International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

An Effective Data Management Framework for Healthcare: Big Data Perspective: A Review

Sakshi Raj Singh[1], Dr Mohammad Zunnun Khan[2]

Computer Science and Engineering, Integral University, Lucknow, India
srssingh2022[at]gmail.com
zunnunkhan[at]gmail.com

Abstract: ***This paper provides a broad overview of the relationship between big data and healthcare. In the healthcare industry, big data architecture and techniques can be used to manage data growth. The first step toward a better understanding of how big data affects healthcare is to conduct an empirical study. A growing number of healthcare professionals are utilising big data. Machine learning and big data are difficult to predict in the health care industry. When it comes to disease diagnosis, machine learning and big data analytics have ignored privacy and security.***

Keywords: Big Data Analytics, Healthcare, Electronics Health Record, Medical Record

1. Introduction

The healthcare industry has decided to abandon a conservative approach to diagnosis and treatment after decades of doing so. Big Data - based solutions are becoming increasingly popular as a result of the rise of chronic diseases, globalisation, technological advancements, and a push for evidence - based medicine.

Patient - centered care can be delivered using Big Data solutions that provide a complete picture of each patient. Patients will benefit from improved medical care as a result. In an EBM treatment decision - making process, no doctors are involved; instead, scientific evidence is used to produce a quantifiable result.

The concept of big data has been useful in a variety of industries, and its technologies have been put to good use in a variety of fields. [1] The use of big data in health care has a lot of potential. Because of the massive amount of data that must be sorted through in order to make sense of it, data management is critical in healthcare.

The use of mobile and wearable sensors has aided the proliferation of data sources in healthcare. Traditional data analysis methods are becoming increasingly ineffective as the volume and variety of medical data grows. Prescriptive analytics, as well as descriptive, diagnostic, and predictive analytics, are used in the healthcare industry.

Descriptive Analysis: Data can be analysed using descriptive statistics. Reporting current events requires describing and critiquing the current state. There are a variety of methods that can be used at this level of analysis. Typical descriptive analytics tools include histograms and charts.

Diagnostic Analysis: It's time to take stock. One of the most important aims of this article is to provide an explanation for how and why certain events took place. Clustering and decision trees are frequently used in the diagnosis of patients, as are other diagnostic analysis techniques.

Predictive Analytics having the ability to predict what will happen in the future. Prediction, trend identification, and the calculation of uncertain outcome probabilities are just a few of the skills demonstrated. The likelihood of a patient having a complication can be predicted using data. Predictive models are frequently built using machine learning.

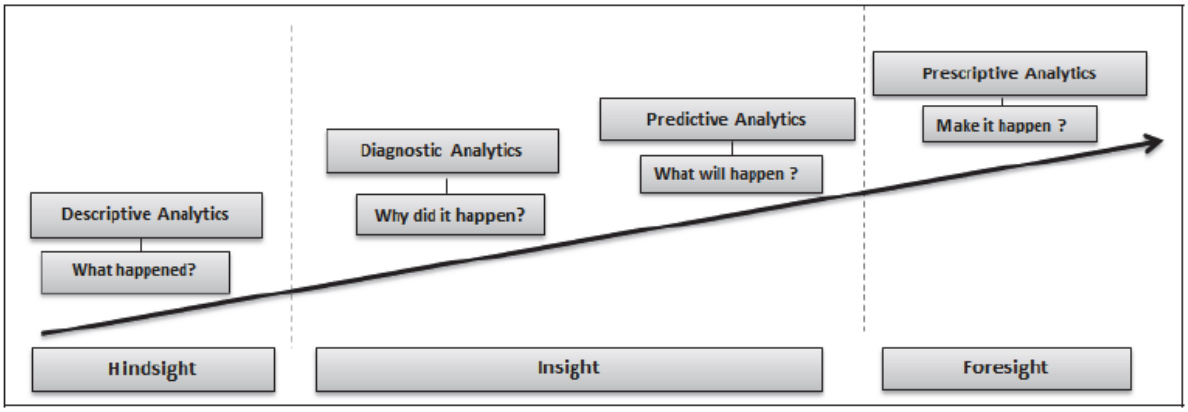

Figure 1: Analytics for healthcare domain

Data Management - Based Healthcare Systems

It's impossible to deny that researchers are in awe of the potential of big data to improve healthcare analytics. A wide range of diseases can now be diagnosed and treated using health informatics because of recent advances in big data health informatics [13]. Security and privacy concerns have been raised as a result of this investigation.

Raghupathi et al. [14] discussed the challenges and architecture of big data health analytics. Big data healthcare systems are vulnerable, according to another study, if safety and privacy issues aren't taken care of right away. In addition, a large amount of data is generated by healthcare systems that can be used to improve patient care. Analyses healthcare data using data mining techniques such as survival analysis and patient similarity. It is possible to perform multidimensional analysis on large medical data sets using this framework.

Data mining is used to analyse a large amount of information in this framework. Smarthealth, a platform for healthcare analytics, was built with the help of ICT. Wiki - Health can be used to analyse sensor data. Data storage and querying are all included in this platform. All aspects of the data lifecycle must be handled by an application. A query and analysis layer is in charge of data storage and retrieval. asserts that privacy and data protection issues have complicated the creation of such platforms. An intelligent healthcare management system.

Data management Applications

Big data doing so a growing number of businesses, organisations, and individuals are turning to big data for a variety of purposes. According to Table 1, the most common uses of big data are shown. it a hot topic since 2010 and is still

Table 1: Applications of Data Management in various categories

Category	Applications
Public Sector	Tax reduction, Social security, Energy exploration, Environmental protection, Power investigation, Public safety.
Healthcare industry	Cost reduction in medical treatments, Prediction of diseases, Eliminate the risk factors associate with diseases, Improves the preventive care , Analyzing drug efficiency.
Education and learning	Students' preferred learning mode, Track students' performance, Provide guidance, Gives real time feedbacks and updates, Improving the learning material, Cross checking of assignments, Digital students assessment.
Insurance industry	Predicting customer behavior, Evaluate the risk of insuring, Monitoring real time claims, Customer retention, Managing premium for the policies, Manage the fraudulent claims.
Transportation sector	Traffic control, Route planning, Intelligent transport systems, Congestion management, Revenue management in private sector, Technological enhancement, Forecasting routes to reduce cast on petroleum.
Industrial and natural resources	Integrating geospatial, temporal, graphical and text data, Analyze consumption of utilities.
Banking	Analyzing big businesses, Prognostic Analytics, Analyzing shopping patterns of customers, Analyzing CRM tactics of competitors, Customer statistics alteration.
Fraud detection	Detect misuse of credit cards, debit cards, Archival of inspection tracks, Treatment for venture credit hazard, Public analytics for business
Entertainment	Manage content for target audience, Measure content performance.

Suggested Electronic Health Record

This section of our paper will go over the requirements for our system and the various components in greater detail. We'll compare and contrast the most popular open - source projects in the final section.

A. System description

We still have a long way to go before we can implement our management system rules after researching and studying the various stakeholders involved in data collection and storage in a centralised server that medical organisations can access. The EHR management system's specifications and requirements are shown in Figure 1.

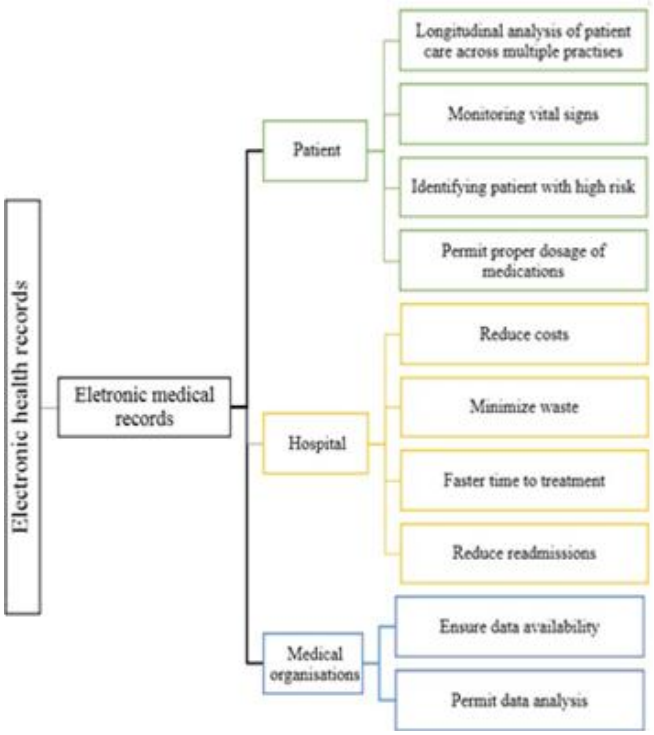

Figure 1: EHR management system needs

This diagram depicts the features that our software will require. Furthermore, it is critical that medical institutions safeguard their own interests.

An ideal system would detect high - risk patients in real time and administer the appropriate medication dose based on the patient's previous medical history. As a result of this implementation, medical facilities will save money and resources, and they will be able to process data more quickly.

B. Architecture overview

An EHR system capable of interacting with multiple Emrs would be included in our design. The design of a system that will allow Emrs and the EHR repository to exchange health data is the focus of this phase of our project. A discussion of electronic medical records will follow an introduction to our architecture (EMRs). Because of the lack of coordination in the sector, health care in developing countries will be used as a case study. Others in the industry perform unnecessary medical tests and procedures.

2. Literature Review

Yu, H., & Wang, D. (2012) more and more patient data is being collected and stored by newer health care IT systems. EMR or PHR data can be used to provide medical advice to patients, while the results of data analysis can be used in scientific research. Because of this, a relational database management system is incapable of meeting the needs of large - scale health care data sets. Archived data stores can benefit from improved performance, scalability, and fault tolerance thanks to a Hadoop - based solution described in this paper. We've laid the groundwork for effective data management. [**1]**

Benhlima, L. (2018) big data - driven healthcare monitoring systems are examined in this paper. Big data processing methods and technologies have also been extensively discussed. Large amounts of data generated in real time by a wide range of medical sources can now be processed by a big data architecture in the healthcare industry. Based on big data analytics, this is the plan. After creating a generic big data architecture for healthcare, it created batch and stream processing to allow for simultaneous generation of accurate predictions and dashboards. [**2]**

Shakil, K. A. et al [2020]Medical technology has advanced to the point where an ever - increasing volume of electronic medical data poses a security risk. Medical records are stored in a variety of formats in the healthcare data management system. There is a massive amount of unstructured and inaccessible data as a result. Many hospitals in the United States have multiple locations. Patients' health data from various locations may need to be combined from time to time for research purposes. A cloud - based healthcare management system can thus facilitate the efficient management of health - care data. However, the most pressing concern with a cloud - based healthcare system is security. This category of criminal activity includes identity theft, tax fraud, bank fraud, insurance fraud, medical fraud, and the defamation of well - known patients. [**3]**

Rallapalli, S. [2016]The analysis of large amounts of data is a major challenge for healthcare organisations. In recent years, the variety of data generated by healthcare devices has increased dramatically. Data must be processed and effectively analysed in order to make better decisions. With cloud computing, data storage, processing, and analysis can all be done on - demand. Traditional data processing systems can no longer keep up with the sheer volume of data. To improve performance and solve scalability issues, we need a better distributed system on the cloud. [**4]**

Khennou, F et al (2016) the amount of data generated on a daily basis is rapidly increasing across a wide range of industries. To properly store, process, and analyse these massive amounts of data, creative problem solving is required. Because of the variety of data we deal with, the time it takes to analyse it, and the speed at which it is collected, the concept of big data has a lot of value in the healthcare industry. In this paper, we hope to present a new health data architecture model. The framework supports an unstructured medical data storage and management framework based on the multi - agent paradigm. With the integration of the mobile agent model into the Hadoop ecosystem, we will be able to instantly connect multiple health repositories. **[5]**

Babar, M. et al (2019) The unbroken amplification of a flexible urban setup is hampered by Big Data processing. In a smart city, making decisions based on the vast amounts of data generated is difficult. The goal of Big Data analytics is to analyse large amounts of data. Conventional methods are no longer effective due to the massive amount of data. **[6]**

Bani - Salameh et al (2021) Software developers and data scientists use big data to uncover new insights and develop

better solutions for improving healthcare and patient safety. Big data analytics (BDA) is gaining popularity due to its importance in healthcare decision - making. The implementation of big data analytics and management in Jordanian healthcare organisations is examined in this article. In Jordan, there are numerous challenges and limitations to managing and analysing big data in the health sector, which are discussed here. This conceptual framework can be used to analyse health big data. According to the conceptual framework proposed in this paper, we could merge the current health information system (HIS) and HIE, which could help us gain insights from our massive datasets and reduce resource waste. By using the framework to process the collected data to develop knowledge and support decision - making, health care quality can be improved for both the community and individuals. [**7]**

Mathew, P. S et al (2015) the amount of data generated in the healthcare industry is growing at an exponential rate, and this trend is expected to continue in the near future. It's not uncommon for the vast majority of healthcare data (e. g., medical records and insurance claims) to be unstructured, stored in silos, and dispersed across multiple systems. It is critical to integrate and factor in these disparate data sets in order to improve healthcare outcomes. Healthcare organisations are unable to take advantage of the wealth of data they possess, either because it is stored in silos in incompatible formats or because of a lack of processing power that prevents them from quickly loading and querying large datasets. [**8]**

Kaur, P., Sharma et al (2018) this paper gives a basic introduction to big data and how it can be applied to healthcare. The use of big data architecture and techniques in the healthcare industry can effectively manage data growth. To better understand the role of big data in the healthcare sector, the first step is to conduct an empirical study. There has been a lot of work done with big data in the healthcare sector. It's difficult to imagine how machine learning and big data will affect the healthcare industry. [**9]**

De Silva et al [2015] the healthcare industry generates a large amount of medical, clinical, and omics data with varying levels of complexity and features. Clinical decision - support is gaining traction as medical institutions and regulatory bodies seek to better manage this data for more effective and efficient healthcare delivery and quality - assured outcomes. The amassing of data across all stages, from disease diagnosis to palliative care, provides additional evidence of the opportunities and challenges of effective data management, analysis, prediction, and optimization techniques as part of knowledge management in clinical environments. **[10]**

About Data Management

The term "Big Data" is frequently associated with the technology that enables its use. The size of the dataset and the complexity of operations required for its processing impose strict memory storage and computational performance requirements. According to Google Trends, the most popular search term associated with the term "Big Data" is "Hadoop" Hadoop is a free and open - source framework for processing large amounts of data using a variety of programming models across a distributed network of computers. HDFS is a file system that allows data dispersed across multiple machines to be accessed without having to deal with the complexity inherent in their dispersed nature. MapReduce is a programming model for efficiently implementing distributed and parallel algorithms, and HDFS is its file system. While MapReduce (Dean & Ghemawat) was created as an Apache open source project, both HDFS (Shvachko et al.2010) and HDFS (Dean & Ghemawat 2008) were proposed by Google (Ghemawat et al.2003). (Ghemawat et al.2003). This demonstrates Google's importance in the development of current Big Data thinking. Hadoop has a number of modules and libraries that can be used in conjunction with HDFS and MapReduce to meet a variety of coordination and analysis needs as well as workflow design requirements for Big Data applications.

Opportunities for Data Management Solutions in Health Care

The data management solutions can be used in the health care to get innovative outcomes in the following areas:

- **Clinical decision support - BDA technologies can be used to predict outcomes [6] or recommend alternative treatments to clinicians and patients at the point of care.**
- **Personalized care** - Predictive data mining and analytic solutions can be used to diagnose disease symptoms in patients before they appear. Patients who are elderly or disabled can wear sensors on their clothing to monitor drug efficacy in real time.
- **Public and population health** - BDA solutions that mine web - based and social media data can predict flu outbreaks and population health trends.
- **Fraud Detection** – Predictive models such as decision trees, neural networks, and regression analysis are used in anti - fraud measures (to name a few). [17].
- **Secondary usage of health data [8]** - Medical records for non - medical purposes Clinical data analysis can now be used to identify patients with rare diseases, research treatment options, and evaluate clinical outcomes.
- **Evidence based medicine [8]:**
  Evidence - based medicine doctors make diagnoses based on statistical studies [8]. Doctors can make the best decisions possible using a combination of intuition and the most recent scientific evidence.

3. Conclusion

The healthcare industry is ready to embrace new technologies in order to improve patient care and lower healthcare costs. As a result of digitization, healthcare providers are now producing massive amounts of digital data. The proper use of health - care data can result in better outcomes at lower costs. Researchers are given analytic tools to help them make the most of the massive amounts of healthcare data they have at their disposal. If the right tools are used, data analytics in healthcare has the potential to produce positive results. In the future, researchers may look into some of the issues raised here.

References

[1] Yu, H., & Wang, D. (2012, August). Research and implementation of massive health care data management and analysis based on HADOOP In *2012 Fourth International Conference on Computational and Information Sciences* (pp.514 - 517). IEEE.

[2] Benhlima, L. (2018). Big data management for healthcare systems: architecture, requirements, and implementation. *Advances in bioinformatics*, *2018*.

[3] Shakil, K. A., Zareen, F. J., Alam, M., &Jabin, S. (2020). BAMHealthCloud: A biometric authentication and data management system for healthcare data in cloud. *Journal of King Saud University - Computer and Information Sciences*, *32* (1), 57 - 64.

[4] Rallapalli, S., Gondkar, R., &Ketavarapu, U. P. K. (2016). Impact of processing and analyzing healthcare big data on cloud computing environment by implementing hadoop cluster. *Procedia Computer Science*, *85*, 16 - 22.

[5] Khennou, F., Khamlichi, Y. I., &Chaoui, N. E. H. (2016, October). Designing a health data management system based hadoop - agent. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)* (pp.71 - 76). IEEE.

[6] Babar, M., Arif, F., Jan, M. A., Tan, Z., & Khan, F. (2019). Urban data management system: Towards Big Data analytics for Internet of Things based smart urban environment using customized Hadoop. *Future Generation Computer Systems*, *96*, 398 - 409.

[7] Bani - Salameh, H., Al - Qawaqneh, M., &Taamneh, S. (2021). Investigating the Adoption of Big Data Management in Healthcare in Jordan. *Data*, *6* (2), 16.

[8] Mathew, P. S., & Pillai, A. S. (2015, March). Big Data solutions in Healthcare: Problems and perspectives. In *2015 International conference on innovations in information, embedded and communication systems (ICIIECS)* (pp.1 - 6). IEEE.

[9] Kaur, P., Sharma, M., & Mittal, M. (2018). Big data and machine learning based secure healthcare framework. *Procedia computer science*, *132*, 1049 - 1059.

[10] De Silva, D., Burstein, F., Jelinek, H. F., &Stranieri, A. (2015). Addressing the complexities of big data analytics in healthcare: The diabetes screening case. *Australasian Journal of Information Systems*, *19*.

[11] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential, " Health Information Science and Systems, vol.2, article 3, 2014.

[12] I. Olaronke and O. Oluwaseun, "Big data in healthcare: Prospects, challenges and resolutions, " in Proceedings of the 2016 Future Technologies Conference, FTC 2016, pp.1152–1157, usa, December 2016.

[13] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare, " BioMed Research International, vol.2015, Article ID 370194, 2015.

[14] J. Sun and C. K. Reddy, "Big data analytics for healthcare, " in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p.1525, Chicago, Ill, USA, August 2013.

[15] M. Bochicchio, A. Cuzzocrea, and L. Vaira, "A big data analytics framework for supporting multidimensional mining over big healthcare data, " in Proceedings of the 15th IEEE International Conference onMachine Learning and Applications, ICMLA 2016, pp.508–513, usa, December 2016.

[16] S. Sakr and A. Elgammal, "Towards a Comprehensive Data Analytics Framework for Smart Healthcare Services, " Big Data Research, vol.4, pp.44–58, 2016.

[17] Y. Li, C. Wu, L. Guo, C. - H. Lee, and Y. Guo, "Wiki - health: A big data platform for health sensor data management, " Cloud Computing Applications for Quality Health Care Delivery, pp.59–77, 2014.

[18] N. Poh, S. Tirunagari, andD. Windridge, "Challenges in designing an online healthcare platform for personalised patient analytics, " in Proceedings of the 2014 IEEE Symposium on Computational Intelligence in Big Data, CIBD 2014, usa, December 2014.