Journal of Cardiovascular Disease Research

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

Big Data Analytics in Healthcare: An Analytical Survey

Kiran Kumain,
Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University,
Dehradun, Uttarakhand India 248002

DOI: 10.48047/jcdr.2021.12.02.90

Abstract:

The topic of big data is one that is expanding rapidly in the modern world, with many new findings and approaches being published over the course of the past few years. Big Data refers to a compilation of data that is not only enormous in amount but also expanding at an exponential rate over time. Because of its enormous quantity and high degree of complexity, none of the conventional methods for managing data are able to store it or handle it in an effective manner. Big data analytics allows firms to make better use of their data and discover new opportunities as a result. The term "big data" refers to datasets that are not only large in size but also high in diversity and velocity, which makes it challenging to manage them using the tools and methods that have traditionally been employed. Reviewing the healthcare analytics that can be performed using big data, as well as the technologies and techniques that can be utilized in the future scope related to big data, is the focus of this particular piece of writing.

Keywords: Big data, big data in Healthcare, Big data analytics, challenges

Introduction:

These days, there is a rise in the quantity of data that is made available as a result of developments in technology and the expansion of the internet. In order to maintain normal operations, it is now important to keep a record of everything from customers' identities and places of residence to the products that are available and the purchases that have been made. The expansion of data collecting systems' storage capacities has made it possible to quickly access enormous amounts of previously inaccessible information. Data is produced at a rate of one byte per second, and in order to derive any benefit from it, it must first be saved, and then analyzed. Because of this, a notion known as "big data" came into being. Big data refers to data sets that are extremely large in volume, are generated very quickly, and contain a significant deal of variation. In terms of the 5Vs, the big data have five distinguishing qualities [1]. As time goes on, there is an increase in the amount of information and specifics that are provided as a result of developments in technology and the internet. For the sake of day-to-day continuity, keeping track of anything from the names and addresses of customers to the items that are accessible to the purchases that have been made to the personnel that has been hired has become vital.

Because of advancements in both storage capacities and methods of data collection, large volumes of previously inaccessible data are now within reach of virtually everybody. Data is being produced at an exponential rate every second, and in order to derive any value from it, it must first be stored and then analyzed. In addition, the cost of storing data has decreased, which means that businesses and other types of organizations need to derive as much value as they can from the massive volumes of data that are collected every day. The information and insights gained from big data, when properly stored and analyzed using the appropriate tools, have the potential to make essential aspects of our social infrastructure, such as healthcare and public safety, more aware, interactive, and effective. The term "big data" refers to datasets that are too large to be stored, managed, or analyzed using the capabilities that are typically included in database software. Therefore, conventional methods of managing and visually representing the data that has been stored are not suitable for handling this kind of data.

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

5V's of Big Data:

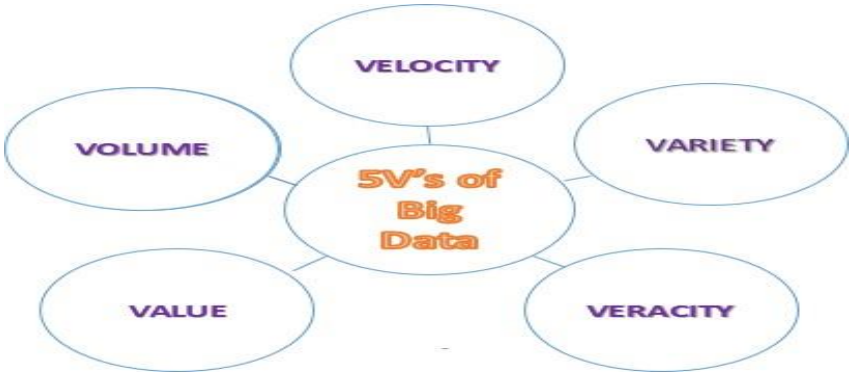

Figure -1: 5V's of Big Data

The term "volume" alludes to the enormous amount of data that is produced each day from a variety of fields. The amount of data being stored and processed used to be a significant factor in both areas. The amount of big data can be measured in terabytes or petabytes, and it can also be characterized by the number of records, transactions, reports, or files it contains. The term "velocity" refers to the rate at which data is generated from a variety of sources, including but not limited to business processes, networks, sensors, mobile devices, etc. The different kinds of data are what we mean by variety. The term "big data" refers to the collection of data from a variety of sources and formats, including structured, unstructured, and semi-structured data. The inconsistency and uncertainty that can be found in the dataset are referred to as its veracity. The information that has been gathered might be incomplete, might contain errors, or might not be able to provide genuine and helpful insight. For instance, in the field of medicine, if data regarding what medications a patient is taking is missing, the patient's life may be in jeopardy, and this could put other patients at risk as well. The term "value" refers to the usefulness that the gathered data can bring for the company, and it is directly related to what businesses are able to accomplish with the data that they have gathered. A meaningful form of the data that is presented in an unstructured fashion is what is meant by value. The 5Vs of Big Data, which have been defined in detail above, are, in essence, the features of big data. We can make a connection between these 5Vs and the data that is kept, handled, and also represented by hospitals and other healthcare industries.

Big Data in Healthcare:

The advancement of information technology and the capacity to store more data have prompted nations and governmental institutions to computerize health records. This has led to the production of the Electronic Health Record (EHR), which collects data in a variety of structures, including structured, unstructured, and semi-structured formats. Big data analytics in the medical field enables the examination of massive datasets including information on thousands of patients, as well as the determination of clusters and correlations between the datasets [2]. Big data analytics has the potential to play a significant part in the evolution of healthcare services at a time when the healthcare sector is placing a premium on innovations that will result in better outcomes for patients.

- Hospital performance management
- Prevent epidemics, cure disease, and reduce costs

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

- Increase transparency and efficiency in early disease diagnosis
- Enhancing clinical outcomes
- Engaging patients and family and more personalized treatment to patient

Management and analysis of big data in Healthcare:

An electronic health record (EHR) makes it possible for all of the permitted parties engaged in a patient's care, including physicians, labs, pharmacies, emergency facilities, nursing homes, state registries, and patients themselves, to share medical information with one another. HER must assure patient data protection. Users and/or entities should be authenticated, and their access permissions should be determined based on that. The vast majority of wearable devices enable the gathering of biochemical, physiological, and motion-sensing data [3, including data on things like heart rate, steps walked, and blood pressure]. In order for it to be able to collect patient health data and have the capability of exchanging data.

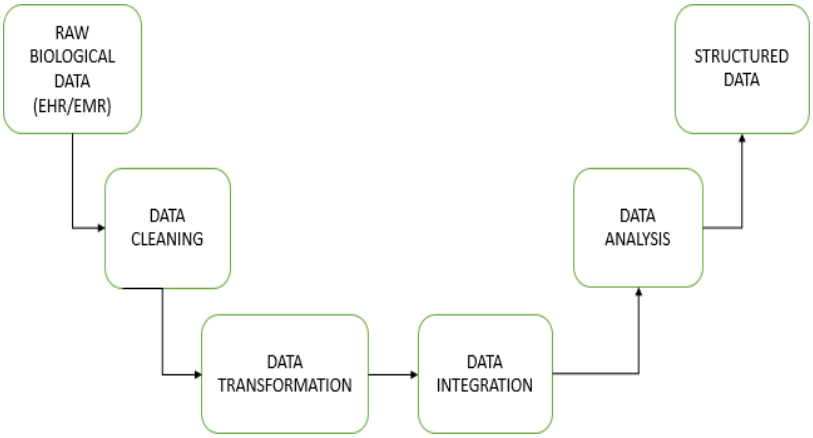

Figure-2: Data management and processing in healthcare

Big data analytics is a methodical process that examines, cleans, transforms, and models this type of data to aid in the discovery of knowledge and the improvement of decision-making. In the field of bioinformatics, the term "data acquisition" refers to the process of acquiring information from a variety of dispersed sources, including healthcare, genetics, and so on. When it comes to data collecting, a method that is well-designed is preferable. There are a great variety of methods for data collecting that are employed nowadays. After the data have been acquired, they are typically transported to large data warehouses for the purposes of storing and further processing. After the data have been recorded, it is necessary to clean it. Under these conditions, activities involving the preprocessing of data play an important role. The process of correcting or deleting inaccurate, corrupted, or duplicate data from inside a dataset is referred to as "data cleaning." Following the completion of the cleaning procedure comes the data integration step. The use of software, applications, and organizational procedures to extract data from numerous sources into information that is coherent and intelligible in order to do data analysis more quickly and effectively is what is meant by the term "big data integration." The algorithms used for data mining are responsible for managing memory, costs, and the accuracy of performance. The algorithms for data mining can be divided into three categories: descriptive, predictive, and prescriptive. Data analytics that are descriptive look for patterns and correlations that can be used to characterize the qualities of the data. Examples of descriptive techniques include clustering and the detection of sequential pattern sequences. Data mining methodologies such as classification, linear regression, and others are used in predictive data analytics. These

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

techniques are used to make predictions about the future based on the factors that are already present in the dataset. The process of making decisions can be improved with the assistance of prescriptive data analytics. Because of the interpretation and analysis of the results of the data analytics, the unstructured data may be visualized in the structured formats. This is one of the most important roles that interpretation and analysis play in comprehending the outcome of the data analytics. When performing each of the numerous types of analyses, a unique set of algorithms is utilized on the datasets in order to produce a statistical or mathematical model. The model that was thus developed is able to determine the link that exists between an independent feature and other dependent features contained in the dataset [4, 5].

Technologies used in Big Data Analytics:

There are a number of big data tools and technologies available in the current market which are essential for storing, managing, manipulating and visualizing huge amount of data in the system or computers.

1. Hadoop: Hadoop is an open-source framework. Hadoop helps in storing EHRs, EMRs and processing large data Hadoop's capability to store large unstructured data sets in NoSQL databases and using Map Reduce to analyse this data helps in the analysis and detection of patterns in the field of Healthcare. A programming language suitable for working on big data (e.g. Python, R, Java). Hadoop enables parallel processing of data and it uses clustered architecture [6].
2. Apache Spark: One of the best uses of Spark in Healthcare is the analysis of patient records along with past clinical data which included diseases, treatments, EHRs reports, (memory calculation)etc. Also, saves costs for both the hospitals and patients.
3. NoSQL: Mongo DB provides cross-platform capabilities. NoSQL is based on the consistency, availability and partial tolerance (CAP) theorem. NoSQL Database is a non-relational Data Management System that does not require a fixed schema. It avoids joins, and is easy to scale. The major purpose of using a NoSQL database is for distributed data stores with enormous data storage needs.

Literature Survey:

In the topic of big data, there have been and are still a great number of studies being conducted. The field of data science has benefited from the contributions of a great number of data scientists and data analysts. The following is a list of some of the research articles that have been published. The name of the paper, significant aspects covered in the paper, and avenues for further reading are all mentioned in the literature overview or review.

In recent years, due to the explosion of data in a variety of industries, big data has been given a tremendous amount of attention. This is because big data is an advanced form of technology. Big data plays an important function in a variety of disciplines, including healthcare, education, business, and manufacturing industries. This role includes helping to organize and store different types of data. In the field of healthcare, a wide variety of big-data analytics tools and approaches have been created in order to manage the enormous volumes of data that are involved. The article titled "Big data analytics for the healthcare industry: impact, applications, and tools" discusses these elements as well as the issues that they present. This article discusses the influence that big data is having in the healthcare industry, as well as the numerous tools that are available in the Hadoop ecosystem for managing it. There is also an explanation of the theoretical architecture of big data analytics for healthcare, which includes the data gathering history of various departments, the genomic database, electronic health records (EHR), electronic medical records (EMR), and clinical records and support systems [7].

The discipline of big data in the healthcare sector analytics is currently the subject of a significant number

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

of ongoing studies all over the world. The current study, titled "Big Data Analytics in Healthcare: A Systematic Literature Review," is one that conducts a review of the previous research in the field of big data analytics in the healthcare industry. In this publication, the results of 41 separate investigations are analyzed, and they are presented within an all-encompassing framework. These findings from this study imply that BDA in healthcare can be observed from healthcare awareness, hospital management systems, therapies, and healthcare technology services. These are the findings. Computing in the cloud, block chains, and machine learning are all examples of applications of new technology in the healthcare industry that give a promising new direction for research. Because of the usage of these technologies, medical professionals and other providers of healthcare as well as trained individuals in the field of big data analysis are now able to manage the data in an effective manner and achieve positive results. Therefore, these many kinds of SLR will be advantageous for the next generation in every possible way [8].

According to the findings of previous research, the implementation of Big Data Analytics in medical facilities can result in a wide variety of positive outcomes. The objective of the research project titled "The use of Big Data Analytics in healthcare" is to investigate the potential applications of big data analytics in the medical field. The research is predicated on an in-depth analysis of the existing body of literature, in addition to the presentation of a selection of the results obtained through actual research on the application of Big Data Analytics in clinical settings. This paper's primary contribution is to offer an analytical overview of employing structured and unstructured data (Big Data) analytics in medical institutions in Poland. More specifically, the overview focuses on Poland. In their daily operations, medical facilities often make use of both organized and unstructured data. The schema for structured data is already set, and the data itself is both comprehensive and freeform. On the other hand, unstructured data, often known as Big Data, does not conform to the standard format for data processing. Therefore, in order to process this kind of medical data, preliminary processing needs to be done in order to provide ever superior visualization. This paper also addressed and explained the difficulties that the healthcare business has as well as potential solutions for those difficulties. As is well known, the healthcare industry is a complicated system with many different parties involved, including patients, medical professionals, healthcare facilities, pharmaceutical corporations, and healthcare decision-makers. It is very important to use the tools, techniques, and platforms as well as for medical data processing in order to overcome the obstacles and issues that include a lack of adequate understanding of a vast volume of data, sharing the data, and updating the data [9]

Applications:

Some of the applications of big data in healthcare are as follows:

1. Predicts patients at higher risk quickly efficiently: While considering the population wide data for a particular area, specifically predictive analytics pinpoints which segment of the patients are at higher risk for diseases and hints for early intervention to protect them.
2. Eases patient diagnostics with EHRs: Enabling effective patient diagnostics with every patient having their own EHRs. This EHRs includes the demographics, medical history, and diagnostic test results of current and previous illnesses along with other details.
3. Health data from sensors: Wearable sensors, just as the name implies, are integrated into wearable objects or directly with the body in order to help, monitor provide clinically relevant data for care. For example Electromyography (EMG), Electrocardiogram (ECG).
4. Cancer prediction and cure: Big data can help to fight cancer more effectively. Healthcare providers will have enhanced ability to detect and diagnose diseases in their early stages and regulate drug doses to minimize side effects and improve effectiveness.

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

Challenges:

In healthcare, real-time big data analytics is important. Some literatures discuss challenges in the development of big data in healthcare applications. The key challenges are as follows:

1. Individuals and businesses alike that save data on people, products, activities, and the like have a significant interest in protecting their personal information and preventing unauthorized access to that information. There is a possibility that private and confidential information is contained in the healthcare data received by healthcare professionals from individuals and their medical records. Access control, authentication, and validation of the data source are some of the security methods that will ensure the confidentiality of medical records [10]. Strong encryption of the data is another one of these measures.
2. The most obvious problem that comes along with big data is storing and processing the enormous amount of data that is involved. Nowadays, anytime a new storage technology is established, there is a considerable increase in the amount of data that is created. This is because healthcare providers gather and transfer a massive amount of data on a regular basis, and there are also many different types of data that are generated periodically. This assists healthcare companies and hospitals in reducing costs and resolving problems related to data storage.
3. The question of who owns the data offers a significant challenge that persists in the context of big data applications in the healthcare industry and other fields. Even though healthcare providers, public healthcare systems, or hospitals are usually the owners of medical records, the information contained inside those records does not belong to any of those entities. Unless there is a legal structure in place, the problems with ownership should be resolved.
4. Required Abilities: A data analyst is a specialist whose job entails gathering, cleansing, visualizing, transforming, and modeling raw data into the informational building blocks that are utilized by marketers, developers, and healthcare providers. Therefore, individuals need to possess certain abilities in order to operate in the big data industry.

Conclusion:

The use of sophisticated technologies by healthcare providers to acquire insights from their clinical and other data repositories and to make informed decisions could be fundamentally altered by the application of big data analytics. Every day, enormous amounts of data are generated at extremely rapid speeds in the information sector. The ability of big data analytics to facilitate quicker decision-making is a significant factor contributing to the industry's overall growth. Big data is a term that refers to datasets that are not only large in size but also high in diversity and velocity. This makes it challenging for users to manage data by employing conventional tools and methods. Big data technology is being used to store data by a variety of industries, including healthcare management systems, manufacturing industries, and banking systems. This technology is advantageous to all of these sectors. As a result, only a small number of relevant literatures were examined so that an overview of big data analytics in healthcare principles could be provided. due to the fact that it contains all of the characteristics of large data. The techniques of big data analytics are applied to this data in order to get insights that are extremely valuable.

Future scope:

In the not too distant future, we will witness the rapid and broad use of big data analytics across the entirety of the healthcare business as well as within individual healthcare organizations. In order to achieve this goal, a number of issues, including data ownership, data privacy and security, improvements to tools and technology, and talent needs, need to be tackled. Big data analytics is a very encouraging method that integrates, examines, and analyzes vast volumes of complicated heterogeneous data of different sorts, including biomedical information,

ISSN:0975 -3583,0976-2833 VOL12, ISSUE 02, 2021

medical data, electronic medical record data (EHRs), and data from experiments. This process is carried out in the fields of medicine and healthcare. A new high-performance data management system will be developed in the not too distant future, and it will be based on an open source platform such as Apache Hadoop. This system will be able to help a range of datasets, and it will employ memory and other hardware resources in a more efficient method to acquire insight.

References:

[1] Laney D. 3D data management: controlling data volume, velocity, and variety, Application delivery strategies. Stamford: META Group Inc; 2001.

[2] Reisman M. EHRs: the challenge of making electronic data usable and interoperable. Pharm Ther. 2017;42(9):572–5.

[3] 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.

[4] Lyko, K. (2016) 'Big Data Acquisition', New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe pp.39–63

[5] Katherine, G., Herbert and Jason, TL Wang. (2007) 'Biological data cleaning: a case study', Int. J. Information Quality, Vol. 1, No. 1, pp.60–82

[6] Hess, K. (2016) New Hadoop Survey Makes Big Data Predictions for 2016'. Available at: http://www.zdnet.com/article/new-hadoop-survey-makes-big-data-predictions-for-2016.

[7] S. Kumar and M. Singh, "Big data analytics for healthcare industry: Impact, applications, and tools," Big Data Min. Anal., vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.

[8] Sayantan Khanra , Amandeep Dhir , Najmul Islam & Matti Mäntymäki (2020) Big data analytics in healthcare: a systematic literature review, Enterprise Information Systems, 14:7, 878-912, DOI: 10.1080/17517575.2020.1812005

[9] Kornelia Batko, Andrzej Ślęzak (2022) "The use of Big Data Analytics in healthcare," *Journal of Big Data* volume 9, Article number: 3

[10] P. Galetsi, K. Katsaliaki, and S. Kumar, "Values, challenges and future directions of big data analytics in healthcare: A systematic review," Soc. Sci. Med., vol. 241, p. 112533, 2019, doi: 10.1016/j.socscimed.2019.112533

Websites/Sources:

1) https://ieeexplore.ieee.org/
2) www.researchgate.net
3) https://journalofbigdata.springeropen.com