BIG DATA ANALYTICS IN HEALTH CARE: A REVIEW PAPER

Maria Mohammad Yousef

Department of Computer Science, Al-albayt University, Jordan

ABSTRACT

The application of big data in health care is a fast-growing field, with many discoveries and methodologies published in the last five years. Big data refers to datasets that are not only big but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Moreover, medical data is one of the most growing data, as it is obtained from Electronic Health Records (EHRs) or patients themselves. Due to the rapid growth of such medical data, we need to provide suitable tools and techniques in order to handle and extract value and knowledge from these datasets to improve the quality of patient care and reduces healthcare costs. Furthermore, such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. This paper presents an overview of big data content, sources, technologies, tools, and challenges in health care. It also intends to identify the strategies to overcome the challenges.

KEYWORDS

Big Data, Healthcare, Big data challenges, EHRs.

1. INTRODUCTION

Nowadays there is increasing in the details and data presented through the advancements in technologies and the internet. Anything ranging from consumer names and addresses to products available, to purchases made, to employees hired, etc. has become necessary for day-to-day continuity. With the improvement in storage capacities and techniques of data collection, enormous amounts of data have become easily available. Every second, more and more data is being produced and needs to be stored and analyzed in order to obtain value. Furthermore, data have become cheaper to store, so business companies and organizations need to get as much value as possible from the huge amounts of data collected daily.

Data sets increase rapidly because they are frequently gathered by many information-sensing devices such as mobile devices, aerial (remote sensing), software logs and records, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks [1]. Thus, big data is a field that explains methods to analyze, systematically obtain information from, and how to deal with data sets that are too large or complex to be dealt with by traditional data processing applications.

The healthcare industry is one of the most important industries. It is also one of the world's largest and fastest-growing industries it can produce and handles data at a staggering speed, but different electronic health records (EHRs) collect data in different structures: structured, unstructured, and semi structured. This variety can pose a challenge when seeking veracity or quality assurance of the data. The EHRs can provide a rich source of data, ready for analysis to improve our understanding of disease mechanisms, as well as better and personalized health care, but the data structures pose a problem to standard means of analysis. So, there is a need for converting the raw

data into significant and actionable information by using big data analytics tools [2].

Big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software or popular tools and methods [3]. Accordingly, big data in healthcare is overwhelming not only because of its volume of data sets but also because of the variety of data types and the speed at which it must be managed.

The purpose of this systematic review is to provide a summarize of big data analytics in healthcare. First, we define and explain the definition of big data and the characteristics of big data analytics in the healthcare domain. Then we describe the big data types in healthcare. Third, we provide examples of big data analytics in healthcare. Fourth, we compile a list of challenges and opportunities faced by big data analytics in health care. Finally, we offer conclusions and future directions.

2. BACKGROUND

2.1. Defining Big Data

The concept of "big data" is not new, however, the way it is defined is continually changing. Many authors have provided big data definitions such as Zulkarnain et al. [4] define Big Data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze". Likewise, Kaislere et al. [5] say "Big data is data too big to be handled and analyzed by traditional database protocols such as SQL". Moreover, the authors in [6] present big data as a collection of data elements whose size, speed, type, and/or complexity require an attempt to use and discover new hardware and software tools to successfully store, examine, and visualize the data. Accordingly, Big Data points to large, complex datasets that are exceeding the capabilities of the traditional data management system to store, manage and process them.

2.2. Big Data Characteristic

As with all big things, if we want to manage them, we need to characterize them to organize our understanding. The three Vs (volume, velocity, and variety) are known as the main characteristics of big data. These features are key to understanding how we can measure big data. The volume of the data refers to its size, and how huge it is. While the velocity points to the rate with which data is changing, or how often it is created. Finally, the variety involves several formats and types of data, as well as the different kinds of uses and ways of analyzing the data [7]. The characteristics are described below in Fig. 1.

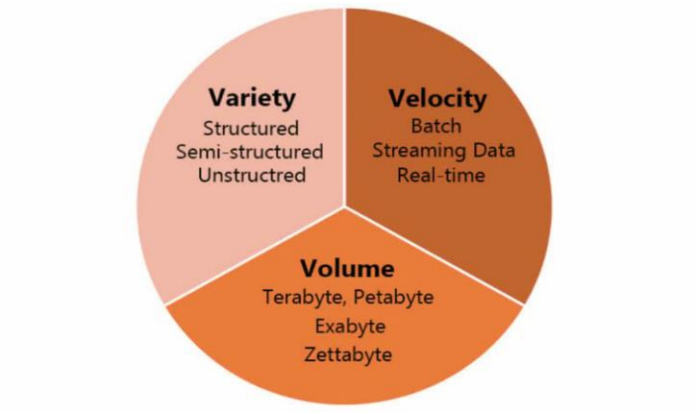

Figure 1. The Big Data Characteristic.

As shown in Figure.1. Big Data can be described by the following characteristic:

- **Data volume**: This is the first and most important attribute of big data. Big data can be quantified by size in TBs or PBs, as well as even the number of records, transactions, reports, or files. The volume of data used to play important role in storage and processing. However, many factors can contribute to the volume rise in data, it could amount to hundreds of terabytes or even petabytes of information generated anywhere. As displayed in [8], the number of data sources for an organization is growing day by day. And therefore, more data sources consisting of enormous datasets increase the volume of data, which needs to be analyzed. As noted in [8], Fig. 2 shows that the volume of data stored in the world would be more than 40 zettabytes (〖10〗^21 Byte) by 2020.

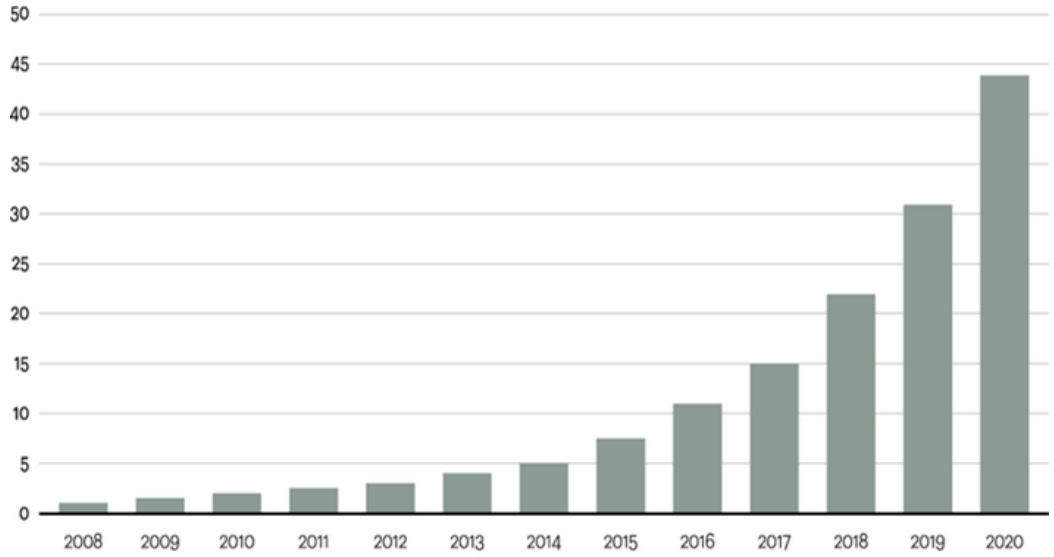

Figure 2. Data volume growth by year in zettabytes

- **Data Velocity:** Points to the speed at which new data is generated and the speed at which data flows around, Hence, increasing speed in data processing, storage, and analysis by relational databases. Moreover, Velocity assists organizations understand the relative growth of their big data and how quickly that data reaches sourcing users, applications, and systems. Some activities are very important and require immediate responses, which is why quick processing maximizes effectiveness. For time-sensitive processes such as fraud detection, Big Data flows must be analyzed and used as they stream into the organizations to maximize the usefulness of the information. An illustration of data that is generated with great velocity would be Twitter messages or Facebook posts.

- **Data Variety:** The next aspect of Big Data is its Variety. Which indicates the type of data that big data can contain. Big data is not always structured data. That means Big Data consists of any type of data, this data may be structured or unstructured such as text, sensor data, speech recordings, video, click streams, log files, and so on. Because Big Data contains data of different types and sources, Dealing with a variety of structured and unstructured data increases the complexity of both analyzing and storing Big Data. One of the goals of big data is to employ technology to take this unstructured data and obtain an understanding of it.

2.3. Big Data in Health Care

In the healthcare field, the progress in information technology and the capability of storing more data have driven countries and governmental institutions to computerize health records and produced the Electronic Health Record (EHR) or Electronic Medical Record (EMR). Big data analytics in medicine and healthcare allows analysis of the large datasets from thousands of patients, identifying clusters and correlation between datasets, Moreover improving predictive models using data mining techniques. As the healthcare industry focuses on improvements in order to save patients' lives, Big Data Analytics can play an important role in improving the services provided to healthcare by:

- Managing hospital performance
- Prevent epidemics, cure disease, and decrease costs.
- Increase transparency and efficiency in early disease diagnosis
- Enhancing clinical outcomes
- Engaging patients and family
- Give patients more personalized treatment and enhance the overall patient experience.

3. RESEARSH METHODOLGY

A systematic review was conducted for obtaining related literature from various sources, focusing on the following aims.

1. Determine the different perspectives to defining big data and its applications in healthcare.
2. To explore the sources of Big Health Data.
3. Discover Big Data analytical techniques and technologies in healthcare.
4. Introduce approaches to reduce the challenges of implementing big data within healthcare.

By considering these goals in detail, this review will make a significant contribution to understanding the overall meaning and the future applying of Big Data techniques and applications in the healthcare domain.

3.1. Information Source

We searched four databases to find related research articles: (1), IEEE Xplore, (2) ScienceDirect, (3) Springer, and (4) Scopus. In searching these databases, we used the main keywords “big data” or “big data analytics”, and “healthcare” or “medicine” or “biomedicine”.

3.2. Selection Criteria

To select the literature for inclusion in the literature review, we depending on the following inclusion criteria:

1. Only papers published in English and between 2013 and 2020
2. Papers that deal with Big Data analytics in healthcare.
3. Papers that discuss the design and use of a big data application in the biomedical and health-care domains.
4. Research papers that discuss the challenges of big data in healthcare.
5. Research surveys that point to the benefit effectiveness of Big Data technologies in the health care domain.

3.3. Exclusion Criteria

The following exclusion criteria were used to filter out irrelevant papers:

1. The paper was a tutorial or a course material.
2. The paper did not discuss any specific big data applications (e.g., general comments about big data).
3. Papers that focus primarily on traditional analytics in healthcare.

3.4. Study Selection

The proposed procedure framework for search and select the research elements is presented in Fig. 3. Also, the selection, examination, and filtering process for studies are described in the next four phases.

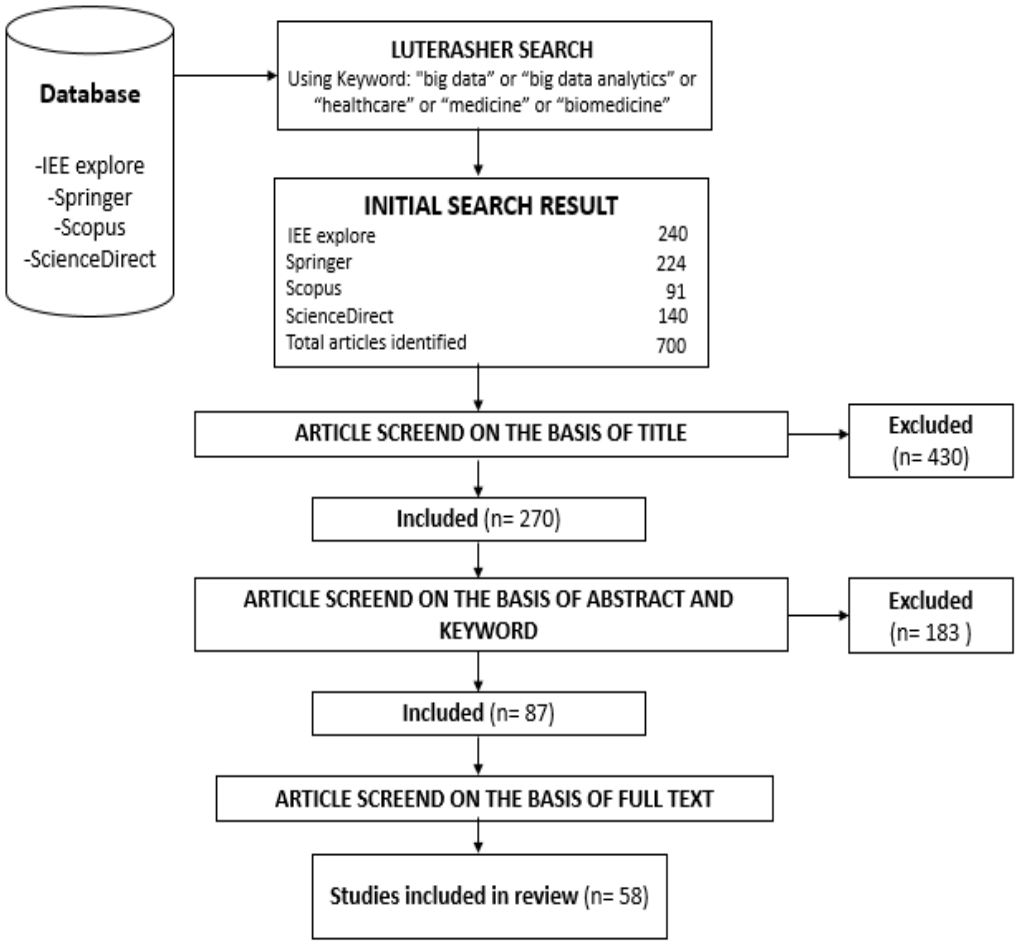

Figure 3. Research Process.

As shown in the previous figure, we depend on several consecutive steps to obtain the most useful and relevant studies for this proposed review by the following steps:

- Began the search process for publications on different databases based on the main keywords such as: "big data" or "big data analytics", and "healthcare" or "medicine" or "biomedicine".

- All potentially related papers were collected by scrutiny of the title to identified papers and selection of the significant articles based on selection criteria. This initial search resulted in 270 from 700 papers.
- In the second search, we perusal the papers that were not eliminated in the previous phase for the review, where we screened the papers based on the abstract, keywords, and the abovementioned inclusion and exclusion criteria, and consequently selected 87 papers.
- Finally, we evaluated the final selection by reading the content of the papers in more detail, and consensus to review 58 papers for this study.

4. RESULT

The literature included in this study contains essentially descriptive papers and studies. Based on the main research goals, the content from these studies was extracted and the papers were classified into many groups: Big Data analytics definition and concepts, sources of Big Data in healthcare, Big Data techniques for healthcare analytics, application and potential benefits of Big Data in healthcare and challenges in Big Data analytics in healthcare. The next section summarizes the conclusions in each of these categories.

4.1. Big Data Analytics Concept

With the evolving of technology and the increasing numbers of data flowing in and out of organizations daily, there has become a demand for faster and more efficient ways of analyzing such data.

The authors in [9] explained that Big Data is ineffective in a vacuum. So, its potential value is only obtained when used in decision making. To enable an organization to acquire knowledge and use it in decision-making, organizations need effective methods to apply large amounts of fast-moving data of various types and forms to analyze and benefit from it. The analytics concept refers to techniques used to analyze and acquire knowledge from big data. Thus, big data analytics can be viewed as a sub-process in the overall process of 'knowledge extraction' from big data.

As discussed in [10], Big data analytics refers to using advanced techniques and tools for analyzing and examining very large and various data sets that combine structured, semi-structured, unstructured data from various sources and in different sizes from terabytes to zettabytes in aims to obtain helpful information included within the data and will also help identify the data that is most important to the business and future business decisions. Instead of: hidden patterns, associations, market trends, and consumer preferences.

4.2. Source of Health Care Big Data

Data that is obtained, collected, and stored in the healthcare industry may be are disorganized and distributed, coming from various sources and having different structures and forms. Healthcare Big Data involves data on physiological, behavioral, clinical, environmental illness, medical imaging, disease administration, medicine prescription records, nutrition, or exercise parameters [11]. However, most of the studies reviewed agreed on common sources of big data in the healthcare field, which are as follows:

- **Electronic Health Records (EHRs):** An electronic copy of a patient's medical record that is maintained by the service provider over time. The EHRs can be containing data related to the results of clinical and administrative meetings between the service provider (doctor, nurse, etc.) and the patient [12]. EHRs may include a range of data including demographics,

medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, personal statistics like age and weight, billing information, and active medical problems [13].

- **Electronic Medical Records (EMRs):** EMRs are similar to EHRs, they are digital records of patient health information; it is a digital version of a patient's information maintained in the form of a chart, and it contains the patient's medical and treatment history from one clinic. Usually, this digital record stays in the doctor's office and does not get shared. If a patient switches doctors, his or her EMR is unlikely to follow. However, this paper chart is stored in clinician offices, clinics, and hospital databases [14].

- **Patient-Reported Outcomes (PROs):** Defined as a report coming directly from patients about their health condition and treatment which are based on a patient's perception of a disease and its treatment. This report includes a range of outcomes such as symptoms, health status, and health-related quality-of-life [15].

- **Data collected from wearable sensors:** The majority of wearable devices allow the collection of biochemical, physiological, and motion-sensing data such as (Heart rate, Steps walked, Blood pressure, etc.). So, it can collect patient health data and have data sharing capabilities [16]. The analysis of this type of data, when integrated with electronic health records, can support health monitoring and diagnosis for different chronic conditions.

- **Data extraction from social networking tools (social media):** Patient posts on online social media such as Facebook, Instagram, Twitter, etc. can be extracted to obtain information about disease trends, patients' satisfaction, happiness, interests, and feelings. Twitter is a common example where data analytics methods have been used for disease monitoring and health-related trends (e.g. [17]).

4.3. Big Data Analytical Techniques and Tools in Healthcare

Different types of healthcare data are difficult to analyze due to their dynamicity and complexity, such as medical images (X-ray, Magnetic Resonance Imaging MRI images), biomedical signals (EEG, ECG, EMG, etc.), audio records, multi-dimensional healthcare data, written prescriptions and structured data from EMRs and EHRs [18]. Moreover, there is a lack of analytical approaches that can handle such unstructured data and help decision making [19]. In this review, we summarize the literature that considers some of the analytical strategies and tools which can apply to healthcare and medicine.

As reported by (Asante-Korang and Jacobs, 2016) [20], there are 4 types of Big Data Analytics: Descriptive, Diagnostic, Predictive, and Prescriptive Analytics. According to the literature, predictive analytics is the most popular in the healthcare industry as they are used to detect early signs of patient deterioration, predict high-cost patients, re-admission, what might happen (when the patient's condition worsens), adverse events, and treatment improvement for diseases affecting the multi-organ system as discussed in [21, 22,23]. Moreover, Healthcare organizations have observed improved quality of care after adopting several Big Data analytics techniques that helped enhance the ability of the healthcare sectors to predict epidemics and treat disease. Table 1. Summarizes some of the Big Data Analytical Techniques used in healthcare.

Table 1. Summarizes some of the Big Data analytical techniques used in healthcare.

Analytic Technique	Description	Healthcare Application	Studies By
Cluster Analysis (CA)	(CA) is a commonly used applied statistical technique in Big Data, that helps to reveal hidden structures and "clusters" found in large data sets. The CA aims to place objects into groups or "clusters" that have been suggested by the data and not defined a priori, the objects in the same cluster are similar to each other in some attributes, while, objects in different clusters tend to be different. K-Mean method is considered to be one example of clustering in big data	-Identify cost change patterns of patients with end-stage renal disease (ESRD) who initiated hemodialysis (HD) by applying different clustering method.	ISMAIL et al.,[21]
Data Mining	Is the ability to extract useful knowledge hidden in the large volume of data by applying new techniques, for discovering understandable patterns and correlations from data and use it in making decisions and Prediction of likely outcomes such as Association Rule Learning and Regression Analysis.	-Determination of epidemics; - detection some diseases - management of healthcare and measuring the effectiveness of certain treatments	Jothi et al.,[22]
Graph Analytics	Graph Analytics are analytic tools used to define the strength and discover the direction of relationships between objects in a graph. In big data, graph analytics focus of understand, codify, and visualize pairwise relationships that exist between two objects at a time and structural attributes of the graph as a whole.	-Analysis of hospital performance across various quality measures	Nisar et al.,[23]
Natural Language Processing (NLP)	Is the subfield of artificial intelligence (AI) that concerned with analyzing, understanding, and interpreting written text and spoken language, as well as using natural languages for communicating with computers	-Extract clinical concept (e.g. diagnosis, procedure, and symptoms) from electronic medical record, patient discharge summaries, and lab report.	Gudivada et al., [24]
Neural Networks	A neural network is a series of algorithms that attempt to identify underlying relationships in a large amount of data through a process that mimics the way that the human brain works. The purpose of a neural network is to learn how to discover patterns from the data. Once the neural network has been trained on samples of data, it can make predictions by detecting similar patterns in future data.	-Prediction of patients future disease -Diagnosis of chronic diseases;	Wang et al.,[25]

Machine Learning	Is a discipline within the area of Artificial Intelligence Moreover, is a field of research that formally focuses on the theory, performance, and properties of learning systems and algorithms Which can provides computers with the ability to recognize patterns from raw data to make predictions such as Decision Tree, Naive Bayes and Random Forest algorithems.	-Microsoft's InnerEye application employs machine learning to differentiate between tumors and healthy anatomy using 3D radiological images that assist medical experts in radiotherapy and surgical planning, among other things.	Qiu et al., [26]

On the other hand, Hadoop and MapReduce are used as tools for processing big data. On the other hand, Hadoop was proposed by Doug Cutting and Mike Cafarella in the year 2002 when they both started to work on an open-source web search engine (Apache Nutch) [27]. Apache Nutch project was the process of building a search engine system that can index 1 billion pages. Apache Nutch producers recognized the limitation of nutch and the challenge to reach a very huge number of webpages on the internet [28]. So, they understood that their project architecture will not be capable enough to work around billions of pages on the web. So they were looking for a possible solution that can reduce the implementation cost as well as the problem of storing and processing huge datasets [29].

Apache Hadoop is an open-source software framework for storing and processing huge clusters of data. Hadoop consists of a two-component Hadoop Distributed File System (HDFS), and Map Reduce framework [30]. The HDFS uses the cluster architecture to facilitate the partitioning of the big data to multiple machines (nodes), such as PCs and servers, which gives the ability to store a huge volume of data on thousands of nodes. The second component MapReduce is defined as a programming model for processing the large data sets stored in the Hadoop File System (HDFS) in a distributed fashion over several machines or nodes. There are two phases in MapReduce, the "Map" phase and the "Reduce" phase [31]. The system splits the input data into multiple chunks, each of which is assigned a map task that can process the data in parallel. Each map task reads the input as a set of (key, value) pairs and produces a transformed set of (key, value) pairs as the output. The framework shuffles and sorts outputs of the map tasks, sending the intermediate (key, value) pairs to reduce task, which groups them into final results.

4.4. Big Data Analytics Challenges in Health Care

Big Data helps organizations, individuals, countries, and the world to create new growth opportunities, but it also poses significant challenges that could offset any potential gains, such as the loss of privacy and confidentiality, and the lack of appropriate IT infrastructure. Also, many of the big data tools are open source and free to use, which could provide the opportunity for intrusive operations, hackers, and data theft. Some literatures [32-38] discuss obstacles in the development of big data in healthcare applications. The key challenges are listed as follows:

1. **Privacy and Security**: Privacy and security are a key concern for individuals and corporations that hold information/data about people, products, activities, etc. Medical data obtained by healthcare providers from individuals and their medical records may contain private and confidential data [32]. Wherefore, protecting the patient's information must be handled with enormous care from harm and hacker. When we use big data, many tools applied to analytics and data processes are open source and do not include all security measures [33]. Therefore, the primary justification for protecting personal privacy is to protect the interests of individuals. In order to overcome these challenges, some approaches are used to enhance the security level and obtain some confidentiality. First, Employing

security measures, including strong encryption of data, validation of the source of data, access control, and authentication, where authentication is one of the measures for securing the data and maintaining confidentiality.

2. **Storage and Processing Issues:** Doubtlessly, the most obvious challenge associated with big data is simply storing and analyzing the huge amount of data. Nowadays, data grow significantly whenever a new storage technology is invented due to the huge amount of data collected and transferred by social media, healthcare providers, business transactions, and stock markets [34]. Moreover, this data is not just high on volume, but it also includes data of varied kinds that is generated periodically. With the rate of data explosion, the biggest challenge in dealing with this big data is that the present or traditional systems are unable to store and process data of this size and kind [35]. Therefore, the storage problem can be solved by making use of cloud computing. This would enable small and medium-sized hospitals and care organizations to eliminate cost and data storage issues.

3. **Data Ownership:** Data ownership represents a crucial and ongoing problem in big data applications in healthcare and other areas. Though petabytes of medical records generally belong to the healthcare providers, governmental healthcare systems, or hospital in which they were created, but the information in it is not owned by them [36]. On the other hand, patients believe that they own the data. This dispute may be ended in the legal system to resolve the ownership issues unless healthcare providers receive written approval from patients before using data for experiences or research objectives.

4. **Skills Requirement:** A data analyst is a professional whose work involves collecting, cleaning, visualizing, and transforming or modeling raw data into the blocks of information that are used by marketers, developers, and even healthcare providers [37]. One of the most important challenges in dealing with big data is the skills required for individuals to works in the big data field. A recent study [38] examined the required skills to deal with big data and concluded that the skills you need to work with big data will involve analytical capabilities.

5. CONCLUSION

The paper first defined what is meant by big data. We presented various definitions of big data, highlighting the fact that size is only one dimension of big data. Other dimensions, such as velocity and variety, are equally important. The studies reviewed showed that big data in the healthcare industry is obtained from several sources such as results of medical examinations, hospital records, medical devices, and records of patients. For better treat disease and diagnosis in medical, the role of big data is one where it can construct better predictive models using tools with the ability to analyze and process this vast amount of data. Finally, a discussion has been made of some challenges that face individuals and organizations in the process of utilizing big data in healthcare, such as data ownership, privacy and security, storage and processing issues, and skills requirements.

6. LIMITATIONS

While the proposed Review covers details about Big Data analytics and its applications in healthcare and medicine, however, we face a few limitations. First, the contents of this research consist of a systematic review of the current state of Big Data technology in healthcare, but it does not get into consideration the technical details concerning the implementation and outcomes achieved in each of the studies reviewed. Second, there is heterogeneity in the documentation since

the literature includes various sources of information on the meaning of Big Data, methods of Big Data analytic, and their techniques and challenges in healthcare. Finally, despite the use of a systematic strategy for review, the inclusion of studies on big data analytics in 'healthcare' for this review was based on personal experience and knowledge, hence the cross-reference literature were also examined for this review.

7. FUTURE OUTLOOK

Big data analytics in medicine and healthcare is a very encouraging process of integrating, examining, and analyzing enormous amounts of complex heterogeneous data with different types: biomedical data, medical data, electronic health records data (EHRs), and experimental data. The combination of such various data makes big data analytics weave many fields, such as bioinformatics, medical imaging, sensor informatics, medical informatics, health informatics, and computational biomedicine. As further work, we plan to study the various improvements in big data analytic systems and databases. Also, we will attempt to produce a new high-performance data management system by depending on open source platform such as Apache Hadoop MapReduce, which can assist heterogeneous datasets and uses memory and other hardware resources in a more efficient way to reveal hidden patterns and novel knowledge from the data in a great execution speed.

REFERENCES

[1] S. Singh, T. Firdaus, and A. K. Sharma, (2015) "Survey on Big Data Using Data Mining", International Journal of Engineering Development and Research, Vol. 3, No.4, pp. 135–143.

[2] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kanshik, (2019) "Big data in healthcare: management, analysis and future prospects", Journal of Big Data. Springer International Publishing, Vol. 6, No.1.

[3] I. K. Subagja, N. Amaliyah, U. Hiermy, B. T. Rahardjo, E. L. Lydia, K. Shakar, and P. T. Naguyen, (2019) "Evaluation of big data analytics in medical science", International Journal of Engineering and Advanced Technology, Vol. 8, (6 Special Issue 3), pp. 717–720.

[4] N. Zulkarnain, M. Anshari, and A. Definition, (2016) "Big Data : Concept, Applications, & Challenges", International Conference on Information Management and Technology, pp. 307–31.

[5] S. H. Kaisler, F. J. Armour, and A. J. Espinosa, "Introduction to the big data and analytics: concepts, (2016) techniques, methods, and applications minitrack", Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 1059–1060.

[6] J. Amudhavel, V. Padmapriya, V. Gowri, K. Lakshmipriya, K. P. Kumar, and B. Thiyagarajan, (2015) "Perspectives, motivations and implications of big data analytics.", In Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology, pp. 1-5.

[7] M. Pospiech, and C. Felden, (2012) "Big data - A State-of-the-Art", 18th Americas Conference on Information Systems 2012, AMCIS 2012, Vol. 5, pp. 3918–3928.

[8] S. Brown, (2020) "Data Characteristics", A multidisciplinary journal of global macro trends BIG, Vol.3, No.6, pp. 19–27.

[9] M. I. Razzak, M. Imran, and G. Xu, (2020) "Big data analytics for preventive medicine", Neural Computing and Applications. Springer London. Vol. 32, No.9, pp.4417-4451.

[10] G. Chen, and M. Islam, (2019) "Big Data Analytics in Healthcare", Proceedings - 2019 2nd International Conference on Safety Produce Informatization, IICSPI 2019, Vol. 2015, pp. 227–230.

[11] S. G. Alonso, I. Torre-Diez, J. Rodigues, S. Hamorioui, and M. L. Coronado, (2017) "A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector.", Journal of Medical Systems, Vol. 41, No. 11, pp.183.

[12] A. P. Ambinder, (2005) "Electronic Health Records By", Journal of Oncology Practice, Vol. 1, No. 2, pp. 57–63.

[13] R. S. Evans, (2016) "Electronic Health Records: Then, Now, and in the Future", Yearbook of medical informatics, pp. S48–S61.

[14] Z. Liang et al., (2014) "Deep learning for healthcare decision making with EMRs", Proceedings - 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014, (Cm), pp. 556–559.
[15] M. Calvert et al., (2015) "Putting patient-reported outcomes on the Big Data Road Map", Journal of the Royal Society of Medicine, Vol.108, No.8, pp. 299–303.
[16] M. Uddin, and S. Syed-abdul, (2020) "Data analytics and applications of the wearable sensors in healthcare: An overview", Sensors (Switzerland), Vol.20, No.5.
[17] J. Zhang, N. Xue, and X. Huang, (2016) "A Secure System for Pervasive Social Network-Based Healthcare", IEEE Access, Vol.4, pp. 9239–9250.
[18] A. Kankanhalli et al., (2016) "Big data and analytics in healthcare: Introduction to the special section", Information Systems Frontiers, Vol.18, No.2, pp. 233–235.
[19] B. Van Calster et al., (2019) "Predictive analytics in health care: how can we know it works?" Journal of the American Medical Informatics Association, Vol.26, No.12, pp. 1651–1654.
[20] A. Asante-Korang, and J. P. Jacobs, (2016) "Big Data and paediatric cardiovascular disease in the era of transparency in healthcare", Cardiology in the Young, Vol.26, No.8, pp. 1597–1602.
[21] A. ISMAIL, S. ABDLERAZEK, I. M. EL-HENAWY, (2020) "Big Data Analytics in Heart Diseases", Journal of Theoretical and Applied Information Technology, Vol. 98, No.11, pp. 1970–1980.
[22] N. Jothi., N. A. Rashid, and W. Husain, (2015) "Data Mining in Healthcare - A Review", Procedia Computer Science. Elsevier Masson SAS, Vol.72, pp. 306–313.
[23] M. U. Nisar, A. Fard, and J. A. Miller, (2013) "Techniques for graph analytics on big data", Proceedings - 2013 IEEE International Congress on Big Data, BigData 2013, pp. 255–262.
[24] V. N. Gudivada, D. Rao, and V. V. Raghavan, (2015) "Big Data Driven Natural Language Processing Research and Applications", Handbook of Statistics. Elsevier Inc, vol.33, pp.203-238.
[25] Y. Wang et al., (2014) "Energy efficient neural networks for big data analytics", Proceedings -Design, Automation and Test in Europe, DATE, pp. 2–3.
[26] J. Qiu et al., (2016) "A survey of machine learning for big data processing", Eurasip Journal on Advances in Signal Processing. EURASIP Journal on Advances in Signal Processing, Vol.1.
[27] M. R. Ghazi, and D. Gangodkar, (2015) "Hadoop, mapreduce and HDFS: A developers perspective", Procedia Computer Science. Elsevier Masson SAS, Vol.48, pp. 45–50.
[28] A. B. Patel, M. Birla, and U. Nair, (2012) "Addressing big data problem using Hadoop and Map Reduce", 3rd Nirma University International Conference on Engineering, NUiCONE 2012, pp. 6–8.
[29] A. Alam, and J. Ahmed, (2014) "Hadoop architecture and its issues", Proceedings - 2014 International Conference on Computational Science and Computational Intelligence, CSCI 2014, Vol.2, pp. 288–291.
[30] K. Dwivedi, and S. K. Dubey, (2014) "Analytical review on Hadoop Distributed file system", Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit, pp. 174–181.
[31] I. A. T. Hashem et al., (2020) "MapReduce scheduling algorithms: a review", Journal of Supercomputing, Vol.76, No.7, pp.4915-4945.
[32] A. L'heureux, K. Grolinger, J. F. Elyamany, and M. A. Capretz, (2017) "Machine learning with big data: Challenges and approaches", IEEE Access, Vol. 5, pp. 7776-7797.
[33] A. Ghazvini, and Z. Shukur, (2013) "Security Challenges and Success Factors of Electronic Healthcare System", Procedia Technology. Elsevier B.V., Vol.11, pp. 212–219.
[34] M. Padgavankar, and S. Gupta, (2014) "Big data storage and challenges", International Journal of Computer Science and Information Technologies, Vol.5, No.2, pp. 2218–2223.
[35] J. Li et al., (2014) "The overview of big data storage and management", Proceedings of 2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2014, pp. 510–513.
[36] P. Kostkova, "Who Owns the Data ?", Open Data for Healthcare, vol.4.
[37] R. Nambiar et al., (2013) "A look at challenges and opportunities of Big Data analytics in healthcare", Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013, pp. 17–22.
[38] A. Gardiner et al., (2018) "Skill Requirements in Big Data: A Content Analysis of Job Advertisements", Journal of Computer Information Systems. Taylor & Francis, Vol.58, No.4, pp. 374–384.