

BIG DATA ANALYTICS IN HEALTHCARE DOMAIN: LITERATURE REVIEW

F.Regina Ananthi¹,Dr.Renuga Devi²

¹Research Scholar, VISTAS, Chennai, Assistant Professor, Guru Nanak College, Chennai,
Mail-id: regina.ananthi11@gmail.com

²Research Supervisor, Assistant Professor, Department of Computer Science, VISTAS, Chennai.
Mail-id: nicrdevi@gmail.com

Abstract - Now a day's very large amount of data has been produced and stored in health care domain. However for advanced clinical research these datas are yet to be improved to unify. The concept of big data in health sector, points to significant challenges. Big data Healthcare refer to collection, analysis and exploitation of patient's physical and clinical data such as Electronic Health Record (EHR) and bio medical data that is too broad and complex. This paper focus on recent research efforts, data sources, big data frameworks, various analysis and challenging issues like privacy and security in related to Healthcare.

Keywords- Big data, Healthcare, EHR, Frameworks, Analysis

1. INTRODUCTION

Huge amount of unstructured data has been generated by Healthcare organization from various sources like clinical, prescription and physical data. Big data helps to improve treatments. It collects large amount of data to validate the patients' habits, detect diseases and predict outcomes. Big data Analytics has an effective pathway that predicts patient's health, cost cut; enhance medicine and its quality. In recent days use of technologies like mobile and wearable sensors has been widespread and as a result large number of data has been flooded in healthcare domain. Therefore to handle very high volume of data on a traditional method has been a challenging one. Big data Analytics in health care domain contain four main types. i) Descriptive ii) Diagnostic iii) Predictive and iv) Prescriptive Analytics.Fig.1

Descriptive Analytics: Characteristics of a data can be analyze and described with the help of descriptive analysis. It is nothing but summarization of information.

Diagnostic Analytics: It explains what happened and why it happened by focusing on past performance.

Predictive Analytics: We determine the future outcomes with the help of predictive analytics. It utilizes variety of data to make the prediction. Predictive models are highly utilized across number of fields with the help of machine learning techniques.

Prescriptive Analytics: It just tells what action to be taken. It is related to both descriptive and predictive analytics.

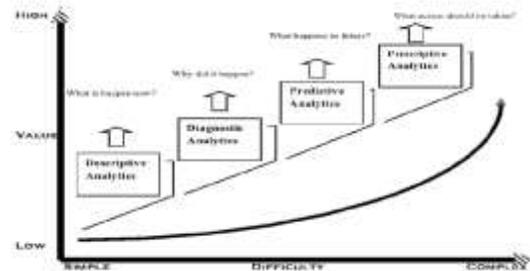


Fig.1 Various Analytics of health care

The content of the paper will be organized as follows:

- In section 2 discuss the related work based on health care.
- In section 3 explains overview of big data architecture in health care domain.
- In section 4 tells various big data frameworks.
- In section 5 discuss privacy and security.
- In section 6 conclusions.

2. RELATED WORK

In [7,8] authors discussed about new innovations and different impacts of big data in healthcare. And also

about health informatics in big data their recent advances and role to handle diseases.

In[9] Healthcare data appears in a variety of formats clinical data in Electronic Health Records (EHR), medical images, sensor data, and genomics, Prescription data, activity claims and cost data from healthcare providers and insurance companies, and patient behavior and sentiment data from wearable devices.

In [1] authors tells about methods to store and retrieve data with the help of big data tools and techniques using Predictive analytics. It also concludes that Hadoop is used to drive sufficient data to healthcare management to predict the needs of people.

In [2] authors discussed predictive analytics including reality mining and vision of big data is utilized for collecting, managing and analyzing healthcare data.

In [3] authors focus on different deep learning data analysis techniques and tools for predicting in healthcare. [4] Main objective of deep learning is to work on large amount of data. The patterns are identified and its features are extracted from complicated unsupervised data automatically without the help of human, which is an important tool in big data analytics.

In [5], authors conclude that, the predicting diseases and cures them is possible by using machine learning algorithms, especially in the field of medicine. They also proved that the performance by predicting gives good results in accuracy and less execution time.

In [6], to improve patient health care outcomes and to achieve best care coordination today's healthcare organizations facing pressure. As a result they are returning towards Predictive analytics. 89 percent of healthcare providers using Predictive analytics and are planning to do same in next 5 years also.

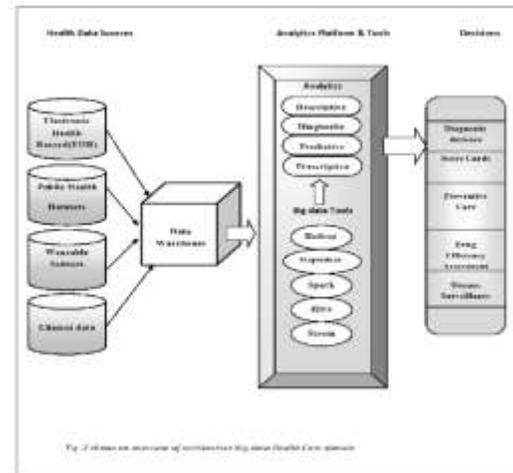


Fig.2 Shows an overview of architecture big data Health Care domain

3. OVERVIEW OF BIG DATA ARCHITECTURE IN HEALTH CARE DOMAIN

Health sector data can be structured or unstructured from multiple locations such as EHR, EMR (Electronic Medical Records), Public health datasets, doctors' notes, Prescriptions and Wearable sensors. Fig. 2 shows the overview of architecture of health care domain. Data can be from any sources. These were raw datas. Data warehouse is an approach were all the data from various different sources are aggregated and that need to be ready for processing. Next step is based on data input approach, tool selection, analytical models several decisions were taken.

4. BIG DATA FRAMEWORKS

All kinds of data such as structure, semi structured and unstructured were supported by big data framework with the help of several features. That feature includes big data mining tools and predictive model design which allows us to make better decision-making process through the relevant information [10].

The following big data platforms are popular in health informatics:

Hadoop:

It is nothing but an open source platform so as to deals with big data analytics. It is easy to use also, adaptable to work with various data sources, either assembling various of data or getting to the data from a database so as to run processor intensive machine learning

process[11]. Numerous huge activities, similar to the assurance of a relationship between's the air quality data and asthma admissions, drug improvement utilizing genomic and proteomic data, and other such parts of healthcare are executing Hadoop. Therefore, with the execution of Hadoop framework, the healthcare analytics won't be kept down.

MapReduce:

Map reduce is a programming model for handling enormous datasets depends on divide and conquer strategy. The divide and conquer technique is actualized in two steps, for example, Map step and Reduce Step. Hadoop executes MapReduce algorithms for preparing and creating enormous datasets. Hadoop deals with two sorts of nodes, for example, master node and worker node. The master node separates the contribution to smaller sub issues and at that point circulates them to wrker node in map step. From that point the master node consolidates the yields for all the subproblems in reduce step. In addition, Hadoop and MapReduce act as an incredible software framework for taking care of huge data issues.

Spark:

Apache Spark is also an open source alternative to Hadoop. It provides more faster and general data processing platform. In-memory processing of data, comparatively, Spark is about 100× faster than Hadoop in multi-pass analytics that is on smaller datasets. This demonstrates processing of huge data with Apache Spark would require a lot of memory.

5. PRIVACY AND SECURITY

Healthcare data is continuously being digitized. The medical data is expanding step by step has arrived at a groundbreaking size everywhere throughout the world. The security and privacy issues are powerful to the point that medical industry isn't skilled to take full profit of it. Due to the developing dangers of misfortune and out streams from individual data it is imperative to verify current Healthcare big data space[12]. The data will be secured by utilizing different innovations like verification, encryption, data mining, and access control.

In big data protection, developers ought to check and break down based on the applications and at that point adjust to protection understandings. Medical big data analytics has different sorts of data protection laws to

keep away from the data spillage of patient. There are a few protection safeguarding strategies in medical big data analytics for example, de-identification, HybrEx, identity-based anonymization.

6. CONCLUSION

The concept of big data in health sector, points to significant challenges. This paper focus on recent research efforts, data sources, big data frameworks, various analysis and challenging issues like privacy and security in related to Healthcare. The main goal of our paper was to make a survey of various big data handling techniques which handle very large amount of data from Healthcare sources and improves overall performance of systems.

REFERENCES

- [1] A. RishikaReddy,P. Suresh Kumar Predictive Big data Analytics in healthcare, 2016 second Inter Conf on CICT,Pp 623-626,2016.
- [2] Hiba Asri, HajarMousannif, Hussan Al Moatassime, Thomas Noel Big data in healthcare:Challenges and Opportunities 2015 Inter Conf on Cloud Technologies and Applications (CloudTech),IEEE,2015
- [3] AnandhavalliMuniasamy,SherishTabassam, MohammadA.Hussain,HabeebaSultana,VasanthiMuniasamy and Roheet Bhatnagar Deep Learning for Predictive Analytics in Healthcare, © Springer Nature Switzerland AG 2020 A. E. Hassanien et al. (Eds.): AMLTA 2019, AISC 921, pp. 32–42, 2020.
- [4] Good fellow, I.Bengio, Y.Courville, A: Deep Learning, 1st edn. The MIT Press Cambridge (2016)
- [5] Basma Boukenze, Hajar Mousannif and AbdelkrimHaqiq, Predicting Analytics in Healthcare System using Data Mining Techniques, David C. Wyld et al. (Eds) : DBDM, CICS, CS CSIP, AI&FL, SCOM, CSE, CCNET-2016pp. 01–09, 2016. © CS & IT-CSCP 2016.
- [6] Jen A. Miller (2019) How Predictive Analytics Is Impacting Patient Care <https://healthtechmagazine.net/article/2019/10/how-predictive-analytics-impacting-patient-care-perfcon> (Accessed Nov 2019)

- [7] Stril A, Dheur S, Hansroul N, Lode A, Meranto S, Moulouguet M, Ruff I (2015) Report on HealthCare.
- [8] T.Huang, L.Lan, X.Fang, P.An, J.Min and F.Wang, Promises and Challenges of Big data Computing in health science, Big data Research, vol.2, no.1,Pp2-11,2015.
- [9] Raghupathi, W., &Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2, 3.
- [10] Naoual El aboudi and Laila Benhlima, Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation, Hindawi Advances in Bioinformatics Volume 2018,Article ID 4059018,10 pages.
- [11] Cheng-Long Ma , Xu-Feng Shang , Yu-Bo Yuan, International conference on machine learning and cybernetics, 2012, vol:4 [12] Mus_ra Siddique, Muhammad Ayzed Mirza, Dr. Mudassar Ahmad, Junaid Chaudhry, and Ra_qul Islam, A Survey of Big Data Security Solutions in Healthcare, Conference Paper · November