IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 10, Iss 08, 2021

CLOUD-BASED DATA ANALYSIS TOOLS IN THE HEALTHCARE FIELD: AN IN-DEPTH EXPLORATION

[#1]**Mr.JANGA RAVI CHANDER,** *Assistant Professor*
[#2]**Mr.PEDDI KISHOR,** *Assistant Professor*
Department of Computer Science and Engineering,
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

Abstract: The healthcare business creates a vast amount of data that is critical to improving patient outcomes and scientific understanding. Traditional on-premises data storage and processing options may struggle to keep up with healthcare data's rising volume, variety, and velocity. The purpose of this study is to look at the benefits and drawbacks of cloud-based technologies for data analysis in the healthcare industry. Inspiring new ideas, improving patient care, and providing more precise medical support are just a few of the goals of the latest AI and cloud computing advances highlighted in this article. This article focuses on the application of Hadoop and Cloud Computing in healthcare for big data analytics, namely for monitoring, forecasting, performance monitoring, and management, as well as in other settings such as the intensive care unit. Many applications, tactics, and future directions for big data analytics are covered. Several cloud platforms, like MMAP, are providing timely, dependable, cost-effective, efficient, and patient-centered solutions to these types of public health challenges. On a national or regional scale, these algorithms may also forecast the health repercussions of specific diseases. AI, Hadoop, and cloud computing infrastructure make it easier to share information. This enables healthcare management to undertake the computations required to find the reasonable, important, and factual trends that influence both disaster preparedness and improvement programs.

Keywords:Big Data; Healthcare; Cloud Computing; Data Analytics; Hadoop.

1. INTRODUCTION

Users can process and analyze data volumes that are too vast for traditional processing and analysis approaches thanks to "big data" technology and tools. Many healthcare-related enterprises offer a diverse range of complex datasets (HCD). Numerous sources, as illustrated in Figure 1, contribute to the amount of knowledge we have on patient healthcare.

The government mainly relies on health statistics supplied by the general populace. Federal and state agencies monitor vaccination rates, COVID-19 results, and the occurrence of insect-borne diseases such as dengue fever and malaria. They also collect health information for regulatory and legal objectives, such as the construction of prescription drug databases. Researchers and students have access to health data sets controlled by the government. As a result of bio-inspired and scientific discoveries, an abundance of public health-care data has been generated, and this data must be categorized, assessed, and projected.

Both public and private research institutions accumulate large amounts of data that can be used to better understand disease spread.

Generic health databases are centralized clinical data warehouses that aggregate data from a wide range of healthcare-related information technology platforms. The primary focus of these databases is clinical and therapeutic work undertaken for patients by various healthcare practitioners. These databases are easy to use, but they provide a plethora of information for studying patient healthcare data using various visualization techniques and statistical models. A patient's medical history in a hospital or clinic is now stored in a computerized format known as an electronic health record (EHR). RTMRs use

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 10, Iss 08, 2021

identity access management (IAM) to ensure that patient information is accessed swiftly and securely in real time.

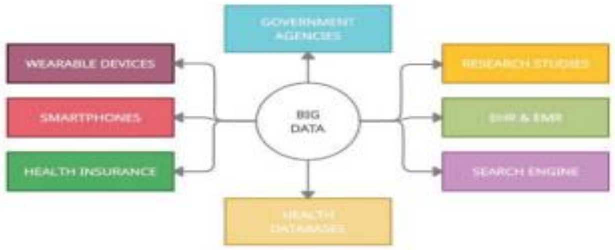

Fig. 1. There are numerous health-related information sources.

Daily medical transactions result in electronic health records. These files contain radiology pictures, drug histories, diagnoses, treatment plans, and dates of vaccines and vaccinations. Online searches are the primary source of information about people's health and well-being.

The healthcare data lake can hold data from healthcare-related social media platforms. This data can help with analytics and predictions. The majority of persons seeking medical assistance do so because they are experiencing similar symptoms; this fact can be used to assess the prevalence of a particular medicine or therapy in a given region. As if it were an epidemic. Sensors are used in wearable devices such as smart watches and exercise bracelets to collect data. These devices contain sensors that can collect vital health data such as pulse, blood pressure, temperature, and SPO2 concentration. These pieces of data can be fed into an intelligent health informatics (IHI) system to learn more about the health of patients.

Health-related apps (HAD) on smartphones store information regarding lab visits and basic diagnoses. The information in these programs on the general population's eating, sleeping, and exercise habits is useful, but it can also make a significant difference in a patient's journey to health. It is possible to predict health problems that could be life-threatening using this data plus information from real-time diagnosis tools. Similarly, health insurance information from companies that provide in-patient health insurance benefits to people contains a wealth of information about the procedures performed on the patient, the medications administered, and the precautions and medications to be taken following surgery. This information can be extremely useful for learning about a patient's entire medical history, assisting with current care, and providing a reputable database for research studies and clinical trials.

There are three main reasons why these input data from healthcare source systems are referred to as "big data." Big data is defined by its quantity, variety, and speed. It is common for the hospital network system to generate large amounts of data. This material comes from a variety of sources and covers a large area. The volume of healthcare data is increasing as more individuals utilize gadgets such as smartphones, fitness bands, smart watches, AI assistants, and smart home sensors. This implies that we must better and more effective methods of storing and analyzing this data.

The input source systems generate many types of data. Different technical systems are responsible for the various forms and characteristics of data. Medical data can now be stored online in places other than spreadsheets and databases. It can be stored in digital images, sensor data, social media messages, and other locations. Because of the volume and variety of this unstructured data, it is difficult to extract, manage, and study. It has been demonstrated that machine learning can effectively tackle this problem. Public healthcare data is classified into three types: semi-structured (like X-rays), unstructured (like interaction and conversation data), and organized (like medical records and pathology reports). Smart devices are gaining popularity, and more and more services are being digitized. This generates intriguing new data streams that may be collected and evaluated.

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper **Journal Volume 10, Iss 08, 2021**

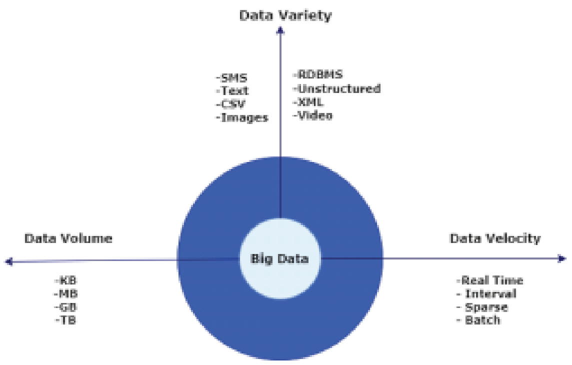

Fig. 2. Data vs. the Big Three

Big data is accumulating at an alarming rate, creating its own set of requirements and issues. The third V in big data is speed. It refers to the amount of data generated, processed, and then made available for display and analysis. Certain hardware is required to handle the astronomical rate at which data is received and created in real time. Big data technology has advanced over the years, making it easier to manage the massive amounts of data that arrive at breakneck speed. The quantity of data created in the last few years has been discovered to be more than the amount created from the beginning of time. In the new digital era, more technology and devices will be employed, which will accelerate the flow of data into a data analytics platform. Figure 2 depicts all three categories of large data that we discussed earlier.

2. LITERATURE REVIEW

Even while some issues remain, big data analytics, which allows healthcare organizations to collect massive volumes of data, is the major item that saves money and improves results. Many doctors, clinics, hospitals, and other health care personnel use cloud computing to better diagnose, treat, and care for their patients. However, other people believe that cloud computing has its own set of boundaries and thresholds that could make the transition less easy.

Hadoop and MapReduce frameworks are utilized for next-generation sequencing research. They are becoming increasingly popular in this industry. Because big data is changing at such a rapid pace, it is critical to devise innovative methods of managing it. Hadoop, No SQL, and Map Reduce are some of the most significant tools for handling large amounts of data. These technologies are quite useful for organizing massive amounts of data. Many fields generate a large amount of data, and experts have investigated these fields and methods for dealing with this data. Cloud software is becoming increasingly used in the medical profession. It has various advantages that customers can take use of, such as adaptability, pay-per-use, and a large access network.

Better big data analytics are now possible in the healthcare sector as a result of the massive increase in medical data from numerous domains, such as physiological signal processing, genetic data processing, and medical picture analysis. This study is divided into two parts: a Hadoop map reduce strategy for large data analytics and a data reduction method that employs a variety of technologies such as Sqoop and Hive. The purpose of this research is to look at new breakthroughs in computational image analysis approaches for prognostic modeling of digital pathology images from the perspectives of discovering them, extracting features, segmenting them, and classifying tissues.

You must be able to analyze data in order to work in healthcare analytics. It's fantastic that Apache Hadoop can process data so quickly. Big data is information that is stored online and processed using the Hadoop map-reduce technology. This report explains how big data and Hadoop are used in healthcare research. The author recommends a Hadoop-based design for healthcare big data analytics. This study investigates the use of big data analytics in intensive care units. The study's author discussed a variety of big data analytics applications, algorithms, and potential future advancements in the care of people in intensive care units. The author discusses the various applications of big data in healthcare, such as management, monitoring, prediction, and performance.

The article covers the architecture of the Hadoop ecosystem and identifies the required tools and

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper **Journal Volume 10, Iss 08, 2021**

components. The issues that arise as a result of big data analytics are also discussed. Machine learning and artificial intelligence are frequently applied in the medical field. They had developed cardiovascular magnetic resonance (CVR) applications. The article discusses how machine learning, particularly deep learning algorithms, can be applied in cardiovascular magnetic resonance. This article discusses the primary applications of machine learning in healthcare, with a focus on translational, public health, and applications. It also discusses the significance of genetic data, data interchange, and privacy. Machine learning tools are utilized to assist clinicians in making infectious illness judgments. Machine learning advances make it easier to investigate infectious diseases in more depth, such as sepsis, HIV, and bacterial resistance.

Machine learning approaches may be valuable for healthcare researchers who seek to improve their prediction of healthcare anomalies because there are so many datasets to choose from and train an approximate algorithm on. When there are many covariates to consider and the outcome of interest is affected by complex nonlinear interactions, machine learning can produce generalizable data-driven estimators. The cloud computing framework facilitates data sharing, allowing healthcare management to perform analytical computations to identify logical, relevant, and genuine patterns that are required for disaster planning and preparation. Furthermore, the framework's conclusions can assist decision-makers in determining and guessing what disaster-related health impacts will occur in any area of any country by looking at that area's geography and previous disasters.

The use of AI/ML and big data analytics in healthcare is one of the most significant advancements for the intelligent healthcare system. The goal of this research is to completely evaluate the m-healthcare system using artificial intelligence and big data analytics. Data mining is critical in this age of huge data. Medical notes and follow-up data can be better stored and retrieved using information technology and data mining. Examine medical data for any probable patterns or correlations to have a better understanding of how to diagnose and treat patients. It can also help anticipate diseases, detect them early, and heal more individuals.

By bringing in massive amounts of data and converting them into a series of events that improve treatment, the effects of big data on the oral health care company will be sufficient to fulfill the demand for precision in determining the optimal care at every follow-up. Some of these include identifying therapy gaps and making patients happier, safer, and better cared for. The oral health sector will advance significantly as a result of this research's new concepts, improved patient care, targeted medical assistance, and other ways. To obtain medical data, massive amounts of data must be organized and made available in a cloud computing environment. This work provides a novel strategy to collect massive amounts of medical data based on a modified immunological approach. Before creating a massive data structure model in cloud computing, its attributes must be investigated; otherwise, an information model cannot be developed. The purpose of designing an immune optimization strategy for clustering is to optimize clustering for very large data sets.

Table 1 Healthcare analytics is used in a variety of fields.

Sr. No.	Title	Year	Related Field	Application
1	W. Raghupathi et al. [9]	2014	Big Data Analytics	Healthcare
2	G. N Reddy et al. [10]	2014	Cloud Computing	Healthcare
3	E. A. Mohammed et al. [11]	2014	Hadoop	Clinical Big Data & Medical Health Informatics
4	J Singh et al. [12]	2015	Hadoop	Big Data Management
5	L Griebel et al. [13]	2015	Cloud Computing	Healthcare
6	Belle et al. [14]	2015	Big Data Analytics	Healthcare
7	Bante et al. [15]	2017	Big Data Analytics	Hadoop Mapreduce and Mitigation process
8	Madhubushi et al. [16]	2017	Image Analysis & Machine Learning	Digital Pathology
9	Merla et al. [17]	2017	Hadoop	Cloud Platform AWS
10	S Kumar et al. [18]	2018	Big Data and Hadoop	Big Data Analytics
11	L. NelsonSanchez-Pinto et al. [19]	2018	Big Data	Intensive Care Unit
12	Bahri, Safa, et al. [20]	2018	Big Data	Healthcare applications
13	Saroha et al. [21]	2019	Big Data	Hadoop Ecosystem
14	Leiner, Tim, et al. [22]	2019	Machine Learning	Cardiovascular Magnetic Resonance
15	Nariman, et al. [23]	2019	Machine Learning	Medical Informatics
16	Peiffer-Smadja et al.[24]	2019	Machine Learning	Clinical Decision support for Infectious Disease
17	Doupe et al. [25]	2019	Machine Learning	Healthcare Informatics
18	Madanian et al. [26]	2019	Cloud Computing & Big Data	Healthcare in Disaster
19	Alotaibi, Sultan Refa et al. [27]	2020	Big data Analytics & Artificial Intelligence	Mobile Health
20	Yang, Jin, et al. [28]	2020	Big Data	Medical Public Database
21	Finkelstein, Joseph, et al. [29]	2020	Big Data	Oral Health
22	Yu, Jing et al. [30]	2020	Big Data & Cloud Computing	Medical Data Clustering
23	Wenke et al. [31]	2021	Big Data Mining	Medical Text Data Mining
24	Li, Wei, et al[32]	2021	Machine Learning and Big Data Analytics	Smart Healthcare System
25	Torre-Bastida, Ana I., et al. [33]	2021	Big Data	Data Fusion
26	Huang, Wei, et al. [34]	2021	Cloud Computing	Prediction of phenotypes from genotypes (MMAP)

Text data is extracted for healthcare purposes utilizing a variety of data analysis methodologies. Medical text data can be handled using KNN, RNN, SVM, and other approaches. To examine medical text data, many techniques such as categorization, grouping, association rules, and others can be employed. Intelligent health care systems require big data insights to function. Machine learning and the Internet of Things will be crucial in future healthcare applications. This study investigates how machine learning algorithms are used to analyze enormous volumes of data in the healthcare area. The advantages and disadvantages of present methodologies, as well as potential issues with future research, are also discussed. This research will assist government agencies and healthcare personnel in staying current on the newest breakthroughs in AI and machine learning-based healthcare analytics.

It is difficult to process large amounts of information. Calculating with massive data takes a lot of effort and time. This essay is intended to serve as a starting point for healthcare personnel and anyone new to the area interested in learning more about bio-inspired technologies. Bio-inspired computing solutions are based on behaviors that have been utilized to tackle difficult modeling, simulation, and optimization problems for thousands of years. There are compelling reasons to believe that bio-inspired solutions should take advantage of the magnitude, ambiguity, and variety of challenges that technology is presently addressing. There are numerous cloud platforms being developed right now for specialized healthcare applications such as MMAP. MMAP is a gadget that uses genetics to predict phenotypes. This tool is highly accurate and features an easy-to-use UI. Table 1 summarizes the research conducted on multiple articles concerning the various applications of healthcare data analytics.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 10, Iss 08, 2021

3. METHODOLOGY

If utilized effectively, healthcare analytics may save costs, enhance coordination and results, do more with less resources, and make care more patient-centered. Creating healthcare analytics may assist businesses in using big data to enhance patient outcomes, figure out how to treat patients and illnesses in the future, and reduce the time it takes to receive value.

Analytics is the application of computer science to examine data and make conclusions from it. It is used to locate, comprehend, and link samples of critical facts. Information designs must also be used to look at numbers and data in a dynamic, methodical, and computationally engaging way. It works best in scenarios with a large amount of recorder data. Analytics is built on the simultaneous use of statistics, programming, and operations research to quantify success. Analytics is increasingly being used by healthcare organizations to extract new insights from data, decipher them, and put them into action. Big data and advanced analytics can be utilized to improve clinical and study results, which can then be used to solve problems in healthcare and improve operations as needed.

Fig. 3. Healthcare Analytics Models of Various Types

Predictive analytics can assist in lowering community expenses by classifying patients as "higher risk patients" and initiating early intervention before a problem occurs. Using descriptive analytics on healthcare data, we can learn about prior trends. Prescriptive analytics, on the other hand, employs particular algorithms to determine what must occur in order for a patient's treatment plan to be successful. This is an excellent reason to incorporate data from other sources, such as socioeconomic, demographic, medical, and disease information that occurs concurrently. Figure 3 depicts the many types of healthcare analytics models in use.

In healthcare, data analytics can be applied in a variety of ways. Figure 4 depicts the various types of data, the analytical models that may be applied to them, data mining methods, and how they are applied in the healthcare industry. The most prevalent forms of data sets utilized in healthcare analytics are web and social media data, data from devices and biometric equipment, transactional data, and data from everyday human behavior. In healthcare analytics, anomaly detection, regression, classification, clustering, and other data mining approaches are applied. Healthcare statistics are used to assist clinicians in making decisions, to administer healthcare systems, to monitor both public and private health, and to ensure the safety of medications.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 10, Iss 08, 2021

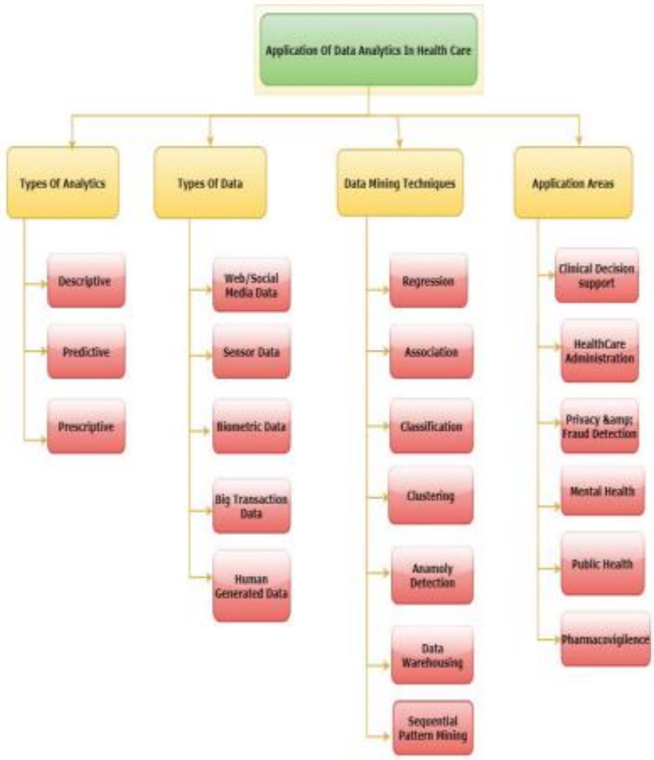

Fig. 4. A healthcare analytics diagram

4. CHALLENGES

One of the primary issues with cloud-based big data apps for healthcare analytics is that they generate a large amount of data. The data sources could be incomplete, incorrect, or influenced by outside causes. Before considering any healthcare analytics tool that leverages patient health data, make certain that patients or test participants have given their permission. It is also critical to combine data from internal and external systems in order to create effective analytics. It becomes quite difficult to process material that is so diverse and inconsistent. When combining data in various forms, it is critical to ensure that judgments are correct and consistent. Handling such a massive amount of data requires a lot of effort and careful consideration in order to keep infrastructure costs low. Another issue is the sheer volume of data. Live feeds of global health information are difficult to retain because they demand extremely quick input/output rates and extremely efficient computer resources.

Because multidimensional data cannot be accurately represented on a graph, it frequently needs more than one chart to derive meaningful conclusions from the many sets of data. Because clinical data is highly sensitive, data security is a critical component of developing a cloud data tool for healthcare analytics. To ensure that data stored in the cloud is secure, the proper policies, authentication methods, and data encryption must be implemented.

5. CONCLUSION

AI-based solutions that use HADOOP's Big Data Analytics characteristics and the ubiquitous use of Cloud Computing are transforming healthcare facilities and services to be more patient-centered. They also provide economical ways to satisfy the community's health needs, even during a crisis. The Hadoop plus Map Reduce framework is used in the healthcare industry to process massive numbers of varied complex datasets (HCD) for next-generation sequencing analysis. Sqoop, Hive, and other technologies are utilized in the data mitigation process to assist with spotting, feature extraction, segmentation, and processing using available analytics models. It is feasible to transform the way the health care industry works with the correct mix of analytics models and data mining techniques, as well as the preservation of heterogeneous data sets (HDS), opening up a whole new universe of possible uses and benefits. These models will enable all governments and countries to develop health policies that are more open, collaborative, goal-oriented, effective, and adaptable to changing demands. It would be possible to uncover new approaches to aid mental health problems by placing all available data sets into the correct Prescriptive analytics system, which is a critical concern. Similarly, a thorough model will be extremely beneficial in ensuring that quality standards in pharma-co vigilance are met globally. The applications market will expand significantly if issues with data storage, security, privacy, latency in data transmission, and network disruption for critical apps are resolved.

REFERENCES

[1] Shilo, S., Rossman, H. & Segal, E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med 26, 29–38 (2020).https://doi.org/10.1038/s41591-019-0727-5.

[2] The Old Bailey and OCR: Benchmarking AWS, Azure, and GCP with 180,000 Page Images DocEng '20: Proceedings of the ACM Symposium on Document Engineering 2020 September 2020 Article No.: 19Pages 1–4https://doi.org/10.1145/3395027.3419595

[3] Banerjee, A., Chakraborty, C., Kumar, A., & Biswas, D. (2020). Emerging trends in IoT and big data analytics for biomedical and health care technologies. In Handbook of data science approaches for biomedical engineering (pp. 121-152). Academic Press.

[4] Bani-Salameh, H.; Al-Qawaqneh, M.; Taamneh, S. Investigating the Adoption of Big Data Management in Healthcare in Jordan. Data 2021, 6, 16. https:// doi.org/10.3390/data6020016

[5] D. E. O'Leary, "Artificial Intelligence and Big Data," in IEEE Intelligent Systems, vol. 28, no. 2, pp. 96-99, March-April 2013, doi: 10.1109/MIS.2013.39.

[6] Z. He, Z. Cai, Y. Sun, et al., Customized privacy preserving for inherent data and latent data, ACM Pers. Ubiquitous Compute. 21 (1) (2017) 43–54.

[7] Cloud-Based Big Data Analytics—A Survey of Current Research and Future Directions Samiya Khan, Kashish Ara Shakil and Mansaf Alam© Springer Nature Singapore Pte Ltd. 2018 V.B. Aggarwal et al. (eds.), Big Data Analytics, Advances in Intelligent Systems and Computing 654, https://doi.org/10.1007/978-981-10-6620-7_57

[8] M.M.E. Mahmoud, J.J.P.C. Rodrigues, S.H. Ahmed, et al., Enabling technologies on cloud of things for smart healthcare, IEEE Access 6 (2018) 31950–31967

[9] M. Daniels, J. Rose, C. Farkas, Protecting patients' data: An efficient method for health data privacy, in: Proceedings of the 13th International Conference on Availability, Reliability and Security, ACM, 2018, p. 9.

[10] Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." Health information science and systems 2.1 (2014): 1-10.