IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

Big Data Analysis for Predictive Healthcare Information System

Neha Maurya[1*], Anirudh Tripathi[2], Pankaj Pratap Singh[3], Amit Kishor[4]

[1,2,3,4]CSE Department , Swami Vivekanand Subharti University , Meerut , India

Corresponding Author: nehamauryamtech@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i6.4751 | Available online at: www.ijcseonline.org

Accepted: 15/Jun/2019, Published: 30/Jun/2019

Abstract-In the era of information, enormous different type of data has become available for decision making. Big data don't refer to that data sets that is big, but also that is high in variety and velocity, which makes them hard to handle using by traditional tools and techniques. The quantity of data that we harvest and eat up is thriving aggressively in the digitized world. Increasing use of new innovations and social media generate vast amount of data that can earn splendid information if properly analysed. This large dataset generally known as big data, do not fit in traditional databases because of its' rich size. Organizations need to manage and analyse big data for better decision making and outcomes. So, big data analytics is receiving a great deal of attention today. In healthcare, big data analytics has thepossibility of advanced patient care and clinical decision support. In this paper, we review the background and the various methods of big data analytics in healthcare. This paper also elaborates various platforms and algorithms for big data analytics and discussion on its advantages and challenges. This survey winds up with a discussion of challenges and future directions.

Keywords: *Big Data, Android, Hadoop, Big Data Mining, Predictive Analytics.*

I. INTRODUCTION

Healthcare big data includes the clinical data, doctor's written notes, prescriptions, electronic patient records (EPR) data, numerous amount of medical journals, diseases data(special diseases, seasonal diseases), medicines data(common medicines) and doctors data(of all speciality). So, huge amount of healthcare data are available for big data scientists. By understanding stencils and trends within the data, big data analytics seems to improve care, save lives and reduce costs. Therefore, big data analytics applications in healthcare take advantage of extracting insights from data for better decisions making purpose. Analytics of big data is the process of inspecting enormous amount of data, from different data sources and in various formats, to deliver insights that can enable decision making in real time. Various analytical concepts (such as data mining) can be applied to analyse the data. Big data analytical approaches can be employed to recognize anomalies which can be found as a result of integrating vast amounts of data from different data sets.In the rest of this paper, firstly we introduce thecommon background, definitions and properties of bigdata.Then various big data platforms and algorithms arediscussed. Eventually, the challenges, future directions andconclusions are presented.

II. BRIEF DESCRIPTION ABOUT THE RESEARCH

- There are lots of diseases that occur every season and are common so with the help of different datasets I am going to analyse what are the chances of any disease to occur in future and going to aware people about it.
- As we all know that due to unpredictive increase in occurrence of diseases availability of doctors get decreases so I am planning to aware people about it and making it easy to cope such kind of problems.
- A feeling bar will be there in order to get information about how an individual is feeling.
- Also providing the list of basic medicines related to symptoms of diseases in order to rectify the diseases

What is a Health Care System?

A **health system**, also sometimes referred to as **health care system** or as **healthcare system**, is the organization of people, institutions, and resources that deliver health care services to meet the health needs of target populations.
There is a wide variety of health systems around the world, with as many histories and organizational structures as there are nations. Implicitly, nations must design and develop health systems in accordance with their needs and resources, although common elements in virtually all health systems are primary healthcare and public health measures.[24] In some countries, health system planning is distributed among market participants. In others, there is a concerted effort among governments, trade unions, charities, religious

organizations, or other co-ordinated bodies to deliver planned health care services targeted to the populations they serve. However, health care planning has been described as often evolutionary rather than revolutionary.[22][23]

III. BIG DATA

- **Definition and properties**

Big Data is a term used for a collection of data sets that are large and complex, which is difficult to store and process using available database management tools or traditionaldata processing applications. The challenge includes capturing, curating, storing, searching, sharing, transferring, analysing and visualization of this data.

- **Big Data Characteristics**

The term big data is described by the following characteristics: *value*, *volume*, *velocity*, *variety*, *veracity* and *variability*,denoted as **6 "Vs"** [1], [2], shown in Figure 1. Besides these 6 "Vs", some authors has defined morethan these 6 properties to describe big data characteristics [3].

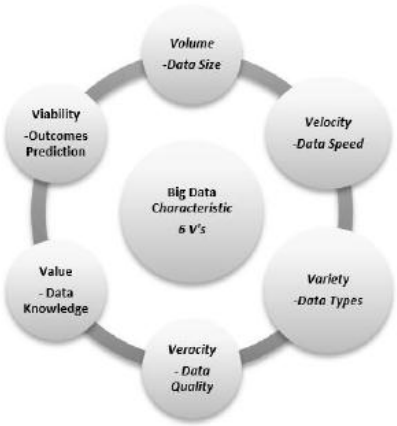

Fig: 1: Characteristics of Bigdata

- **Big Data Analytics**

Applications of big data analytics can improve the patient-based service, to detect spreading diseases earlier,generate new insights into disease mechanisms.Data mining techniques employed on EHRs, identifying the association rules in the EHRs [4] and revealing the disease monitoring and health-based trends. Moreover, integration and analysis of the data with different nature.Nowadays, smart phones are excellent platforms to deliver personal messages to patients to involve them in behavioural changes to improve their wellbeing and health conditions. The mobile phone messages can substitute delivering of medical and motivational advices to the patients [2].

- **Challenges in Big Data Analytics**

Regarding collection of large amount data, some challenging issues should beconsidered. Obtaining high throughputbefore employing of the data mining methods. Different data mining techniques can be applied on theseheterogeneous biomedical data sets, such as: anomaly detection, clustering, classification, association rules aswell as summarization and visualization of those big data sets.These shortcomings might lead to the unreliability of some of the data points, such as missing values or outliers.Despite of these drawbacks of the – EHRs data arevery influenced by the staff who enteredthe patient's data, which can lead toentering missing values, incorrect data as a result of mistakes, misunderstandingor wrong interpretation of the original data [5]. Integration of data from various databases andstandardization for laboratory protocols and values still remain challenging issues [6].High dimensionality of the – The EHRs data which regard to the individuals/patients, makes datamining techniques to be more challenging task.The subsequent stage is the pre-processing of the data, which usually envelop handling noisy data, outliers,missing values, data transformation and normalization. This data pre-processing enables to be applied statisticaltechniques and data mining methods and thus the big data analytics quality and outcomes can improveand can result with discovering of novel knowledge. This novel knowledge obtained by integration of the – EHRs data should results with improving of the implemented healthcare to the patients as well toadvanced decision making by the healthcare decision policy makers.

Hive

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analysing easy. Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

HiveQL Features

- HiveQL is similar to other SQLs
- Use familiar relational database concepts (tables, rows, Schema...)
- Support multi-table inserts via your code
- Accesses Big Data via table
- Converts SQL queries into MapReduce jobs
- User doesn't need to know MapReduce

Comparison with traditional databases

A schema is applied to a table in traditional databases. In such traditional databases, the table typically enforces the schema when the data is loaded into the table. This enables the database to make sure that the data entered follows the representation of the table as specified by the table definition. This design is called *schema on write*. In comparison, Hive does not verify the data against the table

schema on write. Instead, it subsequently does run time checks when the data is read. This model is called *schema on read*.[21] The two approaches have their own advantages and drawbacks. Checking data against table schema during the load time adds extra overhead, which is why traditional databases take a longer time to load data. Quality checks are performed against the data at the load time to ensure that the data is not corrupt. Early detection of corrupt data ensures early exception handling. Since the tables are forced to match the schema after/during the data load, it has better query time performance. Hive, on the other hand, can load data dynamically without any schema check, ensuring a fast-initial load, but with the drawback of comparatively slower performance at query time. Hive does have an advantage when the schema is not available at the load time, but is instead generated later dynamically.

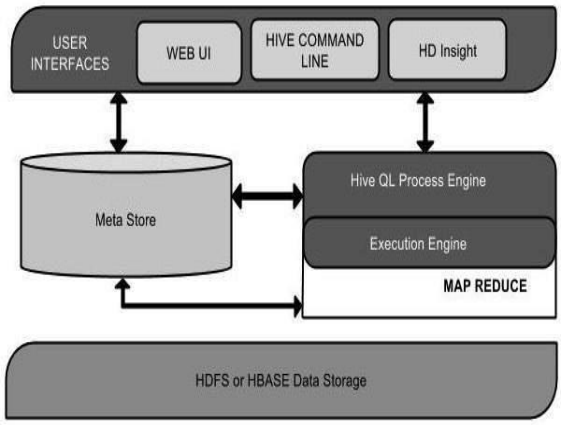

Fig2: Hive Architecture

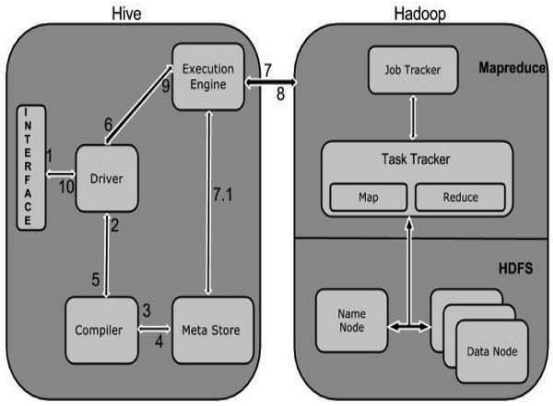

Fig3: Hive Working Diagram

Android

- **Definition and properties**

Android is a mobile operating system (OS) first developed by a Silicon Valley company by the name of Android Inc. A collaboration spearheaded by Google in 2007 through the Open Handset Alliance (OHA) gave Android an edge in delivering a complete software set, which includes the main OS, middleware and specific mobile application, or app.

- **Android Features & Specifications**

Android is a powerful Operating System supporting a large number of applications in Smart Phones. These applications make life more comfortable and advanced for the users. Hardware that support Android are mainly based on ARM architecture platform.

- **User Interface & Navigation**

Your app's user interface is everything that the user can see and interact with. Android provides a variety of pre-built UI components such as structured layout objects and UI controls that allow you to build the graphical user interface for your app. Android also provides other UI modules for special interfaces such as dialogs, notifications, and menus.We construct two modules of our project shown in Fig 5.1 & Fig 5.2.

- **Android Studio**

Android Studio is the official IDE for Android development, and includes everything you need to build Android apps.

[8]According to Canalys, In Q2 2009 Android had 2.8% market share which had grown to 33% market share by Q4 2010 which made Android leader of smart phone OSs worldwide[7]. The market share for commonly used mobile OSs is shown in the following pie chart.

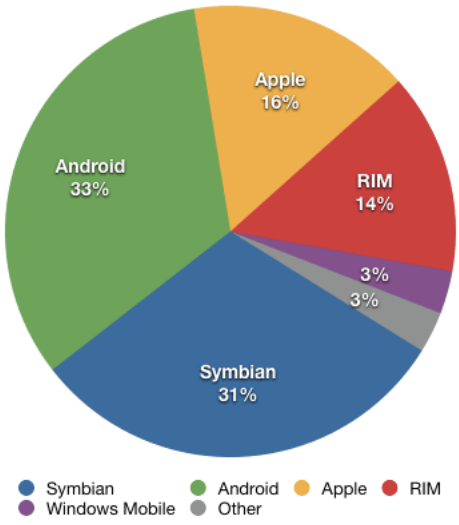

Fig: 4

Module Descriptions:

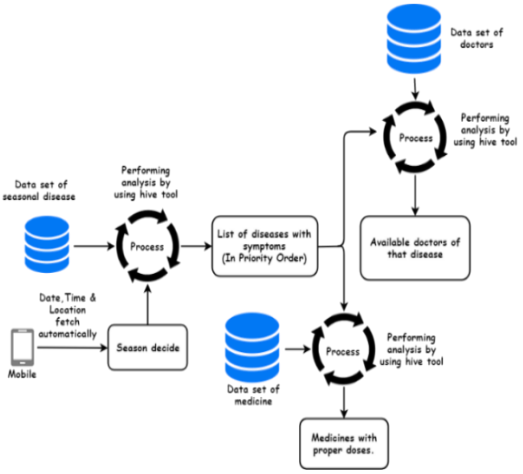

Fig: 5 (Automatic)

Module 1: Mobile will automatically fetch date, time, location which is then utilized in identification of season. Once season is identified then disease analysis will take place using hive tool based on past disease data of that region which will result into list of diseases in a priority order of their occurrence. Then analysis of disease doctor availability take place using hive tool (which will be done on doctors' data of that region) that will result into doctors who will be available in that particular season (for curing disease) And simultaneously basic medicine for curing diseases will be identified from the medicine data of diseases.

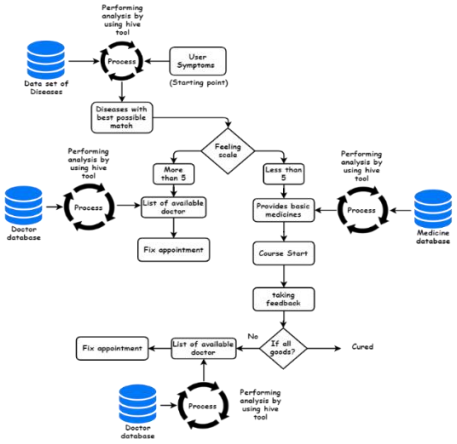

Fig: 6 (Manual)

Module 2: It is the manual section of my application in which user will enter their disease symptoms which will then be utilised for identification of disease from which the user is striving. This process will be done by analysing disease data for symptoms that are entered by user using hive tool and result will be the list of diseases with best possible match in a priority order then a kind of feedback will be taken from the user about their feelings. If they rate their feelings less than 5 then analysis will be done corresponding to that disease on the disease data using hive tool for the identification of basic medicines and user medical course will be started. After few days feedback will be taken again in words and if it is not good then analysis will be done on doctors' data for identification of doctors (of that particular disease) And in case of good feedback the medical course will be stopped. Else if user rate their feelings more than 5 then directly doctors list will be allotted to them by analysing the doctor's data for particular disease.

IV. FUTURE SCOPE OF BIG DATA IN HEALTHCARE

- **Healthcare Data Solutions:** Big Data is used to store huge amount of discontinuous or continuous 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 save time and money spent on finding and collating data.

- **Anti-Cancer Therapy Using Big data:**Cancer has already become one of the leading 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 empowered them to detect and diagnose even the rarest forms of cancer at an early stage itself.

V. CONCLUSION

In the Preamble to this report, fundamental goals of reform are taken into consideration to maintain and improve health and well-being, in order to make basic health coverage universal, and to encourage the efficient use of limited resources. As we know that disease in city/town/country spread rapidly which result into its poor management and rehabilitation due to unavailability of desired number of doctors and lack of alertness about disease as they occur.

The preceding sections of this document have provided a broad framework for attaining these goals via automatic and manual approaches

REFERENCES

[1] Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE J Biomed Health Inform 2015;19:1193–1208.
[2] Archenaa J, Anita EM. A survey of big data analytics in healthcare and government. Procedia Comput Sci 2015;50:408–13.
[3] Borne K. Top 10 big data challenges – a serious look at 10 big data V's. MAPR, 2014:NO4, 80.
[4] Dinov ID, Heavner B, Tang M, Glusman G, Chard K, Darcy M, et al. Predictive big data analytics: a study of Parkinson's disease

using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. PLoS One 2016;11:e0157077.
[5] Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. –Omic and Electronic Health Record Big Data Analytics for Precision Medicine. IEEE Trans Biomed Eng2017;64:263–73.
[6] Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. Biomed Inform Insights 2016;8:1.
[7] https://www.techopedia.com/definition/5415/android
[8] According to Canalys, In Q2 2009 Android
[9] Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. Proteomics 2016;16:741–58
[10]Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. Biomed Inform Insights 2016;8:1. PubMedWeb of ScienceGoogle Scholar
[11]Gaitanou P, Garoufallou E, Balatsoukas P. The effectiveness of big data in health care: a systematic review. In: Metadata and semantics research. 2014:141–53. Google Scholar
[12]Lillo-Castellano JM, Mora-Jimenez I, Santiago-Mozos R, Chavarria-Asso F, Cano-González A, García-Alberola A, et al. Symmetrical compression distance for arrhythmia discrimination in cloud-based big-data services. IEEE J Biomed Health Inform2015;19:125363. Web of ScienceCrossrefPubMedGoogle Scholar
[13]Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE J Biomed HealthInform2015;19:1193–1208. CrossrefGoogle Scholar
[14]Archenaa J, Anita EM. A survey of big data analytics in healthcare and government. Procedia ComputSci2015; 50:40813. Crossref Google Scholar
[15]Borne K. Top 10 big data challenges – a serious look at 10 big data V's. MAPR, 2014:NO4, 80. Google Scholar
[16]Hermon R, Williams PA. Big data in healthcare: what is it used for? In: Australian Ehealth Informatics and Security Conference. 2014:40–9.
[17] R. Chaiken, et. al. Scope: Easy and Efficient Parallel Processing of Massive Data Sets. In Proc. of VLDB, 2008.
[18] HadoopDB Project. Available at http://db.cs.yale.edu/hadoopdb/hadoopdb.html
[19] MicroStrategy. Available at http://www.microstrategy.com
[20] Mysql list partitioning at http://dev.mysql.com/doc/refman/5.1/en/partitioning-list.html.
[21] White, Tom (2010). Hadoop: The Definitive Guide. *O'Reilly Media*. *ISBN 978-1-4493-8973-4*
[22] "Health care system". *Liverpool-ha.org.uk*. Retrieved 6 August 2011.
[23]New Yorker magazine article: "Getting there from here." 26 January 2009
[24]*White F (2015).* "Primary health care and public health: foundations of universal health systems". *Med PrincPract.* ***24****: 103–116.* doi*:*10.1159/000370197
[25]International Journal of Scientific Research in Computer Sciences and Engineering (ISSN: 2320-7639)
[26] International Journal of Scientific Research in Network Security and Communication (ISSN: 2321-3256)

.

Author's Profile

Neha Maurya is pursuing M. Tech. from Subharti Institute of Engineering and Technology, Swami Vivekanand Subharti University, Meerut, India. She received her B. Tech Degree in computer science and Engineering from Uttar Pradesh Technical university, Lucknow, India.

Er. Pankaj Pratap Singh received his B. Tech Degree in Computer Science Engineering from Uttar Pradesh Technical University, Lucknow, India, in 2007 and M. Tech degree in Medical Image and Image Processing from Indian Institute of Technology Kharagpur, Kharagpur, India, in 2010. He is currently working as Assistant Professor in the Department of Information technology, Subharti Institute of Engineering and Technology, Swami Vivekanand Subharti University, Meerut, India. His research interests include IOT, Neural Network, Machine Learning, Deep Learning, Image Processing techniques, Cognitive Science, Computer Network and Data Mining techniques.

Er. Amit Kishor is working as Assistants Professor in the department of Computer Science Engineering and I.T., Subharti Institute of Engineering and Technology, Swami Vivekanand Subharti University, Meerut, India. Currently he is pursuing Ph. D. in Computer Engineering from Department of Computer Science and I.T., Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad.

Er. Anirudh kumar Tripathi is working as Assistant Professor in the department of Computer Science Engineering and I.T., Subharti Institute of Engineering and Technology, Swami Vivekanand Subharti University, Meerut, India.

