SURVEY OF BIG DATA ANALYTICS IN HEALTH CARE

K. AKSHATHA, S. CHANDRAKALA
Department of Computer Science and Engineering,
Sri Krishna College of Engineering and Technology,Coimbatore.
E-mail: 16epcs002@skcet.ac.in
chandrakalas@skcet.ac.in

Abstract

The world is floating by information today. The amount of information that we gather and consume is thriving forcefully in the digitized world. Increased usage of new technical advancements and online networking create unlimited information that can acquire useful insights if appropriately investigated. The dataset which cannot be handled by traditional database System are referred as big data. The size of big data differs based on the data management capacity of the organization. In healthcare, Big data helps in providing the best possible diagnosis and to make smart decisions. This paper discusses the concepts and characteristics of Big Data and the role big data analytics in health care. Challenges of Big data in medical applications are also discussed.

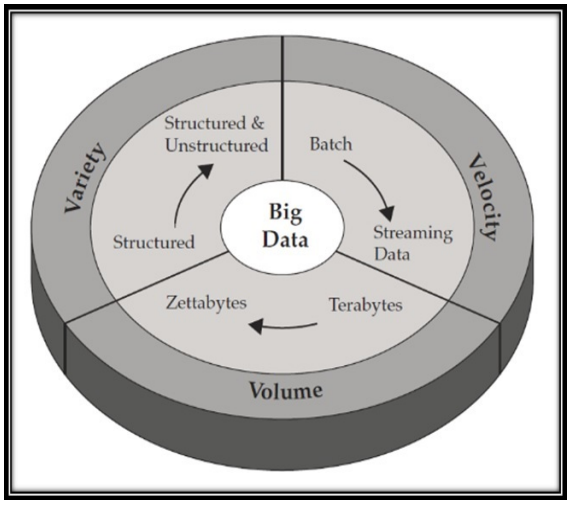
Keywords:Bigdata,CloudComputing, Database, Diagnosis, Health Care.

I. INTRODUCTION

Bigdata Analytics involve the procedure of gathering and sorting huge amount of data to find hidden patterns in it. It helps the organizations from various perspectives like Cost diminishment, smart decision making, etc.... Nowadays, as opposed to putting away the information as printed hard copies, we store them digitally. A large portion of the Information Technology expenses is spent on dealing with the data delivered today. A recent review states that 90% of information accessible online today is recorded in the most recent 5 years. The data put away carefully helps us to safeguard the information for quite a while by diminishing the cost of capacity and social insurance capacities[1]. Big Data can be represented using the 3v's

A. Volume (measure of information orthe span of information collection)

Volume Refers to the unlimited measures of information produced each second, it is not measured in terms of terabytes but rather in zetabytes. It makes data collections too large to store and break down using conventional database system. Big Data tools utilize distributed frameworks, to store and analyze the data's stored in different databases irrespective of their locations. 90% of all information at any point made, was made in the previous 2 years. Starting now and in the near future, the measure of information on the planet will multiply at regular intervals. (I.e.) By 2020, we will have 50 times the measure of information as that we had in 2015. [2]



B. Velocity (speed of information in and out or information On movement)

Velocity Refers to the speed at which new data is produced and the speed at which the data moves around. For Example: Consider the messages delivered in the social networking sites, which becomes viral in few seconds. The speed at which information is made at present is practically impossible: Every moment we transfer 100 hours of video on YouTube. Also, consistently more than 200 million messages are sent, around 20 million photographs are seen and 30,000 transferred in Flicker, just about 300,000 tweets are sent and very nearly 2.5 million inquiries on Google are performed.[3] The organizations are challenged to adapt to the rapid speed the data is made and utilized in reality.

C.Variety (scope of information sorts, areas and sources)

Variety Refers to the diverse sorts of information we can now utilize. In the past, we just centered around organized information that perfectly fitted into spreadsheet tables or databases are called as structured data, for example: budgetary information. To be honest, 80% of the world's information are unstructured (content, pictures, video, voice, and so on.),[4].Semi-structured data falls in between both structured data and semi-structured data. With Big data, we can now break down and unite data's of various sorts, for example: messages, social networking discussions, photographs, sensor information, video or voice recordings.

II. SOURCES OF BIGDATA IN HEALTH CARE

The data to be collected and analyzed by healthcare organizations may come from hospitals, emergency care facilities,Pathology(clinical laboratories), Radiology(medical images), research and other non-traditional data sources.Gathering,sorting and analyzing the data is difficult because, the data are shattered over various locations.Most of the medical data recorded are unstructured data like images,audio,video,..and the quality of the data is not assured.Some of sources of big data in health care is listed below,

* ***Genomic data***

Genomic data include the genetic factors like DNA.It is used to analyze the genetic factors through which better diagnosis can be provided.It requires some specific data analytic tools to process genomic data.

* ***Electronic Health records (EHRs)*** gather, store, and display the data.for example, past medical summaries, Allergic medications, vaccinations,pharmaceuticals, fundamental signs, lab records and radiology tests, advance notes made by human services suppliers.[8]

* ***Electronic medical records (EMRs)*** are not indistinguishable to EHRs. It contains patient information's that are accessible to a specific doctor.

* Patient registries kept up by insurance companies are, frequently connected to the EHR.Various health care providers from different geographical scale can track the records to provide better diagnosis. [5]

* Clinical information distribution centers get patient-level information from various clinical data frameworks, for example, EHRs and different sources recorded previously

* Healthcare sector additionally creates large data sets by taking care of and keeping up beneficial records about patients.

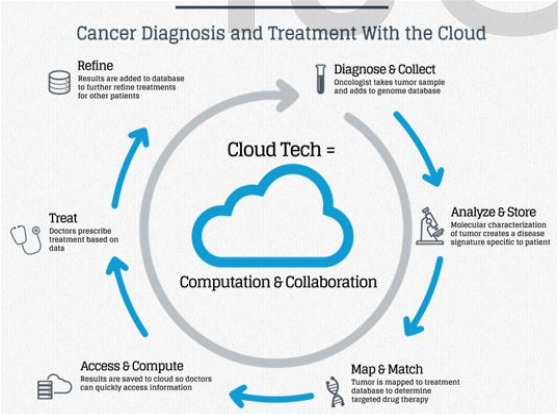
III.APPLICATION OF BIG DATA IN HEALTHCARE

1. Identify the patients, who frequently take medical advice and provide them programs to improve their health and to monitor their status.[4]
2. Providing people with needed medical resources to make smart

decisions for maintaining their own health by tracking their behavior[4].

3. Identifying the medical treatments which don't provide expected improvement or of high cost and to find and replace the most efficient way.[4]
4. Reducing the readmission of the patients by identifying the external factors that trigger the translation.
5. Monitoring the patients by using "at home health monitors" and helping them in case of any abnormalities.
6. Helps to take preventive measures in case of disease outbreaks by detecting the most vulnerable patients.
7. Integrating the clinical,financial and operational data to track the utilization of resources.
8. Providing easy,secure and remote access to the clients.

Case Study:1 Cancer diagnosis and treatment [12]



1. Collect the tumor sample and add the test samples in the genome database.
2. The characterization of the tumor sample identifies the disease to the specific patient.
3. The Tumor is then mapped to the diagnosis details present in the database.
4. The results are saved to the close so that they can be accessed by the doctors as soon as possible.
5. Doctors prescribe the medication based on the analysis report.
6. The outcome is recorded in the database for further reference of other patients.

Case study:2 Neurological disease diagnosis using health care data analytics[11]

Computer aided diagnosis (CAD) system is used to provibe automatic diagnosis of neurological disorders.The CAD system consist of techniques like data pre-processing, feature extraction and classification.

A. Pre-processing: The collected raw data is processed to remove unwanted noises,which in turn reduces the computation time taken by the CAD algotithm.

B. Feature extraction: The biomarkers of the disease is extracted from the source data.

C. Classification: The processed medical data is compared with the data extracted from the original source, to identify whether the patient condition is normal or abnormal.

The following are the few neuro diseases that can be identified using data analytics

1. Epilepsy Diagnosis
2. Alzheimer Diagnosis
3. Brain Tumour Diagnosis
4. Sleeping disorder diagnosis
5. Drug addiction related diagnosis
6. Autism diagnosis

III. DIFFICULTIES OF BIG DATA IN HEALTHCARE

Various Challenges are involved in analyzing the data in health care .Some of them are briefly explained below,

Quality or Nature of information: Nowadays, a lot of information is being gathered and yet there is no Quality

Control(QC). For example: MRI Scan without visual clearance is useless.[9]

Data Analyst: The amount of data collected and analyzed are increasing rapidly.There will not be sufficient data analysts with the necessary skill set to handle the data in the near future.It is mandatory to provide the data analysts with efficient tools to enhance the process.[7]

Data Compliance and Policies: Patient medical records are more profitable but,data privacy and security should be taken into concern.[9] Only 56% of people are ready to share their personal medical records.

Restricted Variations: There is no 'one-arrangement fits-all' model. Developments that may work in one nation may not work in another. Arranging a worldwide enterprise now incorporates, a world separated by the fringes, security laws, temperature ranges, workforce accessibility and shifting expenses for dependable media transmission benefit.

Change Management: Administration change is in truth an unavoidable challenge. The bigger the doctor's facility, the greater the obstacle.The decision made by one management will not be acceptable by other management.

Data Recovery: Data loss happens in all industries,due to some unexpected accidents like fire.Frequent data backup to cloud storage may help.[6]

IV. FUTURE CHALLENGES

1. Future challenges include handling clinical data that are shattered among various departments like labs, data vendors[10].
2. Providing more accurate results based on the information gathered.
3. To improve data security.
4. To design a global framework to analyze the data.

V. CONCLUSION

This paper,covers the basic concepts and characters of big data. Along with the applications and challenges of big data in health care .It also include the future challenges of bid data in health care.

VI. REFERENCES

[1] Lidong Wang and Cheryl Ann Alexander,"Big Data in Medical Applications and Health Care",American medical journal. May 4, 2015.

[2] Priyanka K,et al, / (IJCSIT) International Journal of Computer Science and Information Technologies,"A survey on big data analytic sin health care" Vol. 5 (4) , 2014, 5865-5868.

[3] Gemson Andrew Ebenezer and Durga "BIG DATA ANALYTICS IN HEALTHCARE: A SURVEY",ARPN Journal of engineering and applied sciences Vol:10(8), May 2015.

[4] IBM Corporation," Data-driven healthcare organizations use big data analytics for big gains" ,2013.

[5] White SE," A review of big data in health care: challenges and opportunities", August 15,2015.

[6] Alexandros Labrinidis and H. V. Jagadish,"Challenges and opportunities with big data," Proc.VLDB Endow. 5, pp. 2032-2033, August 2012.

[7] Bottles, K. and E. Begoli, 2014. Understanding the pros and cons of big data analytics. Physician Exec., 40: 6-12.

[8] https://www.forbes.com

[9] Hsieh, J.C., A.H. Li and C.C. Yang, 2013. Mobile, cloud and big data computing: Contributions, challenges and new directions in telecardiology. Int. J. Environ.Res. Public Health, 10: 6131-6153.DOI: 10.3390/ijerph1011613.

[10] Wullianallur Ragupathi and Viju Raghupathi, "Big data analytics in healthcare promise and potential",HISS, 2014.

[11] Siuly, S. & Zhang, Y. Data Sci. Eng. (2016) 1: 54. doi:10.1007/s41019-016-0011-3Medical Big Data: Neurological Diseases Diagnosis Through Medical Data Analysis.

[12] Carrick Carpenter, delivery director for Dell Services Cloud Computing - Healthcare. *Healthcare Industry Expertise Key in Dell.* Cloud Solutions for Medical Practices and Providers