

Big Data in Health Care Analysis: A Survey

P. Sharmila¹, P. Rathiga^{2*}

¹Navarasam Arts and Science College for Women, Arachalur, Tamil Nadu, India

²Dept. of computer Applications, Navarasam Arts and Science College for Women, Arachalur, Tamil Nadu, India

Corresponding Author: rathiganavarasam@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7i8.297300> | Available online at: www.ijcseonline.org

Accepted: 15/Aug/2019, Published: 31/Aug/2019

Abstract-The fields of science, engineering and technology are producing data at an exponential rate important to Exabyte(s) of data everyday life. Big data helps us to travel and re-invent many areas not limited to education, health and law. This paper surveys big data with underlining the big data analytics in healthcare analysis.

Big data methodologies can be used for the healthcare data analytics which provide the better decision to accelerate the business profit and customer affection, acquire a better understanding of market behaviors and trends and to provide E-Health services using Digital imaging and communication in Medicine. This paper presents an overview of big data and Health care system benefits, Tools and challenges and review on it.

Keywords: Big data, Health care analysis, Challenges, Tools.

1.INTRODUCTION

The healthcare industry traditionally has produced large amounts of data, driven by record keeping, compliance and regulatory requirements, and patient care [1]. While most data is stored in hard copy format, the current style is toward fast digitization of these huge amounts of data. By definition, big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with old-style software and/ or hardware; nor can they be easily managed with traditional or common data management tools and methods. A familiar current example of using Big Data for health purposes is Google Flu Trends. Within the domain of medical care, technological advances such as on-line health forums and digital data commons⁶ are also opening up new opportunities [2].

In general, it refers to the group of large and complex datasets which are difficult to process using traditional database management tools or data processing applications. These are available in structured, semi-structured and unstructured set-up in petabytes and beyond. Formally, it is defined from 3Vs to 4Vs. 3Vs refers to volume, velocity and variety. Volume refers to the huge amount of data that are being generated daily whereas velocity is the rate of growth and how fast the data are gathered for being analysis. Variety provides information about the types of data such as structured, unstructured, semi structured etc. The fourth V refers to veracity that includes availability and accountability. The main objective of big data analysis is to process data of high volume, velocity, variety and veracity

using various traditional and computational intelligent techniques [3]. After having the 4 V's into account there comes one more V which stands for Value. The bulk of Data having no Value is of no good to the company, unless you turn it into something useful.

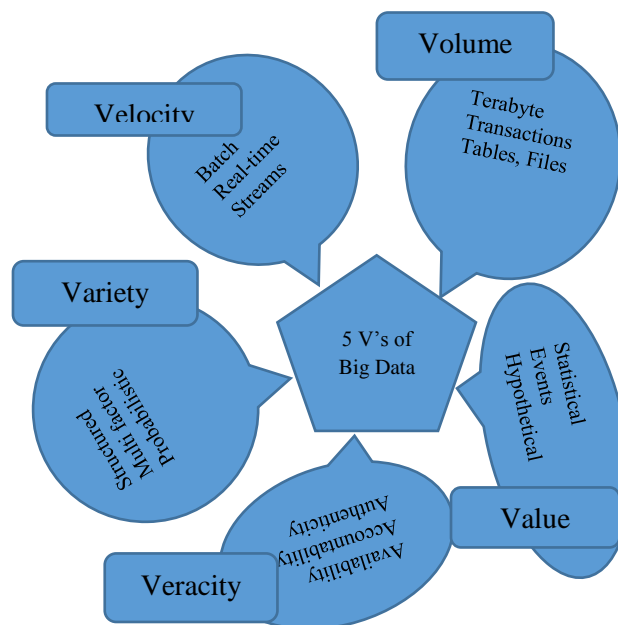


Figure 1.1 5 V's of Big Data

II. OVERVIEW OF THE HEALTH CARE SYSTEM

A health care system is the cluster of people, clinical society and resources that deliver health care services to the patients to make better decision making system. We described the health care system via five attributes which are given below:

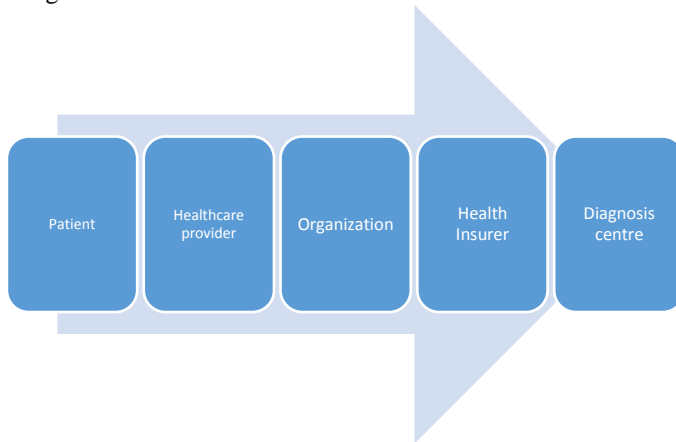


Figure 2.1 Five attributes of health care system

III. OPEN RESEARCH ISSUES IN BIG DATA ANALYTICS

Big data analytics and data science are becoming the research focal point in industries and academia. Data science aims at researching big data and knowledge extraction from data. Applications of big data and data science include information Science, uncertainty modeling, uncertain data analysis, machine learning, statistical learning, pattern recognition, data warehousing, and signal processing. Effective integration of technologies and analysis will result in predicting the future drift of events. Main focus of this section is to discuss open research issues in big data analytics. The research issues pertaining to big data analysis are classified into three broad categories namely internet of things (IoT), cloud computing, bio inspired computing, and quantum computing. However it is not limited to these issues. More research issues related to health care big data can be found in Husing Kuo et al. paper [4].

IV. BENEFITS OF HEALTHCARE BIG DATA ANALYTICS

- ❖ **Advanced patient care:** Electronic health records (EHR) gathers all related demographic and medical data consisting of lab tests, clinical data, diagnoses, medical conditions, allergy information, etc.
- ❖ **Faster time to treatment:** Big data analytics tools support expedite the process by factoring in unique Circumstances, such as lifestyle choices and

demographics, along with the patient's symptoms to help physicians make a more precise diagnosis and formulate the best treatment regimen in real-time.

- ❖ **Reducing energy costs:** Improved monitoring capabilities as part of a Big Data solution make it so healthcare facility staff and managers can pinpoint areas of with high energy use, and determine where it is really required and where energy is being wasted, and then take steps to combat those inadequacies.
- ❖ **For personage/patients/Individuals:** With the help of BDA, the specific treatment is given for a patient based on his genomic data, location, weather, lifestyle, medical history, response to certain medicines, allergies, family history etc. Then the specific line of treatment can be constructed for every individual. By using the BDA techniques, it is easier to predict the disease in earlier stage and treated effectively.
- ❖ **Detecting fraud:** Medical abuse and insurance fraud are two most significant problems facing the healthcare industry. The efficient use of big data in healthcare can solve these problems. By analyzing a larger dataset of claims history, fraud patterns can be identified and even predicted before they occur. Health analytics can process large amounts of data effectively in order to reduce fraud, waste and medical abuse.
- ❖ **Supply Expenditure:** With the use of Big Data Analytics, predictions based on past year's records can be made as to what the estimate for this year will be. Predictive analytics allow hospitals to save a large portion of their money by accurately predicting demand for medical supplies.
- ❖ **Reduced hospitalizations and readmissions:** Predict the patients staying and readmission information. New healthcare plans will be developed to prevent hospitalization. Various questions can be answered by analyzing the data using BDA tools and techniques regarding disease treatment. The hospital management can take and manage administrative decisions in the better way.
- ❖ **Risk stratification:** This data analytics tool helps hospitals track and identify the sickest and often the costliest patients in a proactive way.
- ❖ **Finding cure for diseases:** No two persons in the world would have the same genetic sequences, which is the reason why particular medication seems to work for some people but not for others. Since in one genome there are millions of things to observe, it is almost impossible to study them in detail. Big data in healthcare has, however, revolutionized the field of genomics. Big data analytics can uncover unknown

correlations, hidden patterns, and insights by examining large data-sets [5].

V. CHALLENGES

The important challenge in the health care industry is the collection and processing of massive and heterogeneous data generated through record keeping, compliance and patient related data. The data such generated should be analyzed effectively to predict answers to the challenges that arise. There should be a technology to be built that performs real time analysis on the collected data sets. This will provide instant results to the patients on analyzing with suitable parameters. This will be achieved by means of applications of big data. It helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning algorithm.[6]

VI. TOOLS

The enormous amount of data collected can be classified into useful trends and patterns. Thus it must be preserved, studied and processed. [7]. Following are some of the majorly used Big Data taming tools

Table 5.1 Big Data tools and features

.S.No.	Tools	Features
1.	Map Reduce	It is computational paradigm which works on mapper and reducer tasks which can be executed and re-executed on any node in the cluster.
2.	Hadoop	Framework for similar processing in circulated environment on commodity hardware, comprise a set of primitives to execute batch processing.
3.	Kaggle	It helps organizations and researchers to post their data & statistics. It is the best place to analyze data seamlessly.
4.	Apache Sqoop	It is a tool to interchange the data between Hadoop and relational database.
5.	Hive	Hadoop framework that allows anyone to make queries against data stored in a Hadoop cluster.
		PIG consists of a "Perl-like" language that allows for

6.	PIG	query execution over data stored on a Hadoop cluster, instead of a "SQL-like" language.
7.	Wibi Data	Wibi Data is a combination of web analytics with Hadoop, being built on top of HBase, which is itself a database layer on top of Hadoop.
8.	Splunk	It creates an index of your data as if your data were a book or a block of text. It is just like text search process
9.	Cloudera	The fastest, easiest and highly secure modern big data platform. It allows anyone to get any data across any networks within single, scalable platform.
10.	Openrefine	It provides an environment with cluttered data, cleaning it and transforming it from one into another, also allows extending it with web services and external data. It is a very powerful big data tool.

VII. FUTURE SCOPE OF BIG DATA IN HEALTHCARE

- **Healthcare Data Solutions:** Big Data is used to store vast amount of discontinuous or unbroken data systemically. This makes it easier for healthcare practitioners to access data whenever they want so that they can make informed decisions. It also helps to save time and money spent on finding and collating data.

- **Anti-Cancer Therapy Using Big data:** Cancer has already become one of the foremost causes of temporality and morbidity across the world today. With predictive analytics, pre-existing conditions and habit patterns can be used to predict how unsafe an individual is to cancer. For healthcare providers, big data has permitted them to detect and diagnose even the rarest forms of cancer at an initial stage itself.[8]

VIII. CONCLUSION

Big Data analytics increasingly provide value to healthcare by improving healthcare quality and outcomes and providing cost-effective care. Big Data analytics integration in the healthcare system will improve the tools and techniques and also many challenges can be overcome by

various strategies. Using Big Data analytics in healthcare there will be growth in the healthcare sector, advanced patient care and preventing human errors.

REFERENCE

- [1]. Raghupathi W: "Data Mining in Health Care". In Healthcare Informatics: Improving Efficiency and Productivity. Edited by Kudyba S. Taylor & Francis; 2010:211-223.
- [2]. Professor Alex Pentland, Dr Todd G Reid, and Dr Tracy Heibeck, "Big Data and Health Revolutionizing Medicine and Public Health", Report of the Big Data and Health Working Group 2013, pp:6
- [3]. M. K.Kakhani, S. Kakhani and S. R.Biradar, "Research issues in big data analytics", International Journal of Application or Innovation in Engineering & Management, 2(8) (2015), pp.228-232.
- [4]. MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, "Health big data analytics: current perspectives, challenges and potential solutions", International Journal of Big Data Intelligence, 1 (2014), pp.114-126.
- [5]. Jasleen Kaur Bains "Big Data Analytics in Healthcare- Its Benefits, Phases and Challenges" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 4, April 2016.
- [6]. N. Vanjulavalli, M. Saravanan. "The Impact of Big Data Applications in Health Care Industry", International Journal of Computer Sciences and Engineering, Vol.-7, Special Issue, 4, Feb 2019, PP:325-327.
- [7]. P.Muthulakshmi, S. Udhayapriya, "A Survey on Big Data Issues and Challenges", International Journal of Computer Sciences and Engineering, Vol.-6, Issue-6, Jun 2018, pp: 1238-1244
- [8]. Neha Maurya, Anirudh Tripathi, Pankaj Pratap Singh, Amit Kishor, "Big Data Analysis for Predictive Healthcare Information System", International Journal of Computer Sciences and Engineering, Vol.-7, Issue-6, June 2019, PP: 47- 51.

AUTHOR'S PROFILE

Dr. P. Sharmila, is working as Principal of Navarasam arts and Science College for Women, Arachalur, Erode, and Tamil Nadu. She did Ph.D. from Bharathiar University, Coimbatore. Her area of research is Data Mining and Software Engineering. She has a teaching experience of 20 years. She has published 3 research papers in National and International Journals & presented 28 papers in National and International seminars and Conferences. She has delivered expert lectures on Data Mining and Software Engineering.



Dr. P. Rathiga, is working as Head of the Department in Navarasam arts and Science College for Women, Department of Computer Applications, Arachalur, Erode, and Tamil Nadu. She did Ph.D. from Bharathiar University, Coimbatore. Her area of research is Network Security and Wireless Network. She has a teaching experience of 13 years in Navarasam arts and Science College for women, Erode. She has guided 15 M.Phil. Scholars. She has published 8 research papers in National and International Journals & presented 18 papers in National and International seminars and Conferences. She has delivered expert lectures on Cryptography and Network Security.

