifically, big data analytical tools and techniques have the potential to improve the quality of medical services and reduce the medical cost of patients by exploring the association and understanding the nature of healthcare data. In 2016, Kohli et al. discuss how electronic health records (EHR) facilitate integration of patient health history for planning safe and proper treatment [6]. More about big data and healthcare big data definition are presented in Table 1.

2. Systematic Literature Review (SLR) Method

The purpose of the research process for conducting a systematic literature review (SLR) (based on the relevant articles and studies published in academic journals) focuses on the following objectives:

- Analysing different perspectives about the concept of big data in healthcare
- Exploring the origins of healthcare big data
- Identifying tools and techniques for healthcare big data analytics
- Highlighting the potential advantages and applications of big data in healthcare
- Drawing attention to overcome the big data challenges in healthcare

By discussing these goals in depth, the systematic review aims to assist in understanding the overall context of big data and its applications in the healthcare sector.

2.1. Research Questions. The following are the key research questions that are to be addressed for conducting the SLR of the proposed study:

- RQ1. What are the characteristics of big data in the healthcare domain?
- RQ2. What are the challenges and opportunities of healthcare big data?
- RQ3. What are the features of big data analytics in healthcare?
- RQ4. What techniques are used for big data analytics in healthcare?
- RQ5. What are the applications of big data analytics in healthcare?
- RQ6. What research has been pursued in healthcare big data since 2015?

2.2. SLR Protocol. Based on the SLR protocol designed in [12], this literature review follows the below mentioned guidelines.

2.2.1. Search Strategy. The two main electronic research databases: ScienceDirect and IEEE Xplore, were used to search for the collection of relevant articles related to the proposed research. However, some good and relevant works published by Springer publ. are also included in the present study.

2.2.2. Search String. The keywords defined by the authors for search process were “Big data,” “Healthcare,” and “Big data analytics” in context to the research domain. To conduct an SLR, the search process was carried out to identify the relevant articles for addressing the research questions based on predefined keywords using Boolean operators.

2.2.3. Selection Criteria. The authors agreed to select articles based on the following inclusion-exclusion criteria:

(1) Inclusion Criteria

- The articles relevant to healthcare big data and big data analytics
- The articles published during year 2015 to 2019
- The articles from journals publications only
- The articles written in the English language

(2) Exclusion Criteria

- The articles not in the range of 2015 to 2019
- The articles other than journal publications

2.2.4. Study Selection Process. The methodology for the literature review process was performed in different stages. The details of the study selection process of SLR are shown in Figure 1. Initially, all the articles relevant to big data, healthcare big data, and big data analytics were selected in the preliminary stage of screening as per the searching keywords. Based on inclusion-exclusion criteria, these articles were screened in the first stage, and irrelevant articles which were not published between 2015 and 2019 were excluded. During the second stage of screening, the selected articles were further screened on the basis of title, abstract, and keywords. The articles which were not associated with the proposed study were excluded. Finally, in the last stage of screening, these articles were further screened on the basis of abstract using the Boolean AND operator applied to all the three authors’ defined searching keywords. As a result, 34 articles relevant to the research domain were selected from 8355 articles, for further study by the authors.

2.2.5. Quality Assessment. During the review, quality assessment plays a significant role in the SLR protocol. The quality assessment of articles was done by all authors after the analysis and evaluation of abstracts of selected articles. These articles were selected with respect to each defined key research question based on inclusion-exclusion criteria.

TABLE 1: Big data and healthcare big data definition.

Sources	Definition
[7]	Healthcare big data can be defined as digitalized version of health information which is so vast and complex that they are not easy to manage using traditional software and/or hardware, nor can they be easily handled using conventional data management tools and methods
[8]	Big data means enormous amount of digital data that organizations and governments collect about individuals and their general surrounding environments where the generated data are about 2500 petabytes or even more
[9]	Big data in healthcare refers to the data sets with $\log(n \times p) \geq 7$, and that they have high variety and high-speed characteristics
[10]	Big data can be defined as wealth of information described by massive volume, high velocity, and wide variety in order to have specific technology and analytical techniques to transform it into worth
[11]	Healthcare data generally incorporates electronic health records (EHRs) such as patient’s medical history, physician notes, clinical reports, biometric data, and other medical data related to health, as well as social media posts such as blog posts, tweets, Facebook postnotifications, and publications in medical journals

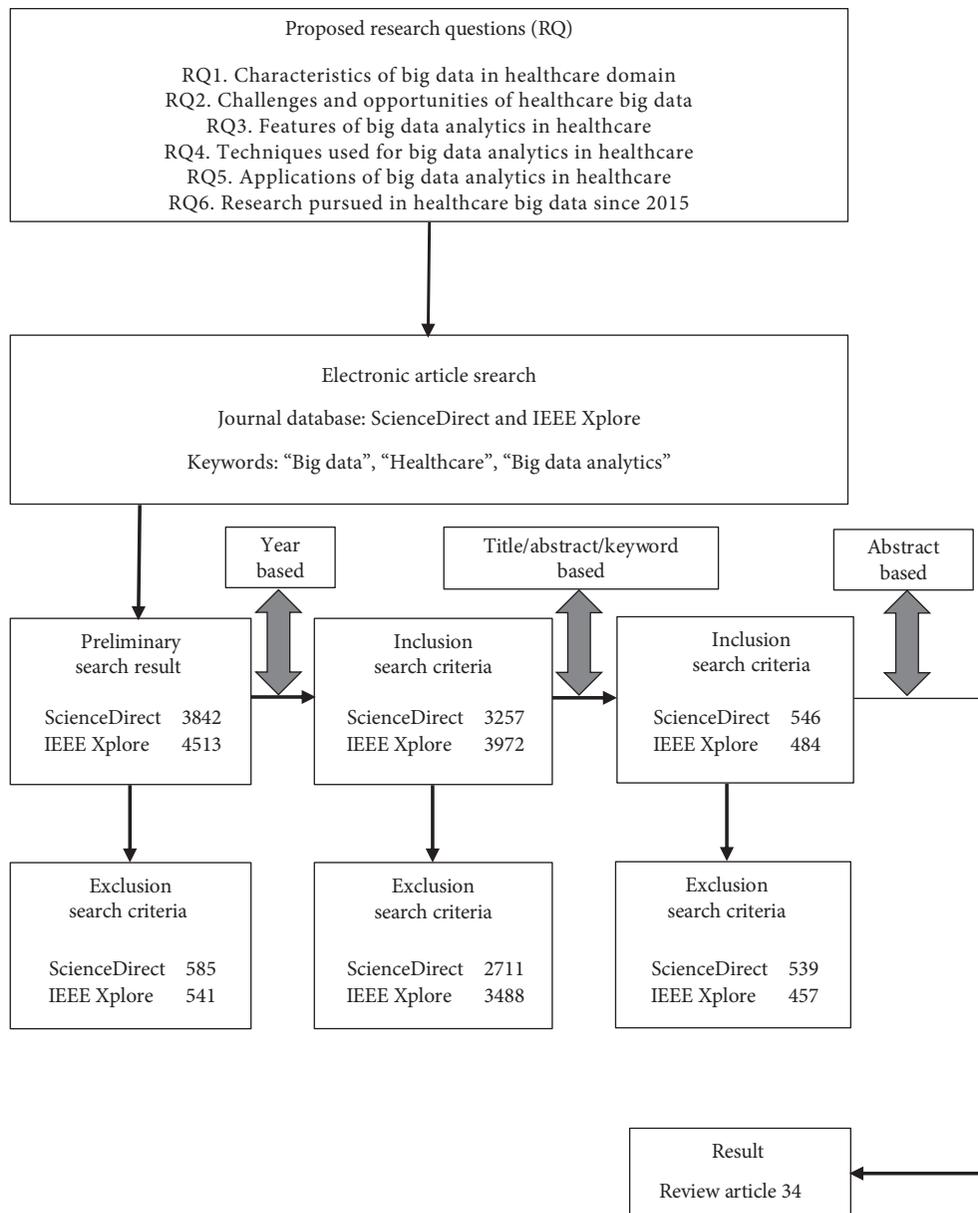


FIGURE 1: SLR process.

2.2.6. Results and Discussion. During the SLR process of the proposed research article, a collection of review articles related to defined research questions based on authors' defined search string (keywords) were identified by performing a search operation on the two most common electronic databases: ScienceDirect and IEEE Xplore. Around 7699 articles were filtered for the years 2015–2019 from the preliminary stage. Based on the title, abstract, and keywords, a total of 1030 articles were selected in the next stage. All of these articles were finally screened on the basis of the abstract using the Boolean AND operator applied to all three searching strings (keywords). As a result, 34 articles with respect to each defined research question were selected for further study by the authors according to the inclusion-exclusion criteria.

Table 2 shows the three screening stages of articles. Based on the main research objectives, the contents from these articles were extracted, and the proposed research article was organized into different sections: comprehensive overview of big data in the healthcare domain, sources of healthcare big data, challenges of big data in healthcare, big data analytics in healthcare, and application and potential benefits of big data in healthcare.

2.3. Trend of Big Data Research in Healthcare Domain. With the rapid growth of data, big data has given researchers an exposure to utilize it in more noticeable manner for decision-making in several healthcare applications. The trend of big data in the field of healthcare domain for the year 2015–2019 is described in Figure 2 with respect to Tables 2 and 3 of the revised version of the article. Figure 2 shows the increasing tendency of doing innovative research studies (published in reputed journals) in the area of healthcare big data.

3. Big Data: A Comprehensive Overview

3.1. Big Data in General. Big data refers to a collection of extensive and complicated data sets that are hard to handle using conventional database systems. As per the *zdnet.com*, big data pertains to the tools and techniques that allow an organization to generate, exploit, and maintain vast amounts of data with storage facilities. Each one of us is continuously producing enormous amount of data. And, big data is being generated by every computerized system as well as social networking sites. It is transmitted by the digital system, sensor devices, cameras, handheld devices, smartphones, and their applications [13]. Big data arrives at an unprecedented rate, large data size, and greater diversity from various sources. To extract significant worth from such large amount of data, we need high computational power, analytical capabilities, and expertise. This explosion of data attempts to change the opinion of people to think about everything in terms of big data. In recent times, transactional data, web-based data, sensor data, and electronic medical data keep developing with rapid speed. These data can be classified into web-based data, sensor-based data, demographic data,

transactional data, and machine-generated data [14] (as stated below):

Web-based data are acquired from social networking sites such as Facebook, Twitter, and Blogs

Machine-generated data are extracted from sensor-based devices and other gadgets

Transactional data are retrieved from biometrics, vital sign, radiology, and other medical images

Human-generated data comprise E-mails, doctor's prescriptions, and digitalized version of medical reports

This remarkable development in data growth has led to this new concept known as big data. In article [15], it is stated that big data is a complex set of data that has a significant impact on the ability of conventional data warehouses to store, maintain, perform, and analyse data. A formal definition of big data has been provided in [10]. It is stated there as follows: big data is a wealth of information described by huge quantity, high velocity, and wide variety in order to have specific technology and analytical techniques to transform it into worth. Looking at it another way, the McKinsey Global Institute defines big data as data sets whose size exceeds the capability of conventional database systems to collect, store, maintain, and analyse data. According to the authors in [16], big data is the assemblage of data collected from different sources such as corporate databases, websites, maps, movies, and public databases.

3.2. Characteristics of Big Data. The common characteristics of big data are illustrated in the following:

Volume: this implies data size usually measured in terabytes (TB = 10^{12} bytes), petabytes (PB = 10^{15} bytes), and zettabytes (ZB = 10^{21} bytes), and so forth

Velocity: this indicates the rate of generation of data

Variety: this refers to the nature of data which big data can include such as structured, semistructured, and unstructured data

Veracity: this refers to the trustworthiness of the data

Value: the term itself is related to the worth of data being extracted

Apart from the abovementioned features of big data, several researchers and scientists have introduced new features to big data due to various applications available; i.e., the big data definition keeps changing according to the advancement of technology, data storage, and data transmission rate, as well as other system capabilities. The different explanations for the definition of big data are from 3Vs to 4Vs [17, 18], 5Vs [19], and 10Vs [20]. In particular, these dimensions are expanding as time goes by; and we currently have 42 distinct dimensions for big data till 2017 as per [21], and also the dimensions will keep on expanding as the big data evolves further. Figure 3 describes the generic notion of big data.

TABLE 2: Screening stages of research articles.

Electronic search		Article selection based on search string			Result
		Big data	Big data, healthcare	BDA	
Journals	ScienceDirect	2852	642	348	3842
	IEEE Xplore	4080	129	304	4513
<i>First screening based on year (2015–2019)</i>					
Journals	ScienceDirect	2424	543	290	3257
	IEEE Xplore	3578	123	271	3972
<i>Second screening based on title/abstract/keywords</i>					
Journals	ScienceDirect	464	34	48	546
	IEEE Xplore	429	32	23	484
<i>Final screening based on abstract</i>					
Journals	ScienceDirect		7		34
	IEEE Xplore		27		

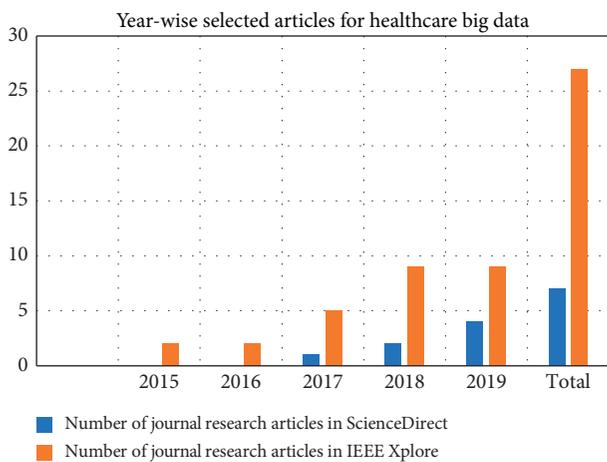


FIGURE 2: Trend of healthcare big data research from 2015 to 2019.

TABLE 3: Number of research articles year-wise.

Year	Number of journal research articles	
	ScienceDirect Articles	IEEE Xplore Articles
2015	0	2
2016	0	2
2017	1	5
2018	2	9
2019	4	9
Total	7	27
Outline	Studies which discuss the significance of BDA and the usefulness of big data in the field of healthcare	

3.3. *Big Data Definitions.* Big data and healthcare big data definitions are given in Table 1.

4. Big Data in Healthcare Domain

4.1. *Healthcare Big Data.* A pioneering renovation is taking place in the healthcare industry. The healthcare industry is generating a large volume of healthcare data due to the advancement in technology and digitization of medical records. In recent years, health information technology

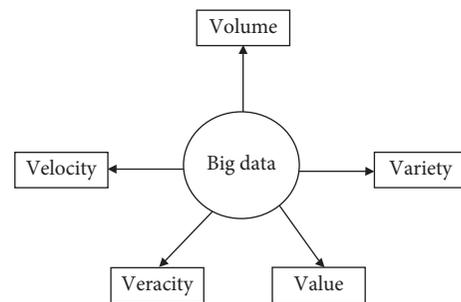


FIGURE 3: Generic concept of big data.

(HIT) has developed the power to generate, store, and transmit data electronically worldwide within seconds and also has the potential to deliver tremendously better productivity and service quality to healthcare. It allows each stakeholder in healthcare sectors to possess his/her own database of patients’ medical records in a digital form. The healthcare sectors have produced huge amounts of healthcare data by keeping records, consent and regulatory requirements, and patient care [11]. All these data together form healthcare big data. To be more specific, healthcare big data can be defined as electronic medical records (EMRs) which incorporates patient’s medical history, physician notes, clinical reports, biometric data, and other medical data related to health, as well as social media posts such as blog posts, tweets, Facebook postnotifications, and publications in medical journals [11]. Importantly, the exponential growth of healthcare data is another major issue in the current healthcare information systems (HISs). This transformation is not only about the large volume of healthcare data; however, we are also experiencing an exponential rise in the velocity at which these data are generated, as well as large diversity of medical data.

The evolution of advancement in technologies like sensor systems, cameras, and smartphones is a significant source of healthcare data. Everyday new sources of data are introduced. This makes it much more difficult to process or analyse big data in healthcare using common database management tools. Typically, when massive volume of healthcare data are captured, stored, and analysed properly in order to gain insight, it will enhance the healthcare service

outcome through smart decisions and also reduce healthcare costs. However, effective data analytical tools and techniques as well as powerful computing systems are required for this purpose. Healthcare big data analytics (BDA) in particular has started to emerge as a promising tool for taking care of issues in numerous healthcare disciplines. In addition, the role of a data analyst is to mine the big data, exploring the association and understanding trends and patterns of healthcare data. This enhances the health and improves the quality of life of an individual, as well as provides appropriate early-stage treatment at low cost.

The amount of data stored in healthcare sectors continued to increase curiosity about healthcare big data. There is an enormous amount of data ready to be analysed. One of the principle motivations behind big data is to focus on healthcare. The basic motive of nations around the world is to improve the healthcare facilities and decrease the medical costs. However, the revolution of massive volume of data in healthcare remains a barrier for achieving this goal. Electronic healthcare data from all around the world were estimated at 500 petabytes in 2012, reaching 25 petabytes by 2020 [22]. Thus, healthcare can be described as a wide variety of services offered by medical professionals to people, families, or societies to encourage, maintain, or restore better health. The quality of the healthcare system is significant because it determines hospital sustainable growth and helps people to maintain the optimal state of health. In certain cases, the quality of healthcare services is too high, and it ends up costly for patients. Consequently, it is essential to address the key healthcare procedures and related quality parameters that act in collaboration to ensure the best possible outcomes for patients and reduce the healthcare costs.

4.2. Sources of Healthcare Big Data. This section deals with several important sources of healthcare data. Big data in healthcare can revolutionize the medical field through early-stage disease detection using adequate analytical tools and techniques by incorporating and analysing health-related information in a comprehensive manner. Currently, the evolution of advancement in technologies like sensor systems, cameras, wearable devices, and mobile applications is widely used in the domain of the medical field [23, 24]. As a result, more medical information is being explored in a consistent manner. Data in medicinal services are fragmented and dispersed, originating from disparate sources with multiple formats [25]. The facts confirm that information on health is large and heterogeneous. The reason is on the ground that they originate from various internal and external sources accessible at multiple locations. External sources include web data, social media data, and machine-generated data, and internal sources include transactional data, biometric data, and human-generated data. Various healthcare data and their sources are summarized in Table 4.

4.3. The 5Vs of Healthcare Big Data Characteristics. In this section, the important Vs about healthcare data are briefly stated. The five key characteristics that have been found in

most literature [12, 35] to define healthcare big data are as follows:

Volume. Based on the general discussion of big data, healthcare data are a perfect case of big data. The volume refers to the data size that grows exponentially day to day, and by 2020, the volume of big data may reach to 44 zettabytes [36]. Compared to most of the industries, the healthcare sector generates massive amounts of data in the form of electronic medical records (EMRs), biometric data, clinical data, radiology images, genomics, etc. All these data collectively form healthcare big data [37–39]. Obviously, the utilization of several tools such as Hadoop, MapReduce, and MongoDB is getting more popular among healthcare organizations due to their ability to store and measure massive volume of data [40, 41].

Velocity. Velocity refers to the speed at which data are generated, as well as data acquired from various healthcare systems [42].

Variety. Variety refers to the heterogeneity and diversity of data. The healthcare industry generates and collects data at a staggering rate from different sources such as social networking sites, sensor devices, cameras, and smartphones. However, these healthcare data may be in any one of the forms, structured, unstructured, or semistructured. Example of structured data is clinical data, whereas data such as physician notes, images, social media data, mobile data, and radiograph films are unstructured or semistructured. Figure 4 depicts the types of healthcare data, along with examples.

Veracity. The veracity characteristics of healthcare data refer to the trustworthiness of the data, which in this context is equivalent to quality assurance of data. It gives the degree of authenticity about healthcare knowledge.

Value. Value is the most important and distinctive characteristics of all the 5Vs of healthcare big data, as it has the ability to transform healthcare data into worth of information. Its concept is exactly in line with that of healthcare data.

5. Big Data Challenges in Healthcare

The evolution of big data introduces several challenges, constraints, and problems due to exponential growth of healthcare data. Big data is constantly changing, and this change of data presents a lot of challenges in storing, analysing, and retrieving the massive volume of data. Certainly, the conventional database systems could not be used to store, process, and extract the information due to its massive size and diversity of data.

The main challenges encountered by healthcare BDA are as follows:

- Quality and storage of data
- Data analysis of good quality
- Expertise in data analytics

TABLE 4: Important sources of healthcare data.

Healthcare data	Features	Sources
Clinical data	Data within electronic health records (EHRs) can be either structured (e.g., EMRs and clinical data), unstructured (e.g., clinical trials data), or semistructured (e.g., claims data) forms	[26]
Patient-generated data	Biometric data, social media data, online data (e.g., blogs, Facebook posts, and Twitter)	[27]
Sensor data	Data produced by sensor-based devices (e.g., vital signs, ECG, and handheld devices)	[28–30]
Genomic data	Gene typing (e.g., gene expression and DNA sequence)	[11, 31, 32]
Clinical research data	Health product data (e.g., drug information)	[33]
External data	Insurance data (e.g., financial data) Biometric data (e.g., fingerprints)	[34]

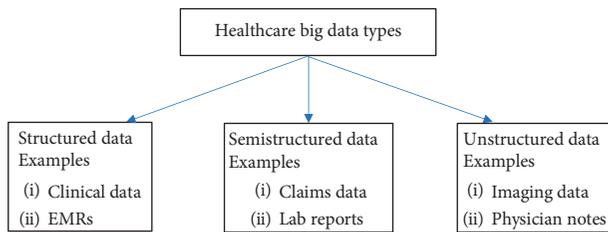


FIGURE 4: Types of healthcare big data.

Data security and confidentiality
Multiple sources of data

Healthcare big data challenges encountered are no different. Big data characteristics are the main issues that need to be addressed. It is vital to move towards big data technology in order to provide better medical facilities. Big data technology, however, introduces a potential risk to certain categories.

5.1. *Issues in Healthcare Big Data.* Big data issues that generally occur in the healthcare organizations are covered by four main categories [35, 43]:

Data Governance. Data management and regulation is the governance of data. As the healthcare sector moves towards data analytics, data governance is a major challenge. Healthcare data generated are diversified in nature, requiring standardization and governance.

Economic Challenges. The facilities in the medical field between patients and healthcare professionals throughout clinical visits depend on the paid service. Subsequently, advancement in technologies associated with this process places a burden on the medical community and generates an unnecessary impact for the personnel against such unpaid services.

Big Data Technology Challenges. Big data in healthcare is enormous and highly fragmented which causes problems in quality of information, as well as technology-wise, big data creates a barrier to accomplish the healthcare vision [44].

Security and Privacy Issues. In the era of big data, the privacy of healthcare data must be seriously considered due to the potentially sensitive information about individual healthcare stakeholders. Healthcare data are highly sensitive data which must be secured from unauthorized access so that it cannot be made publicly available, as well as healthcare fraud can also be prevented from attackers. Therefore, data security is one of the most important challenging tasks in the healthcare domain.

While studying and analysing several published research papers with reference to the SLR protocol, this research focuses on how recent developments in ICT (information and communication technologies) together with big data techniques can be effectively incorporated to address these challenges of healthcare big data and make a significant contribution towards healthcare services [45–51]. Based on [17, 52], we the authors classify healthcare BDA into three categories, namely, descriptive analytics, predictive analytics, and prescriptive analytics. Among these different BDA techniques, this literature review reflects that there are various tools, for example, Hadoop [53] and MapReduce [54] that have been developed for healthcare big data management. These are described in Section 6. A few of the well-known BDA techniques used in the areas of healthcare are described in Table 5. The categories provided in Table 5 are drawn from the literature [12, 66–68].

6. Big Data Analytics in Healthcare

Healthcare BDA has a potential to improve the quality of care and reduce the medical cost of patients by finding the associations from massive volume of healthcare data, thereby offering a wider perspective of clinical expertise based on medical evidence and various tests. Advanced analytical tools and techniques used in healthcare systems provide services that satisfy a growing need and enable healthcare agencies to process massive volume of data, analyse it in real time, and extract knowledge from medical records of all patients. In 2017, Palanisamy and Thirunavukarasu have presented various analytical avenues that exist in the patient centric healthcare system from the perspective of various stakeholders [69]. The main goal of the article is to assist researchers and data scientists to make

TABLE 5: BDA techniques.

BDA techniques	Healthcare application areas	Examples	Sources
Machine learning	Early detection of diseases	Predicting epidemics, disease monitoring	[12, 55]
Data mining	Prediction of heart disease at early stages	Health analytics, determination of epidemics	[56–59]
Neural network	Diagnosis of chronic diseases	Predicting of patients' future disease, patient safety	[60–62]
Pattern recognition	Improvement of public health disease surveillance	Empowering public health, health literacy, improving quality of care	[63, 64]
NLP	Improve the quality of care and accuracy of clinical decisions	Cost reduction, high-risk factor identification	[57, 65]

informed healthcare decision and enhance the performance of the healthcare centre, so that people live a healthier lifestyle. In particular, this includes numerous analytical techniques such as machine learning, pattern recognition, statistical analysis, visualization, and data mining to interpret feature relationships and discover knowledge. BDA is based on the concept of data mining that incorporates various analytical techniques to evaluate and explore large volume of data to extract significant and useful information. Researchers may find ample information about BDA and healthcare from the articles [66, 70–72].

6.1. Types of Healthcare Big Data Analytics. BDA mainly perform three types of analytics, namely, *descriptive analytics*, *predictive analytics*, and *prescriptive analytics*. The descriptive analytics facilitates to explore insights and allows healthcare practitioners to understand what is happening in a given situation [73, 74]. In the context of healthcare data, the descriptive analytics analyses the data gathered in order to interpret, understand, summarize, and visualize significant health-related information. On the other hand, predictive analytics assist healthcare stakeholders to identify the healthcare services and responding appropriately according to the requirements of patients. It also enables clinicians to be capable of making patient-related decisions on the basis of system predictions [73, 74]. Predictive analytics involves various statistical techniques used to analyse and extract valuable insights from big data [17]. Hadoop/MapReduce is one of the most widely used techniques to develop a predictive model for healthcare systems. Prescriptive analytics is comparatively a modern type of analytics that combines descriptive and predictive analytics [75]. Though predictive analytics recommends what will happen in the future, prescriptive analytics provides the best course of action to be taken by healthcare providers in the future [73, 74]. By incorporating clinical and genomic data, prescriptive analytics continuously repredicts the healthcare services and improves the predictive accuracy in order to provide more appropriate diagnoses and treatments for healthcare providers [76, 77].

The medical industry is flooded with enormous volumes of data that require validation and analysis. BDA has a power and capability to perform essential computing and analytical ability to process large volumes of healthcare data. It facilitates medical professionals, clinical researchers, and healthcare stakeholders to improve their results through the use of their internal and

external sources of big data [78, 79]. As per the healthcare providers, the assessment of patient data, which incorporates patient medical history (EHR), doctors' prescriptions, diagnostic reports, biometric data, clinical tests, and other medical data related to health, assists them to follow the advancement of a recommended course of treatment and interrupt the course so that changes can be made if necessary. Thus, it helps to eliminate unnecessary visits and reduce readmission rates. Furthermore, the drug company and other medical organizations take benefit of analytical advantages in designing marketing strategies. Indeed, pharmaceutical industries can study their current market status by capturing and analysing the healthcare data such as sales record and interpretation of drug information prescribed by healthcare professionals for each patient and disease to develop the strategic goals. Therefore, the health insurance company can develop an appropriate health plan for every patient by analysing their demographic data, clinical trials, and statistical data related to health factors [69].

An enormous amount of data are accumulated in the healthcare sector from patients' medical histories, clinical trials, and diagnostic reports. Like healthcare big data, data analytics can be characterized by volume, velocity, and variety known as 3Vs [3, 17]. BDA is the use of advanced analytical techniques to analyse, extract, and discover meaningful patterns and insight from large data sets [80, 81]. BDA plays a crucial role in enhancing healthcare facilities and increases patients' clinical outcomes. It therefore has the ability to improve the quality of care and life styles and reduce medical costs. Based on the systematic review on the current state of big data research by Wang and Hajli, BDA in the context of healthcare can be characterized as the capability to acquire, store, process, and analyse large amounts of health data in different forms and provide meaningful information to users, which enables them to explore business values and insights in a timely manner [82].

6.2. Necessity of Healthcare Big Data Analytics. BDA in healthcare is needed to enhance the healthcare quality by taking the associated healthcare services into account:

Provision of Personalized Healthcare. Big data in healthcare can revolutionize the medical field through early-stage disease detection and reduce medical cost for the patients using appropriate analytical tools in a

comprehensive manner. This helps to develop a personalized healthcare system for healthcare stakeholders [83, 84].

Early Detection of Spread of Diseases. This concentrates on early prediction of viral (infectious) diseases (*i.e.*, before spreading) on the basis of social network analysis. More and more social media of the patients suffering from a disease in a specific geographical area are monitored to identify the development and spread of viral disease. This assists the healthcare experts to counsel the sufferers to take the necessary preventive action.

Monitoring the Clinical Performance. There is a lot of enthusiasm to evaluate clinical performance in order to screen and enhance the quality of healthcare services. The reform of the hospital is of major concern in the strategic plan of the healthcare sectors. This can be achieved by monitoring and setting up the hospital in accordance with medical council's standards.

6.3. Big Data Analytical Techniques in Healthcare. In the past, traditional technologies and data warehouses were used by the data analyst to store, process, and manage data. However, the revolution of massive volume of data in healthcare cannot be handled using conventional database systems, tools, and techniques. Nowadays, many advanced technologies with high computing power and storage capacity have been developed in order to address the low performance and difficulty of traditional systems. Accordingly, in [85], "big data technologies can be referred to as advanced technologies that have a high computing power and analytical ability to process large volumes of data collected from various sources to extract insight from it." As per the authors in [86, 87], big data techniques cover a wide range of fields such as machine learning, statistical analysis, and image analysis. A few of the well-known BDA techniques used in the areas of healthcare are shown in Table 5. The categories that are generated in Table 5 are taken from the literature [12, 66–68, 86]. Big data plays a significant role across all domains such as government organizations, trade associations, healthcare industries, education, and research and development. BDA also empowers the secondary use of clinical data in the healthcare sector [88]. Big data acceptance has shown enormous growth from 17 percent in 2015 to 53 percent in 2017 according to Forbes [89].

In the current digital era, healthcare is one of the sectors that generates a large volume of healthcare data, and these healthcare big data can be characterized by its volume, velocity, and variety known as 3Vs [3]. Data mining techniques can be applied on this massive amount of healthcare data so as to identify new interesting patterns and valuable insights for quality medical services. The Hadoop is an open source software framework for BDA in healthcare as well as the most popular implementation of the MapReduce programming model [90]. It allows distributed storage and processing of large variety of healthcare big data whether it is structured, semistructured, or unstructured such as patient's EHR, physician's notes, laboratory data, clinical trials

reports, and insurance data as compared to conventional database systems. Figure 5 shows a general conceptual architecture of big data analytics [7].

6.4. Platform and Tools for Healthcare Big Data Analytics.

There are currently several techniques available for performing BDA. The few tools and techniques that support the Hadoop distributed platform are being discussed below [91, 92]:

Hadoop Common. It refers to the set of common utilities that assist other modules of the Hadoop framework. Hadoop Common is a fundamental part of the Apache platform in addition to the HDFS, YARN, and MapReduce. Hadoop Common is generally called as Hadoop Core.

Apache HDFS. HDFS refers to the Hadoop Distributed File System that can be used to process unstructured data on commodity hardware predominantly. HDFS is the primary data storage where each file is divided into blocks of fixed size and distributed across numerous servers (nodes). HDFS employs the master/slave architecture using NameNode (master node) and DataNode (slave node) [93, 94].

Hadoop MapReduce. MapReduce is a programming framework that enables us to process massive amount of data in parallel in a distributed computing environment. This framework consists of two main functions, namely, Map and Reduce that can effectively manage structured as well as unstructured data [95, 96]. As the name MapReduce indicates, reducer function occurs after the completion of the mapper function.

Apache Hive. Hive is a data warehouse framework designed to query and analyse huge amount of data stored in Hadoop HDFS. It is an ETL (extract, transform, and load) tool for the Hadoop ecosystem. Hive is built on top of the Hadoop platform and provides a declarative language similar to SQL known as the Hive query language (HiveQL) that enables SQL programmers to perform data analysis conveniently [97].

Pig. Apache Pig is a parallel computing framework that runs on the Apache Hadoop platform. Pig Latin is the language for this platform which is used for analysing large volumes of data due to its distributed architecture. In fact, Pig Latin is like the SQL language and is easy to learn. The main distinction is that Pig Latin can process semistructured and unstructured data [93, 98].

HBase. Apache HBase is an open source, multidimensional distributed database system in a Hadoop ecosystem. It runs on the top of the Hadoop Distributed File System (HDFS). HBase can store large volumes of data usually measured in terabytes (TB) to petabytes (PB) and does not support a structured query language like SQL; indeed, HBase employs a NoSQL approach.

Mahout. Apache Mahout is an open source distributed framework that supports BDA on the Hadoop platform and is designed for machine learning using the MapReduce program. The Apache Mahout enables us to

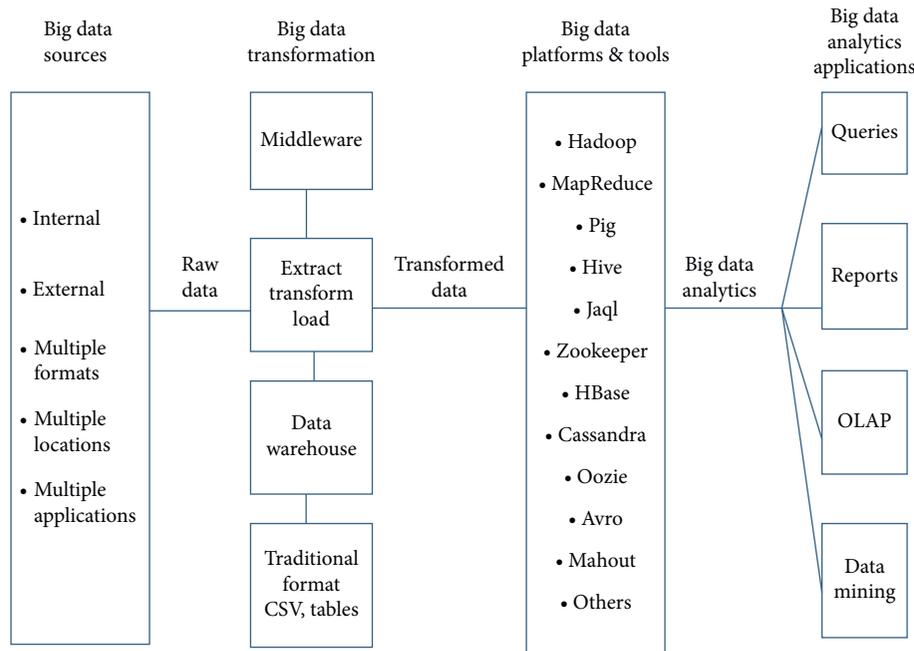


FIGURE 5: A conceptual architecture of big data analytics [7].

develop collaborative filtering, classification, clustering, association mining, and statistical algorithms related to machine learning with the help of data science techniques [93, 99].

For more about the tools and techniques, one may refer to [7, 53].

7. Big Data Benefits in Healthcare Sector

Healthcare sectors extending from a single physician's office to a large set of networks of healthcare service providers have a potential to acknowledge significant benefits by digitizing, integrating, and effectively using big data analytical tools and techniques in healthcare.

Based on the recently published studies [65, 66, 100], following are some of the major benefits:

Clinical operations: the information on healthcare helps to determine methods of diagnosing and treating patients that are more clinically important and cost-effective

Patients: healthcare information can help patients to make the right decision at the right time and improve patients' health while reducing the healthcare cost

Healthcare providers: the data acquired from medical organizations assist the stakeholders to develop new healthcare strategies for patients to minimize the unnecessary hospitalizations

Research and development: healthcare data support researchers and scientists to enhance healthcare services through more precise and appropriate treatments

Public health: healthcare data also assist to assess the health risks as well as analyses trends of diseases to enhance public health surveillance

8. Application of Big Data Analytics in Healthcare

The buzzword big data in the digital world is highly in demand in every sector especially in the field of healthcare. This has laid a foundation for various applications in the healthcare sector. Healthcare BDA has a potential to improve the quality of care and reduce the medical cost of patients by discovering the associations from massive volume of healthcare data, thereby offering a wider perspective of clinical expertise based on medical evidences and various tests [101]. Healthcare BDA also helps the clinicians and policy makers to develop public policy and service delivery based on open health prescribed data, disease prevalence data, and economic deprivation data [102]. As per the authors in [100, 101, 103, 104], the major areas for the applications of BDA in healthcare are as follows:

Healthcare Monitoring. Healthcare data analytics can be used to continuously monitor the health status of the users (patients) in order to enhance their lifestyle [93].

Healthcare Risk Prediction. A deep analysis of healthcare data helps healthcare stakeholders and medical practitioners to develop solutions for risk prediction. It also enables clinicians to be capable of making patient-related decisions on the basis of system predictions [73, 74]. Data analytics in healthcare can also be used to identify and manage high-risk and high-cost patients [105].

Behavioural Monitoring. Another prospective implementation of BDA in healthcare is monitoring of patients with abnormal behaviour [66]. In 2005, Nambu et al. proposed the home healthcare system to capture the behavioural data of patients for diagnosing their health conditions [106].

Fraud Detection and Prevention. One of the major and important application of data analytics in the healthcare sector is fraud detection and prevention. As per the authors in [107], data mining and machine learning techniques are mainly used for fraud detection in healthcare.

Clinical Decision Support Systems. In the medical field, clinical decision support systems are designed to facilitate healthcare professionals in making clinical decisions to diagnose diseases based on patient's health condition [108, 109].

Personalized Healthcare Recommendation System. Big data plays a significant role in the healthcare domain to develop a personalized recommendation system to give precise and relevant medical recommendation (advice) to an individual (patient) based on their current health status and medical history [110]. The authors in [111] proposed an intelligence-based health recommendation system using BDA to study and research health records of patients, assess risk and the severity of different diseases, and then provide recommendations based on outcomes of prediction. The authors in [112] suggested a clinical recommendation system that is beneficial for patients to access accurate recommendations based on their own health status.

Drug Discovery and Clinical Trials. Healthcare BDA is widely used by the pharmaceutical industry for drug discoveries so that it can help physicians, pharmaceutical developers, and other healthcare professionals for getting the right drug to the right patient at the right time [107, 113, 114].

Image Informatics and Telediagnosis. Imaging informatics is the study of methods for generating, managing, and representing imaging information in various biomedical applications. It is concerned with how medical images are exchanged and analysed throughout complex healthcare systems [115, 116]. The authors of the study [117] introduce a novel telemammography system for early detection of breast cancer with the help of image processing and machine learning techniques. Computer-aided diagnosis plays a significant role in medical imaging [118].

Healthcare Knowledge System. According to [119], a knowledge management system is developed based on healthcare big data in order to support clinical decision-making and disease diagnosis. The healthcare knowledge system is based on a variety of databases such as electronic health record (EHR), medical imaging data, and unstructured clinical notes and genetic data.

Public Health Information. As per [115, 120, 121], BDA in healthcare can also be used to track and monitor public health status for decision-making and policy development.

Based on the studies of different authors, it is revealed that the BDA in healthcare has a potential to improve the quality of healthcare, decreasing the readmission rates and

reducing the medical cost of patients by exploring the association and understanding the nature of healthcare data [7, 93, 122]. Furthermore, image processing, signal processing, and genomics are presently the three main areas for the application of data analytics in the healthcare domain [123].

9. Conclusion

This systematic review focuses on the existing literature to study healthcare big data based upon defined keywords and research aspects in the healthcare domain. The proposed research uses an SLR protocol and guidelines to review the systematic study of the past and the cutting-edge articles of the big data in healthcare. The purpose of an SLR protocol is based on the following objectives:

- Analysing different perspectives about the concept of big data in healthcare
- Exploring the origins of healthcare big data
- Identifying tools and techniques for healthcare big data analytics
- Highlighting the potential advantages and applications of big data in healthcare
- Drawing attention to overcome the big data challenges in healthcare

The present study will help the researchers with a useful base for future work to understand the overall context of healthcare big data and its applications. The limitation of the proposed research is that the electronic search process was performed in only two journal databases from 2015 to 2019, and the rest of the databases were skipped while accessing the quality of journal articles which can be addressed in future research.

Data Availability

Data sharing is not applicable to this article as no data sets were generated or analysed during the current study.

Conflicts of Interest

The authors declare that no conflicts of interest exist regarding this publication.

References

- [1] D. P. Augustine, "Leveraging big data analytics and Hadoop in developing India's healthcare services," *International Journal of Computer Applications*, vol. 89, no. 16, pp. 44–50, 2014.
- [2] M. Cox and D. Ellsworth, "Application-controlled demand paging for out-of-core visualization," in *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, pp. 235–244, IEEE, Phoenix, AZ, USA, October, 1997.
- [3] D. Laney, "3D data management: controlling data volume, velocity and variety," *META Group Research Note*, vol. 6, 2001.

- [4] C. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.
- [5] X. Wang, Y. Wang, C. Gao, K. Lin, and Y. Li, "Automatic diagnosis with efficient medical case searching based on evolving graphs," *IEEE Access*, vol. 6, pp. 53307–53318, 2018.
- [6] R. Kohli, S. S. L. Tan, and S. S.-L. Tan, "Electronic health records: how can IS researchers contribute to transforming healthcare?" *MIS Quarterly*, vol. 40, no. 3, pp. 553–573, 2016.
- [7] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, p. 3, 2014.
- [8] M. Vivekanand and B. M. Vidyavathi, "Security challenges in big data," *International Journal of Advanced Research in Computer Science*, vol. 6, no. 6, 2015.
- [9] E. Baro, S. Degoul, R. Beuscart, and E. Chazard, "Toward a literature-driven definition of big data in healthcare," *BioMed Research International*, vol. 2015, Article ID 639021, 9 pages, 2015.
- [10] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of big data based on its essential features," *Library Review*, vol. 65, no. 3, pp. 122–135, 2016.
- [11] K. Priyanka and N. Kulennavar, "A survey on big data analytics in health care," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5865–5868, 2014.
- [12] S. Nazir, M. Nawaz, A. Adnan, S. Shahzad, and S. Asadi, "Big data features, applications, and analytics in cardiology—a systematic literature review," *IEEE Access*, vol. 7, pp. 143742–143771, 2019.
- [13] S. Sagioglu and D. Sinanc, "Big data: a review," in *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS)*, IEEE, San Diego, CA, USA, pp. 42–47, May 2013.
- [14] I. Lee, "Big data: dimensions, evolution, impacts, and challenges," *Business Horizons*, vol. 60, no. 3, pp. 293–303, 2017.
- [15] S. Tiwari, H. M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: insights to industries," *Computers & Industrial Engineering*, vol. 115, pp. 319–330, 2018.
- [16] G. Silahtaroglu and N. Yılmaztürk, "Data analysis in health and big data: a machine learning medical diagnosis model based on patients' complaints," *Communications in Statistics-Theory and Methods*, pp. 1–10, 2019.
- [17] A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [18] C. H. Lee and H.-J. Yoon, "Medical big data: promise and challenges," *Kidney Research and Clinical Practice*, vol. 36, no. 1, pp. 3–11, 2017.
- [19] R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives," *Computers & Industrial Engineering*, vol. 101, pp. 572–591, 2016.
- [20] K. F. Tiampo, S. McGinnis, Y. Kropivnitskaya, J. Qin, and M. A. Bauer, "Big data challenges and hazards modeling," in *Risk Modeling for Hazards and Disasters*, pp. 193–210, Elsevier, 2018.
- [21] L. Wang and C. A. Alexander, "Big data in medical applications and health care," *American Medical Journal*, vol. 6, no. 1, pp. 1–8, 2015.
- [22] J. Sun and C. K. Reddy, "Big data analytics for healthcare," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '13*, Chicago, IL, USA, 2013.
- [23] A. T. Lo'ai, R. Mehmood, E. Benkhelifa, and H. Song, "Mobile cloud computing model and big data analysis for healthcare applications," *IEEE Access*, vol. 4, pp. 6171–6180, 2016.
- [24] I. García-Magariño, R. Lacuesta, and J. Lloret, "Agent-based simulation of smart beds with internet-of-things for exploring big data analytics," *IEEE Access*, vol. 6, pp. 366–379, 2017.
- [25] W. B. Rouse and N. Serban, *Understanding and Managing the Complexity of Healthcare*, MIT Press, Cambridge, MA, USA, 2014.
- [26] S. Yang, M. Njoku, and C. F. Mackenzie, "Big data approaches to trauma outcome prediction and autonomous resuscitation," *British Journal of Hospital Medicine*, vol. 75, no. 11, pp. 637–641, 2014.
- [27] N. P. Terry, "Protecting patient privacy in the age of big data," *SSRN Electronic Journal*, vol. 81, p. 385, 2012.
- [28] R. B. Shrestha, "Big data and cloud computing," *Applied Radiology*, vol. 43, no. 3, p. 32, 2014.
- [29] A. Rizwan, A. Zoha, R. Zhang et al., "A review on the role of nano-communication in future healthcare systems: a big data analytics perspective," *IEEE Access*, vol. 6, pp. 41903–41920, 2018.
- [30] L. Carnevale, R. S. Calabrò, A. Celesti et al., "Toward improving robotic-assisted gait training: can big data analysis help us?" *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1419–1426, 2018.
- [31] P. Y. Wu, C. W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, and M. D. Wang, "—Omic and electronic health record big data analytics for precision medicine," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, pp. 263–273, 2016.
- [32] F. Celesti, A. Celesti, J. Wan, and M. Villari, "Why deep learning is changing the way to approach NGS data processing: a review," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 68–76, 2018.
- [33] K. Miller, "Big data analytics in biomedical research," *Biomedical Computation Review*, vol. 2, pp. 14–21, 2012.
- [34] S. C. Helm-Murtagh, "Use of big data by blue cross and blue shield of North Carolina," *North Carolina Medical Journal*, vol. 75, no. 3, pp. 195–197, 2014.
- [35] B. K. Sarkar, "Big data for secure healthcare system: a conceptual design," *Complex & Intelligent Systems*, vol. 3, no. 2, pp. 133–151, 2017.
- [36] M. S. Hajirahimova and A. S. Aliyeva, "About big data measurement methodologies and indicators," *International Journal of Modern Education and Computer Science*, vol. 9, no. 10, p. 1, 2017.
- [37] A. Widmer, R. Schaer, D. Markonis, and H. Müller, "Gesture interaction for content-based medical image retrieval," in *Proceedings of International Conference on Multimedia Retrieval*, pp. 503–506, April 2014, Glasgow, Scotland.
- [38] J. A. Seibert, "Modalities and data acquisition," in *Practical Imaging Informatics*, pp. 49–66, Springer, New York, NY, USA, 2009.
- [39] A. S. Panayides, M. S. Pattichis, S. Leandrou, C. Pitris, A. Constantinidou, and C. S. Pattichis, "Radiogenomics for precision medicine with a big data analytics perspective," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2063–2079, 2018.

- [40] G. Adrián, G. E. Francisco, M. Marcela, A. Baum, L. Daniel, and G. B. de Quirós Fernán, “Mongodb: an open source alternative for HL7-CDA clinical documents management,” in *Proceedings of the Open Source International Conference (CISL’13)*, Buenos Aires, Argentina, 2013.
- [41] K. Kaur and R. Rani, “Managing data in healthcare information systems: many models, one solution,” *Computer*, vol. 48, no. 3, pp. 52–59, 2015.
- [42] B. Cyganek, M. Graña, B. Krawczyk et al., “A survey of big data issues in electronic health record analysis,” *Applied Artificial Intelligence*, vol. 30, no. 6, pp. 497–520, 2016.
- [43] D. V. Dimitrov, “Medical internet of things and big data in healthcare,” *Healthcare Informatics Research*, vol. 22, no. 3, pp. 156–163, 2016.
- [44] F. Firouzi, B. Farahani, M. Ibrahim, and K. Chakrabarty, “Keynote paper: from EDA to IoT eHealth: promises, challenges, and solutions,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 12, pp. 2965–2978, 2018.
- [45] S. Sakr and A. Elgammal, “Towards a comprehensive data analytics framework for smart healthcare services,” *Big Data Research*, vol. 4, pp. 44–58, 2016.
- [46] Y. Zhang, M. Qiu, C. W. Tsai, M. M. Hassan, and A. Alamri, “Health-CPS: healthcare cyber-physical system assisted by cloud and big data,” *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2015.
- [47] S. K. Sharma and X. Wang, “Live data analytics with collaborative edge and cloud processing in wireless IoT networks,” *IEEE Access*, vol. 5, pp. 4621–4635, 2017.
- [48] M. S. Hadi, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, “Patient-centric cellular networks optimization using big data analytics,” *IEEE Access*, vol. 7, pp. 49279–49296, 2019.
- [49] S. El-Sappagh, F. Ali, S. El-Masri, K. Kim, A. Ali, and K. S. Kwak, “Mobile health technologies for diabetes mellitus: current state and future challenges,” *IEEE Access*, vol. 7, pp. 21917–21947, 2018.
- [50] Z. Hong, W. Chen, H. Huang, S. Guo, and Z. Zheng, “Multi-hop cooperative computation offloading for industrial IoT-edge-cloud computing environments,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 12, pp. 2759–2774, 2019.
- [51] D. C. Yacchirema, D. Sarabia-Jácome, C. E. Palau, and M. Esteve, “A smart system for sleep monitoring by integrating IoT with big data analytics,” *IEEE Access*, vol. 6, pp. 35988–36001, 2018.
- [52] M. H. U. Rehman, V. Chang, A. Batool, and T. Y. Wah, “Big data reduction framework for value creation in sustainable enterprises,” *International Journal of Information Management*, vol. 36, no. 6, pp. 917–928, 2016.
- [53] S. Kumar and M. Singh, “Big data analytics for healthcare industry: impact, applications, and tools,” *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 48–57, 2018.
- [54] H. Jiang, Y. Chen, Z. Qiao, T.-H. Weng, and K.-C. Li, “Scaling up MapReduce-based big data processing on multi-GPU systems,” *Cluster Computing*, vol. 18, no. 1, pp. 369–383, 2015.
- [55] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [56] K. R. Ghani, K. Zheng, J. T. Wei, and C. P. Friedman, “Harnessing big data for health care and research: are urologists ready?” *European Urology*, vol. 66, no. 6, pp. 975–977, 2014.
- [57] J. Roski, G. W. Bo-Linn, and T. A. Andrews, “Creating value in health care through big data: opportunities and policy implications,” *Health Affairs*, vol. 33, no. 7, pp. 1115–1122, 2014.
- [58] T. U. Mane, “Smart heart disease prediction system using improved K-means and ID3 on big data,” in *Proceedings of the 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, IEEE, Pune, India, pp. 239–245, February 2017.
- [59] K. Rahimi, D. Bennett, N. Conrad et al., “Risk prediction in patients with heart failure,” *JACC: Heart Failure*, vol. 2, no. 5, pp. 440–446, 2014.
- [60] D. Al-Jumeily, A. Hussain, C. Mallucci, and C. Oliver, *Applied Computing in Medicine and Health*, Morgan Kaufmann, Burlington, MA, USA, 2015.
- [61] F. J. Martin-Sanchez, V. Aguiar-Pulido, G. H. Lopez-Campos, N. Peek, and L. Sacchi, “Secondary use and analysis of big data collected for patient care,” *Yearbook of Medical Informatics*, vol. 26, no. 01, pp. 28–37, 2017.
- [62] A. Jindal, A. Dua, N. Kumar, A. K. Das, A. V. Vasilakos, and J. J. P. C. Rodrigues, “Providing healthcare-as-a-service using fuzzy rule based big data analytics in cloud computing,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1605–1618, 2018.
- [63] D. D. Luxton, “An introduction to artificial intelligence in behavioral and mental health care,” in *Artificial Intelligence in Behavioral and Mental Health Care*, pp. 1–26, Academic Press, Cambridge, MA, USA, 2016.
- [64] N. Mehta, A. Pandit, and S. Shukla, “Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study,” *Journal of Biomedical Informatics*, vol. 100, p. 103311, 2019.
- [65] Y. Wang, L. Kung, and T. A. Byrd, “Big data analytics: understanding its capabilities and potential benefits for healthcare organizations,” *Technological Forecasting and Social Change*, vol. 126, pp. 3–13, 2018.
- [66] N. Mehta and A. Pandit, “Concurrence of big data analytics and healthcare: a systematic review,” *International Journal of Medical Informatics*, vol. 114, pp. 57–65, 2018.
- [67] C. S. Kruse, R. Goswamy, Y. Raval, and S. Marawi, “Challenges and opportunities of big data in health care: a systematic review,” *JMIR Medical Informatics*, vol. 4, no. 4, p. e38, 2016.
- [68] S. Nazir, M. Nawaz Khan, S. Anwar et al., “Big data visualization in cardiology—a systematic review and future directions,” *IEEE Access*, vol. 7, pp. 115945–115958, 2019.
- [69] V. Palanisamy and R. Thirunavukarasu, “Implications of big data analytics in developing healthcare frameworks—a review,” *Journal of King Saud University—Computer and Information Sciences*, vol. 31, no. 4, pp. 415–425, 2019.
- [70] G. Harerimana, B. Jang, J. W. Kim, and H. K. Park, “Health big data analytics: a technology survey,” *IEEE Access*, vol. 6, pp. 65661–65678, 2018.
- [71] A. Celesti, O. Amft, and M. Villari, “Guest editorial special section on cloud computing, edge computing, internet of things, and big data analytics applications for healthcare industry 4.0,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 454–456, 2019.
- [72] J. Qadir, M. Mujeeb-U-Rahman, M. H. Rehmani et al., “IEEE access special section editorial: health informatics for the developing world,” *IEEE Access*, vol. 5, pp. 27818–27823, 2017.

- [73] G. Phillips-Wren, L. S. Iyer, U. Kulkarni, and T. Ariyachandra, "Business analytics in the context of big data: a roadmap for research," *Communications of the Association for Information Systems*, vol. 37, no. 1, p. 23, 2015.
- [74] H. J. Watson, "Tutorial: big data analytics: concepts, technologies, and applications," *Communications of the Association for Information Systems*, vol. 34, no. 1, p. 65, 2014.
- [75] D. Delen, *Real-World Data Mining: Applied Business Analytics and Decision Making*, FT Press, Upper Saddle River, NJ, USA, 2014.
- [76] M. Riabacke, M. Danielson, and L. Ekenberg, "State-of-the-art prescriptive criteria weight elicitation," *Advances in Decision Sciences*, vol. 2012, Article ID 276584, 24 pages, 2012.
- [77] Z. Pang, H. Yuan, Y.-T. Zhang, and M. Packirisamy, "Guest editorial health engineering driven by the industry 4.0 for aging society," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1709-1710, 2018.
- [78] D. Rajeshwari, "State of the art of big data analytics: a survey," *International Journal of Computer Applications*, vol. 120, no. 22, 2015.
- [79] S. Shafqat, S. Kishwer, R. U. Rasool, J. Qadir, T. Amjad, and H. F. Ahmad, "Big data analytics enhanced healthcare systems: a review," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 1754-1799, 2018.
- [80] M. Chen, S. Mao, Y. Zhang, and V. C. Leung, *Big Data: Related Technologies, Challenges and Future Prospects*, Springer, Berlin, Germany, 2014.
- [81] Z. Zhou, W. Gaaloul, P. C. K. Hung, L. Shu, and W. Tan, "IEEE access special session editorial: big data services and computational intelligence for industrial systems," *IEEE Access*, vol. 3, pp. 3085-3088, 2015.
- [82] Y. Wang and N. Hajli, "Exploring the path to big data analytics success in healthcare," *Journal of Business Research*, vol. 70, pp. 287-299, 2017.
- [83] M. Viceconti, P. Hunter, and R. Hose, "Big data, big knowledge: big data for personalized healthcare," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1209-1215, 2015.
- [84] Y. Zhang, L. Zhang, E. Oki, N. V. Chawla, and A. Kos, "IEEE Access special section editorial: big data analytics for smart and connected health," *IEEE Access*, vol. 4, pp. 9906-9909, 2016.
- [85] J. Gantz and D. Reinsel, "Extracting value from chaos," *IDC Iview*, vol. 1142, pp. 1-12, 2011.
- [86] C. Ngufor and J. Wojtusiak, "Learning from large-scale distributed health data: an approximate logistic regression approach," in *Proceedings of the ICML 13: Role of Machine Learning in Transforming Healthcare*, Atlanta, GA, USA, 2013.
- [87] R. Zhang, G. Simon, and F. Yu, "Advancing Alzheimer's research: a review of big data promises," *International Journal of Medical Informatics*, vol. 106, pp. 48-56, 2017.
- [88] I. Cano, A. Tenyi, E. Vela, F. Miralles, and J. Roca, "Perspectives on big data applications of health information," *Current Opinion in Systems Biology*, vol. 3, pp. 36-42, 2017.
- [89] <https://www.forbes.com/sites/louiscolombus/2017/12/24/53-of-companies-are-adoptingbig-data-analytics/#50bf384239a1>.
- [90] M. Bakratsas, P. Basaras, D. Katsaros, and L. Tassioulas, "Hadoop mapreduce performance on SSDs for analyzing social networks," *Big Data Research*, vol. 11, pp. 1-10, 2018.
- [91] P. Zikopoulos, D. Deroos, K. Parasuraman, T. Deutsch, J. Giles, and D. Corrigan, *Harness the Power of Big Data the IBM Big Data Platform*, McGraw Hill Professional, New York, NY, USA, 2012.
- [92] P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Osborne Media, New York, NY, USA, 2011.
- [93] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big data technologies: a survey," *Journal of King Saud University—Computer and Information Sciences*, vol. 30, no. 4, pp. 431-448, 2018.
- [94] V. Rajaraman, "Big data analytics," *Resonance*, vol. 21, no. 8, pp. 695-716, 2016.
- [95] M. Idris, S. Hussain, M. Ali et al., "Context-aware scheduling in MapReduce: a compact review," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 17, pp. 5332-5349, 2015.
- [96] H. Senger, V. Gil-Costa, L. Arantes et al., "BSP cost and scalability analysis for MapReduce operations," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 8, pp. 2503-2527, 2016.
- [97] <https://www.datasciencecentral.com/profiles/blogs/the-hadoop-ecosystem-hdfs-yarn-hivepig-hbase-and-growing>.
- [98] A. K. Bhadani and D. Jothimani, "Big data: challenges, opportunities, and realities," in *Effective Big Data Management and Opportunities for Implementation*, pp. 1-24, IGI Global, Pennsylvania, PA, USA, 2016.
- [99] N. Khan, I. Yaqoob, I. A. T. Hashem et al., "Big data: survey, technologies, opportunities, and challenges," *The Scientific World Journal*, vol. 2014, Article ID 712826, 18 pages, 2014.
- [100] S. Bahri, N. Zoghalmi, M. Abed, and J. M. R. S. Tavares, "Big data for healthcare: a survey," *IEEE Access*, vol. 7, pp. 7397-7408, 2019.
- [101] S. R. Sukumar, R. Natarajan, and R. K. Ferrell, "Quality of big data in health care," *International Journal of Health Care Quality Assurance*, vol. 28, no. 6, pp. 621-634, 2015.
- [102] B. Cleland, J. Wallace, R. Bond et al., "Insights into antidepressant prescribing using open health data," *Big Data Research*, vol. 12, pp. 41-48, 2018.
- [103] N. Agnihotri and A. K. Sharma, "Proposed algorithms for effective real time stream analysis in big data," in *Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP)*, pp. 348-352, IEEE, Wanknaghat, India, December 2015.
- [104] J. S. Rumsfeld, K. E. Joynt, and T. M. Maddox, "Big data analytics to improve cardiovascular care: promise and challenges," *Nature Reviews Cardiology*, vol. 13, no. 6, pp. 350-359, 2016.
- [105] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123-1131, 2014.
- [106] M. Nambu, K. Nakajima, M. Noshiro, and T. Tamura, "An algorithm for the automatic detection of health conditions," *IEEE Engineering in Medicine and Biology Magazine*, vol. 24, no. 4, pp. 38-42, 2005.
- [107] R. Platt, R. Carnahan, J. S. Brown et al., "The U.S. Food and drug administration's mini-sentinel program," *Pharmacoepidemiology and Drug Safety*, vol. 21, pp. 1-303, 2012.
- [108] E. S. Berner, *Clinical Decision Support Systems*, vol. 233, Springer Science+ Business Media, LLC, New York, NY, USA, 2007.
- [109] C. S. Mayo, J. M. Moran, W. Bosch et al., "American association of physicists in medicine task group 263: standardizing nomenclatures in radiation oncology," *International Journal of Radiation Oncology*Biophysics*Physics*, vol. 100, no. 4, pp. 1057-1066, 2018.

- [110] A. Kos and A. Umek, "Wearable sensor devices for prevention and rehabilitation in healthcare: swimming exercise with real-time therapist feedback," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1331–1341, 2018.
- [111] A. K. Sahoo, S. Mallik, C. Pradhan, B. S. P. Mishra, R. K. Barik, and H. Das, "Intelligence-based health recommendation system using big data analytics," in *Big Data Analytics for Intelligent Healthcare Management*, pp. 227–246, Academic Press, Cambridge, MA, USA, 2019.
- [112] T. R. Hoens, M. Blanton, A. Steele, and N. V. Chawla, "Reliable medical recommendation systems with patient privacy," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 4, pp. 1–31, 2013.
- [113] M. A. Hamburg and F. S. Collins, "The path to personalized medicine," *New England Journal of Medicine*, vol. 363, no. 4, pp. 301–304, 2010.
- [114] G. Wang, K. Jung, R. Winnenburg, and N. H. Shah, "A method for systematic discovery of adverse drug events from clinical notes," *Journal of the American Medical Informatics Association*, vol. 22, no. 6, pp. 1196–1204, 2015.
- [115] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, "Big data application in biomedical research and health care: a literature review," *Biomedical Informatics Insights*, vol. 8, Article ID BIL.S31559, 2016.
- [116] T. Saheb and L. Izadi, "Paradigm of IoT big data analytics in healthcare industry: a review of scientific literature and mapping of research trends," *Telematics and Informatics*, vol. 41, pp. 70–85, 2019.
- [117] L. Syed, S. Jabeen, and S. Manimala, "Telemammography: a novel approach for early detection of breast cancer through wavelets based image processing and machine learning techniques," in *Advances in Soft Computing and Machine Learning in Image Processing*, pp. 149–183, Springer, Cham, Switzerland, 2018.
- [118] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [119] G. Manogaran, C. Thota, D. Lopez, V. Vijayakumar, K. M. Abbas, and R. Sundarsekar, "Big data knowledge system in healthcare," in *Internet of Things and Big Data Technologies for Next Generation Healthcare*, pp. 133–157, Springer, Cham, Switzerland, 2017.
- [120] T. Heart, O. Ben-Assuli, and I. Shabtai, "A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy," *Health Policy and Technology*, vol. 6, no. 1, pp. 20–25, 2017.
- [121] P. Galetsi, K. Katsaliaki, and S. Kumar, "Values, challenges and future directions of big data analytics in healthcare: a systematic review," *Social Science & Medicine*, vol. 241, p. 112533, 2019.
- [122] A. Kankanhalli, J. Hahn, S. Tan, and G. Gao, "Big data and analytics in healthcare: introduction to the special section," *Information Systems Frontiers*, vol. 18, no. 2, pp. 233–235, 2016.
- [123] A. Belle, R. Thiagarajan, S. M. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," *BioMed Research International*, vol. 2015, Article ID 370194, 16 pages, 2015.