

Big Data Analytics in Healthcare: A Review

Prableen Kaur
Department of Mathematics
Chandigarh University
Punjab, India

Abstract—In recent years, huge amounts of structured, unstructured, and semi-structured data have been generated by various sectors around the world and, neither of this data is homogeneous. This enormous amount of data, referred to as ‘big data’, has started to play a pivotal role in the evolution of healthcare practices and research. In this paper, we discuss how by rapid digitalization along with other factors, the health industry has been confronted with the need to handle the big data being produced rapidly at an exponential speed. Big data analytics tools play an essential role to analyze and integrate large volumes of data, which otherwise might have become useless or taken more time to give value. The usage and challenges of big data in healthcare is also addressed.

Keywords—*Big Data, Big Data Analytics, Hadoop, Healthcare*

I. INTRODUCTION

Data is a powerful source which is being produced at an ever-increasing rate every day. Mobile phones, social media, search engine data, healthcare data and many more helps in creating this new data that must be stored somewhere to get as much value as possible from it. Our traditional database systems are unable to handle this enormous amount of variety of data coming at an uncontrollable rate. This has led to the creation of a term called, ‘Big Data’, which refers to the voluminous datasets produced that is growing exponentially with time, and it does not conform itself to the structure of traditional databases. Analysis of this data uses newly researched technologies and distributed architecture that makes extraction of value from the dataset possible. For example, the New York Stock Exchange generates about one terabyte of new trade data per day.

One of the notable areas, where big data analytics is making changes is, healthcare. Big data is being generated in healthcare, with respect to the collection, analysis, and leverage of consumer, patient, physiological, and medical data that is too large or complex to be understood by conventional methods of data processing. Big data in healthcare grew as a result of the digitization of healthcare records and the rise of value-based medicine. In order to address healthcare information issues such as size, speed, uncertainty, and truthfulness, health systems must incorporate technology that can collect, process, and interpret healthcare data. Big data carries the promise of enabling a broad variety of medical and healthcare functions, including clinical decision support, disease monitoring, and population health management, owing to mandated standards and the ability to increase the quality of healthcare delivery while lowering costs [1].

II. BACKGROUND

A. Big Data

"Big data" is not a new term, but its definition is constantly evolving. It's a set of data whose size, delivery, diversity, and/or timeliness necessitate the use of new technological frameworks, analytics, and resources to allow new sources of market value. The volume, velocity, variety, veracity, and value of complex data classify it. The size and enormity of the data are determined by its volume. The rate at which data changes, or how frequently it is created, is referred to as its velocity. Variety covers the various formats and sources of data, as well as the various uses and methods of processing the data. The consistency of the data is the subject of veracity. Big data output is classified as excellent, mediocre, or undefined as a result of data inconsistency, incompleteness, uncertainty, delay, deception, and approximations. Finally, value is derived from gaining any valuable knowledge or making sense of the results. Making sense of this sea of data is the challenge of this age. This is where big data analytics can help [2].

B. Big Data Analytics

With the advancement in technology and the growing amount of data coming in and out of organisations on a daily basis, there is a need for quicker and more efficient data analysis. Having a large amount of data on hand is no longer sufficient for making timely and effective decisions. Only if we can better interpret and use the data generated in various libraries by numerous institutions, as well as the data generated by individuals, can we make a change. In other words, without adequate analytics, data would be nothing more than a resource that isn't used. Furthermore, the word "big data" refers not only to the amount of data but also to the data's capacity. The data sets are vast and diverse, making it difficult to interpret and collect the results using current techniques. Here's where big data analytics comes in handy. Big Data Analytics entails gathering data from various sources, combining it in such a way that it can be analyzed by researchers, and then providing data items that are beneficial to the organization's market. Big Data Analytics is the method of transforming vast volumes of unstructured raw data gathered from various sources into data that is useful [2, 3].

C. Big Data Analytics in Healthcare

Medical data is currently being generated from a variety of sources, including cell phones, body area monitors, patients, hospitals, researchers, healthcare professionals, and organisations. Big data in healthcare refers to vast amounts of

data generated by the use of digital technology that capture medical records and aid in the management of hospital results, which would otherwise be too broad and complicated for conventional technologies. The use of Big Data analytics in healthcare has shown a slew of promising results, many of which are life-saving. Electronic Health Records, computer generated/sensor data, health information exchanges, patient registries, portals, genomic databases, and public records are all examples of data types used in healthcare applications. Public reports are important data points in the healthcare system that need effective data analytics to address their medical challenges. This major health data is specially processed for next stage review on medical servers (MS), clinical databank (CDB) and CDRs. Storage infrastructures are mainly used to store, process, interpret, handle, and recover massive volumes of data in order to facilitate people's lives. As a result, it not only provides information to help people understand symptoms, illnesses, and medications, but also to warn them, forecast results early on, and make the best choices possible. Big Data Analytics is a modern method for analysing, managing, and accurately extracting valuable information from vast quantities of data sets that are very close to a specific patient in a brief period of time. Furthermore, this new technology-based system of analysis transforms treatment to the right patient at the right time [4 , 5].

III. LITERATURE REVIEW

- Ashwin Belle along with few others have published their paper, titled "Big Data Analytics in Healthcare" in Biomed International Journal. In this paper, they have discussed how big data is a set of data elements whose size, speed, type, and complexity necessitate the search, adoption, as well as invention of new hardware and software mechanisms in order to effectively process, interpret, and represent the data. Their focused areas of interest were medical image analysis, physiological signal processing, and genomic data processing. An overview of analytical methods that are used in medical image analysis to improve the interpretability of depicted contents, different challenges and existing approaches in the development of monitoring systems that consume both high fidelity waveform data and discrete data from non-continuous sources as well as wide variety of topics that cover big data applications in genomics [6].
- Revanth Sonnati in the paper "Improving Healthcare Using Big Data Analytics" discussed Hadoop data processing as one of the best choices to go with at the current trends, and how it'll provide an extra edge to analyze the data. The aim of this paper was to provide a viable computer approach using big data and analysis to enhance healthcare through the promotion of healthcare science, affordability and accessibility. It focuses to benefit the society with advanced computation techniques to analyze and provide patient-centric health care. It clearly gives the picture of the data flow starting from the raw data along with its types through the Hadoop ecosystem and the analytic engines to achieve the final goal of the system. The paper affirms that given geographic location is also important to analyse healthcare data [7].
- "Big Data Analytics in Healthcare" by M. Ambigavathi and D. Sridharan, has summarised about the evolution of V's and characteristics of big data in healthcare applications. It discusses the key roles of various components involved in the big health data analytics from data mining to the knowledge discovery process. The various big data analytical tools for data analysis and their functionalities are summarized in this paper. Some of the open research challenges and feasible solutions are highlighted in order to reduce the healthcare cost, enhance treatment, and improve the quality of patient care. With the help of analytics tools, data scientists are able to integrate health related information from both internal and external sources [5].
- The paper "Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools" authored by Sunil Kumar and Maninder Singh, discusses the various sources and forms of Big Data which challenge the information technology industry to improve data processing methods. Techniques that merge different sources of data are highly requested. There are a host of conceptual ways to detect anomalies in large quantities of data from various datasets. It has a brief overview about improving outcomes for patients by following a "pathway" and how it will directly impact the patient. It discusses how various applications provided by the Hadoop ecosystem can help the healthcare domain, which involves the utilization of the big data, generated by different levels of medical data and the development of methods for analysing this data and to obtain answers to medical questions [4].
- Manpreet Singh and others in their paper, "BIG DATA ANALYTICS Solution to Healthcare", discusses the usefulness of application of Big Data Analysis using the patient care dataset for better insight in care coordination, health management and patient engagement. It shows their studies in healthcare and how big data provides solutions and huge application to biomedical problems. It concludes that Big Data will bring greater changes to the healthcare system than estimated and be a boon. The impact on health science seems to be amplified by predictive analysis as the science reports can go viral, assist and predict an enormous amount of emergency clinical cases [8].

IV. USAGE OF BIG DATA IN HEALTHCARE

By effectively using big data, the applications of big data analytics can improve the efficiency and quality of healthcare delivery, detect diseases at an early stage, provide evidence-based treatment options to patients, and monitor the quality of the medical and healthcare institutions as well as provide better treatment methods. Big data has use cases in almost all the industry and it has very crucial use cases in the healthcare domain too [9].

Hadoop data processing is one of the best choices to go with the current trends. The computer capability of Hadoop will stimulate existing statistical techniques and approaches to medical science to maximise the efficiency of outcomes [7]. Hadoop based applications are used in Healthcare to provide treatment in Cancer and Genomics, to monitor Patient Vitals, for Hospital Network, for Prevention and Detection of Frauds and much more [4].

A few viable Big Data and Analytics computer solutions are aimed at encouraging testing, affordability and usability in the area of medical care. It offers measurable advantages to transform the area of healthcare by offering ground-breaking computational technologies for analysing and providing patient-oriented healthcare for the good of the population. These proposed objectives are Clinical Decision Support, Disease Management, Patient Matching and Lifestyle Analytics [7].

V. CHALLENGES

Big data analytics is only in the early stages of growth, and existing technologies and approaches are incapable of addressing the issues associated with big data. Big data may be regarded as large, challenging structures. Much analysis would also be necessary in this area to resolve the healthcare system's problems.

It is recognised that a number of Big Data uses are needed to change our operations in health care, to increase care delivery and to deliver customised care to patients. At least, the main tasks for data collection must be supported by the big data analysis network of healthcare. Platform assessment requirements could include availability, continuity, ease of use, scalability, the ability to exploit at various levels of granularity, privacy and security enablement, and quality assurance. Furthermore, although the majority of platforms currently available are open source, the usual benefits and shortcomings of open-source platforms apply. To be effective, big data analytics in healthcare must be packaged in a menu-driven, user-friendly, and straightforward manner. In healthcare, real-time big data analytics is important. It is necessary to resolve the time gap between data collection and retrieval. However, big data presents a number of obstacles that, if overcome, will make life simpler. It is simple to use the term eradicate, but it is not so simple to put it into action. The list below highlights a number of the major data issues that we face today [10].

A. *Data management, security and privacy issues*

Issues such as data integrity and privacy lead to poor data management. Privacy violation and discrimination. Disclosure of Personal Health Information is also a major risk.

B. *Technological issues*

Lack of required infrastructure cannot produce safe conclusions. Social inequality, as data are only open to a small elite of technical specialists who know how to interpret and use it, and to those who can employ them.

C. *Skilled Resource set*

There is a need to have a Data scientist and Data analyst to perform big data analysis. There is already a huge shortage in the required skill set for Big Data Analytics.

D. *Data Ownership*

There is a lot of big data flowing which includes genomics, remote sensing, social media, mobile app and many other data types.

E. *Healthcare Models*

There is a need to have sufficient business case evidence in health to measure investment return.

F. *Limited awareness and support*

It'll cause lack of funding and awareness. Dependency on private funding will support a few big players that will further lead to international economic competitiveness. Funding models have to be revisited to ensure better care.

VI. CONCLUSION

We discussed the ground breaking subject of big data analytics in healthcare in this paper, which has recently gained a lot of attention due to its potential benefits. In the information age we now live in, massive amounts of high-speed data are produced on a daily basis, and among them are intrinsic facts and patterns of secret intelligence that should be extracted and used. Big data analytics is a huge boost to the healthcare industry because it allows for faster decision-making.

Accordingly, few literatures were reviewed in order to include an overview of big data analytics in healthcare principles. We have already highlighted the use of Big Data Analytics in the Healthcare sector, as well as the issues encountered. Healthcare, among most sectors, generates a large volume of data by either manual processes or automated data capturing devices. Since it possesses all of the features of big data, even a large volume of data is called big data. Big data analytics approaches are applied to this data in order to gain very valuable insight into the data.

REFERENCES

- [1] Javier Nieto León, "Big Data In Healthcare 2021." <https://www.htechtrends.com/big-data-in-healthcare/#:~:text=Big data in healthcare refers to the collection%2C,processed by machine learning techniques and data analysts.>

-
- [2] D. Chong and H. Shi, "Big data analytics: a literature review," *J. Manag. Anal.*, vol. 2, no. 3, pp. 175–201, 2015, doi: 10.1080/23270012.2015.1082449.
- [3] Tutorialspoint, "Big Data Analytics Tutorial." https://www.tutorialspoint.com/big_data_analytics/index.htm.
- [4] 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.
- [5] M. Ambigavathi, "Big Data Analytics in Healthcare," 2018 Tenth Int. Conf. Adv. Comput., pp. 269–276, 2018.
- [6] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big Data Analytics in Healthcare," *Biomed Res. Int.*, vol. 2015, pp. 1–16, Nov. 2015, doi: 10.1155/2015/370194.
- [7] R. Sonnati, "Improving Healthcare Using Big Data Analytics," *Improv. Healthc. Using Big Data Anal.*, vol. 6, no. 3, pp. 142–146, 2015.
- [8] M. Singh, N. Delhi, V. Bhatia, R. Bhatia, and D. Specialist, "Big data analytics," pp. 239–241, 2017.
- [9] J. N. Undavia and A. M. Patel, "Big Data Analytics in Healthcare," *Int. J. Big Data Anal. Healthc.*, vol. 5, no. 1, pp. 19–27, 2020, doi: 10.4018/ijbdah.2020010102.
- [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.