

A Review on Leveraging Big Data Analytics in Health Care Sector for Enhanced Diagnosis and Competent Patient Health Care Monitoring System

Sushmitha.S.P¹, Renuka Devi D²PG Scholar, Department of Computer Science, Stella Maris College, Chennai, India¹Assistant Professor, Department of Computer Science, Stella Maris College, Chennai, India²

ABSTRACT: Health care needs well-organized system, techniques and tools to exploit the full benefits of personal healthcare system and healthcare-related data. The sources of health care data are generated by record keeping, patients care and medical diagnosis etc., there is vital requirement of applying Big Data Analytics (BDA) to enormous amount of health care data for enhanced treatment, disease prediction and surveillance. The key objective of this review paper is to provide deep study on the field of Big Data in healthcare sector. BDA is to analyze the immense quantity of dataset. In this paper, we propose the conceptual architecture of BDA in healthcare sector and highlight the applications of BDA, its potential challenges and future directions in healthcare domain.

KEYWORDS: Big Data, healthcare system, BDA, data analytics

1. INTRODUCTION

The BDA is sequential process of groping the immense volume of data to discover new patterns, associations, trends and other useful information which will be used to provide quality data in healthcare at lower cost to all. The conventional DBMS's and RDBMS's are not reasonable for storing large volume of data. These disadvantages are handled by tools and techniques that were offered by Big Data [13]. The conventional data mining techniques are infeasible for handling very large healthcare data and recent researches are more prominent in analyzing the applications such as image processing, signal processing and genomics [3, 2] in health care domain.

1.1 The 5 V's of Big Data in healthcare sector

Volume: This characteristic is about the quantity of data. Every year Healthcare domain generates enormous amount of data. Health related data are first collected, gathered together and the final outcome is enormous volume of data. According to the report [5] says that amount of healthcare data in 2015 was 153 extra bytes and in future it will be enlarged to 2,314 extra bytes by 2020.

Variety: Big data comes from innumerable resources. Generally, health data falls into following three categories they are: structured, unstructured and semi-structured. The sources of data are: medical image processing, medical signal analytics [5], clinical data from clinical decision support systems (CDSS), machine generated sensor data, data from wearable devices, medical claim related data, hospital's administrative data, national health register data, medicine and surgical instruments expiry date identification based on RFID data [1] and social media data like twitter feeds, facebook status, web pages, blogs, articles [2].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

Veracity: This describes about the reliability of data. It is fundamentally a data assurance. Analyzing such a big, variety of data at high rate is complicated task. The key is to ensure whether given data is consistent, integrated and reliable. E.g. Improper data filled by patients, inaccurate readings taken from faulty sensors machines.

Velocity: This refers to the speed rate of data generated at real time. Sensor devices and wearables are used to accumulate the patient's data.

Value: This characteristic is about the quality of data. Generally, EHR's and EMR's are considered as a high value of data. Social media data also has modest value but it is intricate to find out. High value of data leads to better healthcare solution and cost effective. The characteristics are shown in the figure1.

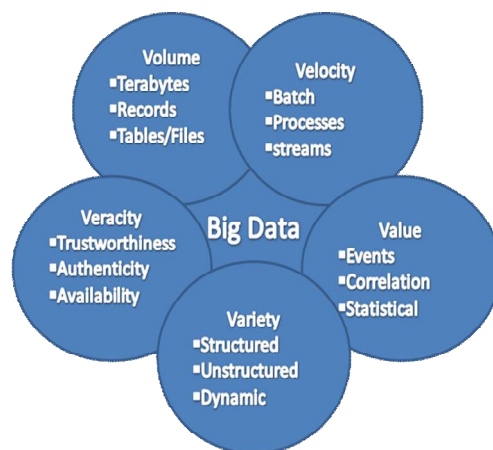


Fig-1: The 5 V's of Big Data

1.2 Stages of Healthcare

There are different levels of healthcare system are primary, secondary and tertiary. These levels are interconnected to each other.

Primary Level: This is the first and lowest level of healthcare system, usually the first interaction between the community members and other levels of health facility. This level can also include general practitioners and persons who provide basic education about food and nutrition, and the importance of water [1].

Secondary Level: This level includes physicians who have basic health training and education. They are usually visited by patients after a referral from a primary health care provider. They are capable of performing minor surgeries [7].

Tertiary Level: This is third and highest level in healthcare system which will provide further advanced treatment and healthcare for patients. They are usually visited by patients after a referral from a primary and secondary health care provider. They are capable of performing major surgeries [7, 1]. It is depicted in the following figure 2.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

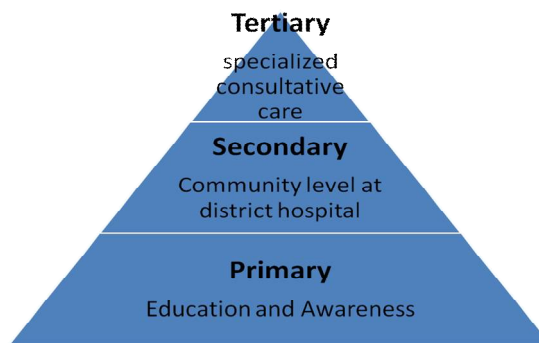


Fig-2: Levels of Health Care System

This review paper is organized as follows. The Section 2 reviews the fundamental conceptual architecture of BDA in healthcare. Section 3 introduces the Infrastructures, platforms and tools. Big data in comparison with conventional databases that are presently used in healthcare is discussed in Section 4. The Challenges and benefits are discussed in Section 5. Application areas of big data analytics in healthcare is briefly introduced in Section 6, big data & IoT in healthcare is discussed in Section 7, followed by conclusions and future enhancement in Section 8.

II. CONCEPTUAL ARCHITECTURE

The conceptual framework of BDA in healthcare is very akin to the conventional medical informatics and analytics. The key difference between the BDA in healthcare and conventional informatics is the steps involved in processing. In conventional health informatics, examination was done by BI(Business Intelligence) tool that can be installed in independent devices. In case of big data, process is divided into many steps and executed across multiple systems using platforms such as hadoop, mapreduce etc.

Source of Healthcare Data:

Besides the algorithms and models similarity between the BDA in healthcare and regular analytics method, they completely differ in usage of user interface tools [8].

Big data on healthcare are accumulated from internal sources (clinical data, data in Electronic Health Records (EHR), Electronic Medical Records (EMR), machine generated sensor data, data from wearable devices, Medical Image data), external sources (government sources, laboratories, pharmacies, insurance companies & HMOs, etc.), often from multiple formats(flat files, relational tables, etc.) ,other applications(data mining, OLAP, database, etc.) and sources and data types include: web and social media, human generated data, Biometric data, Big transaction data. It is shown in the figure 3.

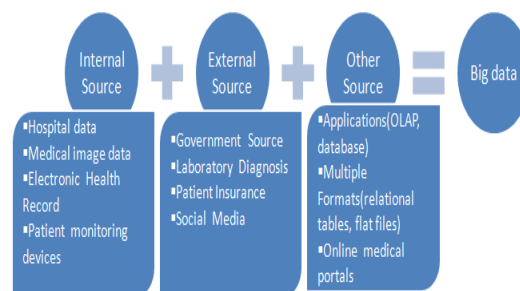


Fig-3: Healthcare Big Data sources

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

The conceptual framework is shown in the figure.4.

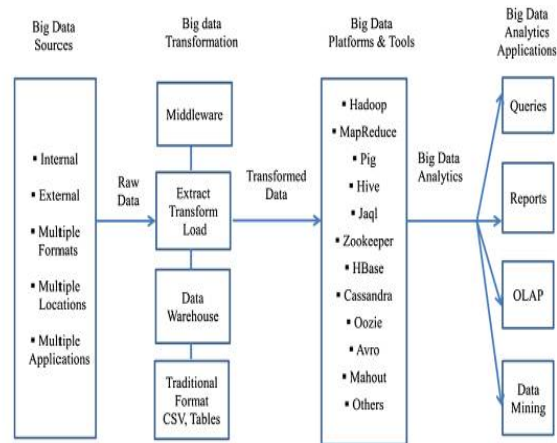


Fig-4: Conceptual Architecture of Big Data

In first phase of big data analytics, the data is accumulated from different big data sources. In the second phase data is in raw state and need to be processed for data transformation. One of the approaches of data warehousing is used to aggregate the data for processing it later. Next phase is used to decide the different data input methods, design for distributed architecture, opt for the platform, tools and analytics models. Finally the distinctive applications of big data analytics includes include queries, reports, OLAP, and data mining is depicted for data visualization [1] [2].

III. INFRASTRUCTURES, PLATFORM AND TOOLS FOR BIG DATA ANALYTICS IN HEALTHCARE SECTORSION

In order to handle and exploit the massive quantity of data, it is essential to select the new infrastructures that can handle the characteristics of big data in healthcare sector (i.e. volume, veracity, variety, velocity, value). The existing cloud infrastructure is capable of managing the big data technologies (like hadoop, mapreduce, lucene etc.) more effectively. It will offer dynamic scalability; parallel computing and infrastructure interoperability which provide the reliable execution (e.g. machine learning algorithms, pattern recognition of images, languages, media, artificial intelligence techniques, semantic interoperability, and 3D visualization and other services) [5].

The Big data tools and platform provide solution to extract, prepare and blend the data in addition to visualizations and analytics. Some of the finest big data tools/platform are listed below:

- **The Hadoop Distributed File System (HDFS):** Designed to handle the unstructured and semi-structured non relational data. It will split up the large data into smaller units and share across various nodes.
- **MapReduce:** It is a core component of Hadoop software framework. This operates on big data and split into manageable parts and parallelly process the large data sets across distributed computer clusters.
- **MongoDB:** It is a No SQL database. It is open-source, cross platform, documented-oriented database written in C++.It is considered as Big Data solutions.
- **Cassandra:** It is distributed No SQL database system. It is designed to handle huge amount of data across many servers. It is providing high availability, performance without any failures.
- **Lucene:** Apache Lucene is a standard open-source community used for text analytics or search engine.
- **HBase:** It is a distributed column-oriented database built on top of the hadoop file system. This uses hash tables and provides random access, stores the data in indexed HDFS files.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

- **Jaql:** It is functional, query language mostly used for JSON query processing on big data. This will convert high level queries to lower level queries to facilitate parallel processing [2].

The most efficient platform for BDA is open-source distributed platform hadoop. It divides large amount of data into smaller units and distributes across various nodes. Hadoop act as data organizer and analytical tool. The main shortcomings of hadoop is it is difficult install, manage and configure. There are many additional platform and tools available to support the hadoop distributed open source (like AWS, Cloudera, Hortonworks and MapR Technologies).

IV. BIG DATA VS. DATABASES PRESENTLY USED IN HEALTHCARE

Big Data differs from conventional database in several ways. Some of the major differences are given below:

Data Structure

Typical relational database follows table format (i.e. every piece of data placed in well-defined manner). In contrast, Big data has minimal structure (i.e. data is extracted from many source in raw form) and stored across multiple systems with help of hadoop distributed file system [15].

Data Types

Unlike conventional datasets (eg. char, int and varchar) which are predefined, big data sets are in structured (e.g. Transaction details, Bank account history) and unstructured form (e.g. Tweet, Facebook messages). These features in addition to its 5V's characteristics makes it impossible to analyze the conventional database.

Cost Effective

Big data is less expensive due to its unstructured nature. Due to lack of predefined structure it is cheaper to own and simpler to create than a traditional database.

Difficulty with Big Data over Traditional Database

In relational database, traditional schemas are used to find the outline of where each piece of data exists. In contrast, there are no traditional schemas and the query languages are much more complicated in big data. In short, big data is cheap but with added difficulty to use. The relational databases are costly but very usable [13].

V. CHALLENGES AND BENEFITS OF BDA

Big data is great platform for processing enormous amount data (like terabytes and petabytes) producing greater result.

For Patients: By using the BDA techniques it is easier to predict the disease in earlier stage and treated effectively.

Fraud Analysis: According to Government medical claims, by using the BDA techniques we can analyze and reduce the act of fraud medical claims.

In Real-Time Healthcare and Clinical Analytics: By using the BDA techniques we can obtain live updates and alert the hospital staffs to act efficiently for problems.

Efficient use of big data will augment the performance and proficient access to information. Big data is not only well known for its opportunities but also it has lot of challenges. While choosing the platform following criteria's should be considered they are availability, feasibility, scalability, continuity, level of security and privacy. Other challenges include healthcare data quality, interoperability, incomplete data and unstructured data. We need to validate and ensure the data which we gathered from source is reliable and of high-quality. Missing or incomplete data can lead to medical errors and big threat to patient's safety. Protection of data is also of the key issue. To improve the data privacy and security we need to strict government policies and rules regarding which data should be shared and which data should not be shared[1][7].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

VI. APPLICATION AREAS OF BDA IN HEALTHCARE

Healthcare analytics produces enormous amount of data and big data in healthcare includes the following application areas, they are: Electronic Health Records (EHRs), Real-time Alerting System, Telemedicine System, Medical Image Processing, and Predictive Analysis and so on.

Electronic Health Records (EHR)

This is one of the most ubiquitous applications in healthcare sector. Every individual (like patients) have their digitalized record. This record includes the patient medical history, allergies and symptoms, laboratory test reports, etc[9]. The evidence-based tools are available to healthcare providers from both public and private sectors and they can use to make decision on patients care. Key features of EHR allows to create and manage health information's in digitalized format by authorized providers and also information shared securely via secure information systems across many providers (e.g. doctors, laboratories, pharmacy and etc).

Real-time Alerting System

This will track and monitors the patient's condition and helps the health practitioners to taking care of patient health (e.g. Clinical Decision Support System is software used to analyze the patient health).This system will bridge the gap between the patient and specialist who is response and take emergency actions in real-time world. For example, in case of sudden high blood pressure for a patient then system will send the alert to respective doctor who will then reach to patient and take action to lower the pressure [9].

Telemedicine System

Telehealth is another example of healthcare analytics. It is a combination of digital information and communication devices (like mobile, laptop, computer etc) to manage and touch with patients in case of any emergency. For example, The Health Resource and Services Administration of U.S Department of Health and Human Services use the electronic information and telecommunication system to manage distance clinical healthcare, patient health [9].

Predictive Analysis in Healthcare

Big data analytics has various techniques like text analytics, multimedia analytics and so on. But one of the determining techniques is predictive analytics which includes modeling, machine learning and data mining. This will analyze the present and historical data to produce predictive future. Optum Labs of U.S research collaborative has composed the electronic information of over 30 million patients and they created a database for predictive analysis. This will help us in predicting disease in earlier stage and preventing the patient sickness. By using predictive analysis tools we can able to predict the disease in earlier stage for example who is at risk of diabetes and advised to make use of additional screenings or weight management.

VII. IoT AND BIG DATA IN HEALTHCARE

IoT is physical gadgets that is accessible by inter-networking such as wearable's, sensors or other objects embedded with electronic signals, software etc. IoT adds more value to healthcare analytics. Big data uses real time instruments to trace the patient's health conditions e.g. devices like sensors and wearables are used to collect the patient data (like blood pressure, heart rate ant etc) from anywhere. Real-time alerting devices provide remote alerts regarding patient's health condition to respective doctors to take the needed action. Smartphone's plays major role in real-time alerting devices. For example, the Smartphone apps like calorie meter are used keep track of calories and these data's are shared with doctors via secure information sharing system which will help them take personalized care and treatment [7].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

VIII. CONCLUSION AND FUTURE ENHANCEMENT

In Big Data analytics, data is processed in many phases to extract the data or discover new pattern from huge information. In this paper we discussed about architecture, challenges and benefits of BDA in healthcare, platform and tools in big data. Generally, healthcare data analytics is the vital topic since we already mentioned about benefits and challenges of BDA. By digitalizing the healthcare information, it will advance the quality of services in terms of analyzing the effectiveness of patient treatment with low cost, strengthen the relationship between patient and doctors and also it will help us to detecting the frauds, drugs abuse quickly and efficiently. Big Data will enhance the healthcare sector to next level with rapid advancement in tools and platform.

REFERENCES

- [1] "Levels of Health Care Referral System", Available: <https://nurseslabs.com/levels-of-health-care-referral-system>
- [2] S.H. El-Sappagh, S. El-Masri, A.M. Riad and M. Elmogy, "Data mining and knowledge discovery: applications, techniques, challenges and process models in healthcare.", International Journal of Engineering Research and Applications, Vol. 3, No. 3, pp. 900-906, (2013).
- [3] D.W. Bates, S. Saria, L. Ohno-Machado, A. Shah and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients", Health Affairs, Vol. 33, No. 7, pp. 1123-1131, (2014).
- [4] Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." Health information science and systems2.1 (2014): 3.
- [5] Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian, "Big Data Analytics in Healthcare", June 2015.
- [6] Gemson Andrew Ebenezer J and Durga S, " Big Data Analytics in Healthcare A Survey", May 2015
- [7] Rohil Shah, Ria Echhpal, Sindhu Nair, "Big Data in Health Care Analytics", October 16 Volume 4 Issue 10 , International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, PP: 134 - 138
- [8] Alkhatib, Mohammad, Amir Talaei-Khoei, and Amir Ghapanchi. "Analysis of Research in Healthcare Data Analytics", 2016.
- [9] Bains, Jasleen Kaur. "Big Data Analytics in Healthcare-Its Benefits, Phases and Challenges." International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Available online at: www.ijarcsse.com 6.4 (2016).
- [10] Househ, M. S., et al. "Big Data, Big Problems: A Healthcare Perspective." Studies in health technology and informatics 238 (2017): 36.
- [11] <https://www.forbes.com/sites/bernardmarr/.../how-big-data-is-changing-healthcare>
- [12] <http://www.bdva.eu/sites/default/files/Big%20Data%20Technologies%20in%20Healthcare.pdf>
- [13] "Examples of Big data in Healthcare", Available: <http://www.datapine.com/blog/big-data-examples-in-healthcare/>
- [14] "Big Data in Healthcare – Real World Use-cases", Available: <http://data-flair.training/blogs/big-data-healthcare-real-world-use-cases/>
- [15] "Big Data in Healthcare Made Simple: Where It Stands Today and Where It's Going", Available: <https://www.healthcatalyst.com/big-data-in-healthcare-made-simple>