Big Data in Healthcare Sector

Aayushya, Sukhdeep Sharma

Department of Computer Science &Engineering
Manav Rachna International Institute of Research and Studies
Faridabad, Haryana, India
y.aayushya081098@gmail.com, sharma.sukhdeep30@gmail.com

Abstract- Healthcare is an important and budding sector where we have a new disease, its treatment and research going on never ending. It involves diagnosing diseases, providing treatment and curing patients using latest methods and procedures. As the field is wide it requires and generates huge amount of data varying from patient records to medicine discovery, treatment reports, new researches and other sections. This data has to be handled efficiently as it helps in assisting doctors and researchers in providing a better and accurate treatment to patient. In this paper we have provided a review on how big data is being used in the healthcare sections covering its applications in detail and also what challenges are posed while handling such large amount of data.

Keywords- Big Data, Diagnostics, IoT, Smart Devices, Hadoop

1. Introduction

Big data is the enormous amount of data generated by performing various applications and functions. It has to be managed as it can be of high usability. Big data analytics is used to examine bulks of data and deriving useful patterns and information from it. In healthcare, big data is used as well as produced in equally high amounts. Under this section we describe what big data is and how its various characteristics are defined with respect to the healthcare sector.

1.1 Big Data

Big data is nothing but collection enormous amount of data which can't be handled by traditional database systems. From past few decades big data had grown interest of many individuals and industries. It has allowed them to easily handle s enormous amount of data in very short duration. It is growing among industries of public and private sectors, everyone is evolving their systems toward big data technologies. It has accepted by many businesses like health, Education, Food etc. It has helped these industries in many ways like it helped them to generate more profit and saving time .Various technologies are adding up or developed to handle big data and to reduce the time and cost for managing it.

Healthcare sectors have been also benefited by this technology they manage their whole data which is produced in large volumes. Which contains records of patients, reports , records of employees, insurance details, etc. This data is then handled and visualised, which further help them to analyse their profits and various other sectors which lead them to affective decision for their organisation.

1.2 Features of big data used in Medical systems databank

With [1],[2] the rapid growth in amount of data produced in past few years has failed the traditional methods of handling databases. This enormous amount of data can only be handled by big data . The characteristics of big data like variety, volume, veracity, velocity and value. These characteristics also known as 5 V's of big data. Relation of these characteristics with big data is mentioned as following:-
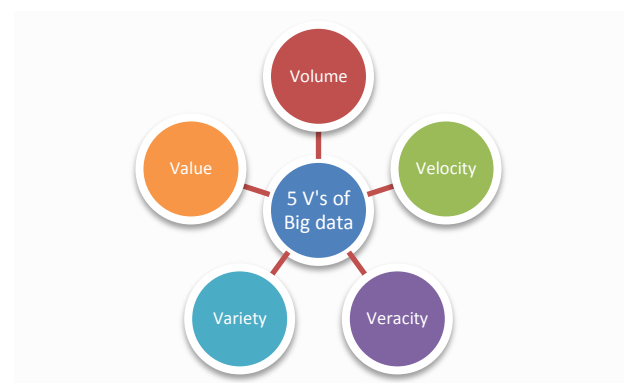

Fig 1 : 5 V's of Big Data

Variety
As the different technologies have been introduced by the time, data is becoming heterogeneous. Data is

produced in different types like structured , semi-structured, unstructured. This can be found in any forms like photos ,text files, videos, etc. This different type of data is managed or handled successfully and efficiently by big data.

In healthcare sectors usually almost information is stored in organised form which is also called structured data. This information can be patients records , Medical information, Doctors information etc. But some data is also found in unstructured format like MRI reports, X-radiation reports, cctv camera footage etc. So that type of information is well managed and saved in the systems which can be used for visualising and classification. Which help the users to produce their information rapidly and efficiently.

Volume

As it is said more than 90% of data present on internet is generated in just a decade. As devices are increasing in large amount which lead to generate vast amount of data. This amount of data cannot be handled by traditional systems. So here's come big data into picture. It has efficiency of handling such enormous amount of data and process it efficiently. It had make this task very easy. We can manage a data or gather information from large data by few clicks.

As data is increasing rapidly in different sectors, data is also increasing in healthcare sector. This is growing in large amount every day. Data like medical researches data, medical institutes information, medical insurance agencies data, government or private hospitals data is example of healthcare data.

Veracity

Veracity is authentication of data. As there is lot of data present on internet. as anybody can put data on internet so there is lot of fake date present over there. This data may affect users in various aspects like it can affect emotionally, religiously, mentally. This data can be generated by anybody like terrorist, atheist or anybody who is mislead or intent to produce hate. So authenticated data is very important for everybody who trust or rely on the internet. We can take example like in any self driving cars information about speed limits or when to use airbag must be true else it can lead to big accidents.

Same in healthcare systems all the data related to organisation should be authenticated or true for making predictions and better decision making.

Velocity

Velocity is how fast can data be generated, managed, stored and processed. Large amount data is produced from various sources. So Main aim is to process data rapidly and it should be updated. Delay in processing a data can be costly and lead to waste of time. So velocity speed of data processing specifies the efficiency of the system. As all the data generated by multiple sources is stored which if further used for analysis. After analysing data queries are responded and results are generated. Big data allows to perform analysis and respond to queries in small unit of time.

Hence, in healthcare systems all the data collected of patients or other units are stored and analysed. So it is required that all the information can be retrieved rapidly and queries can be responded as fast as possible. Which helps the system to attain its efficiency. This can be achieved with help of big data which explains the importance and relation of big data in healthcare sectors.

Value

Planning cost of operation is biggest factor to be kept in mind before performing it. If system takes a time to collect, retrieve , process or analyse any information then it can lead to consumption of time, which leads to wastage of money. As it is said "Time is money". Big data allows to process large amount of data in no time.

In healthcare systems a lot of data is generated from different units. So, it is required to analyse and perform operations on the data in less time. Big data resolve this problem by its efficiency of processing and analysing enormous amount of data in short period of time. And this also helps in saving the cost of processing.

2. Problems with Big Data

In [3] healthcare systems big data have many advantages but with that it also face some problems if we focus on data collection, data security and its usage. Some of the main problems faced in healthcare sectors in terms of big data are:

Security

Information is crucial in any organisation. Information can be used by criminals in many ways which can affect

any individual or organisation privacy. Hackers hack the data and sell it for big amount. So privacy of data is very important. As there can be Patient's data and all its health related data which can be used against that patient by criminal.

Not only in healthcare sectors security is issue in all types of industries. So, before the implementation of big data various protection or data security method should be implemented. Methods like applying Firewall, encrypting the data or any virus cleaning software can be used for security. Along with that regular checkups and monitoring of data can help to achieve data certainty.

Data Classification

Finding a particular data from a such big amount of data is tough because it contain different type of data all together. So it is required to recognized and categorized a data so that big data can work more efficiently. Operation like pooling and contextualizing should be performed so that it become easy for user to find particular data.

Data Modelling

Structure of the data is very important. As makes easy for system to extract relevant information in small amount of time. If model of the data is not defined or if it is not organized then performance of system is compromised. So data should be modelled and structuralised for better efficiency and accuracy.

Data Accommodation

Sometimes in big organisations there is different type of data. So, there is particular set of data which is used by individual unit and is hidden for whole organisation. Hence accumulation of data become way more difficult

Data Personnel

There are very few expert data scientist in static field. Hence finding on for organisation is very difficult task.

Miscommunications Gap

Biggest challenge faced by data scientist in big data is to clearly elaborate their idea to user or clearly understand the view point of user which is also known as miscommunication. In the healthcare systems all the data of various medical units required to put together. So to make more clearer model or effective model time and discussions are required.

Moreover, it creates problems for doctors to search the required information from the database of all the patients or doctor have to search from the start if model is not well formed due to miscommunications.

3. Big Data Applications in Healthcare

As [4] mentioned already there are a number of data sources and tools that are responsible to fetch, collect and store huge data. We have abundant data collected from government agencies, patient portals in hospitals and clinics, research studies, healthcare databases, smart devices (smart phones, smart watches and other wearable devices), electronic data obtained from medical equipments, search engines and other public records (police stations, chemist shops, healthcare applications, social media) to name a few out of a lot of sources.

Big data applications are hence, based on how this enormous amount of data can be managed and subsequently used to obtain useful information, patterns and other benefits. Like we mentioned under the five Vs, the diversity and volume of data and other challenges discussed such as data security and usability, etc., big data applications are developed to overcome these issues to the maximum.

Healthcare model can be developed into a value based model by applying big data tools and characteristics on each application efficiently. Given below are various applications of big data in healthcare.

Diagnostics

We hear a lot in the news about how patients die or suffer from unrecoverable diseases or even catch symptoms of a whole new disease due to incorrect or late diagnosis of symptoms and correct disease. This happens due to lack of up to date devices having correct knowledge of symptoms of different diseases. In order to make a correct judgement, we need information and data related to the disease. Big data here is the data of previously diagnosed patient records, research and study records, etc. stored or fed into the machines and equipments in order to help it make correct judgement in future.

In [1] another way, smart devices help to track symptoms on a daily basis. Devices such as Apple watches, Fitbit, etc. have sensors enabled that help to track heartbeat count, detect sugar level, blood pressure,

breathing issues and other irregularities in order to alert the patient and sends record to the medical team for better diagnosis of the disease.

Hence big data here can be used to detect diseases at an early stage and can therefore be prevented and treated. With the analysis of information obtained from symptoms acquired by the patient, family history, chronic disease, patient's medical history we can decide the best treatment procedure, whether it involves only medication or surgery or requires an operation, side effects of the treatment to be applied, predict the time to recover and if the disease is fully curable or not. Doctors can then go forward with the treatment.

Medicine Recommendation

With the correct diagnosis of a disease it is equally important to prescribe the correct medication. This can be done with the help of predictive data analysis. This analysis is based upon the disease, its symptoms along with the lifestyle of the patients, genetic history of the patient and other social circumstances. These data can be collected and researched upon to create a hyper personalized care routine for the patient to help him cure.

It is also important to keep a record of the prescribed medicines not only for the history of patient but also as a reference for other patients showing same symptoms.

Medical Research

Medical research is a never ending field. As we progress, we encounter new symptoms and diseases and hence, we need to find the cure for it respectively. There is also a need to find new and efficient ways to that are cost effective, more accurate and trustworthy as a cure to existing diseases. The research work never ends and so does the data that is generated from the research. We not only need existing medical data for study but also a large database to store the newly generated data. This data can be converted into a data warehouse and we can perform data mining techniques to extract useful patterns from the study.

Big data ecosystem offers abundant tools to perform research and generate useful patterns and also to visualize the research in the form of graphs and charts to explain the doctors and patients about what the research holds.

Medical research is also used as a study material for students, doctors and other budding researchers as a base of their study.

Predicting Population Health Patterns

Under this application, we take the population as well as the characteristics attached to the population as a whole. Data collected from the population, their surroundings, different geographies and demographics, weather conditions, socio economic conditions, etc. can be used to prescribe better healthcare in that particular area. Every area demands a different healthcare system based on the environment and socio economic conditions. Some areas need more facilities than the other. All this can be determined by monitoring the big data collected for this purpose may be from surveys or government sites.

This also helps in identifying disease trends in those areas and prescribe health strategies to them. In case of a pandemic or disease outbreak, it will help to treat each area as per the demands and the records we have.

Internet of Things

In [5] health care case we can have IoTs in the form of wearable devices or equipments present in the hospital. It not only needs big data to work by obtaining information from it and apply that information to understand the environment where it is placed but also needs a big database to store the data it generates every minute and second. This data is then classified and using big data ecosystem tools, a pattern is generated that makes the data understandable and of use. It is then used by doctors and researchers and also patients to work on related problems. For example, doctors attach small devices in organs such as heart to monitor the heart rate and functioning or sensor devices like a watch that the patient has to wear all the time as it records the functioning and provides feedback to the patient as well as send the real time data to the doctors. The doctors or medical team then monitors the data to see if there is any discrepancy and contact the patient immediately if there is any. Other devices used to collect data can be deployed on various systems are Radio Frequency IDentification (RFID) and Near Field Communication (NFC).

Machine Learning

It is based on big data and is already helping the healthcare sectors to assist them in patient treatment and recovery. There are various predictive algorithms

that help in disease prediction and how to recover. The supervised, unsupervised and semi supervised learning help to create distinction between different symptoms, group symptoms to particular disease, predict related medical treatment and medicines and also help to identify new diseases and predict future adversities of the disease or the treatment the patient is going through. To perform such analysis and prediction, devices require tons of data and generates the same. An example can be how IBM Watson is deployed in healthcare systems and has partnered with big healthcare firms and clinics to work on big data analytics and enhance patient care.

Mobile Devices and Applications

As [6] discussed in other applications, we have various devices like smart phones, smart watches, sensors attached shoes, fitness bands, etc. that are meant for the purpose of keeping us fit. To adapt a fit and healthy lifestyle, people use these devices on a daily basis to keep a track of their weight, calorie count, sugar level, heart rate and metabolism. Healthcare sector encourage people and also recommend them to make use of these devices to be aware of their health. A lot of mobile applications can be downloaded that are meant for the same purpose. This will as a result increase communication between the patient and the medical expert as it helps them to have a direct access to the data generated by these devices. These devices and applications not only help to provide real time health status but also advice healthcare plans and healthy lifestyles to stay fit. The data thus generated is valuable and stored in could platforms and databases to be used in the future whenever required.

Privacy and Fraud Detection

Another [5] application can be handling patient's and hospital's privacy and detecting fraud. There can be a lot of malpractices in medical centres and institutions in order to gain money or for other vicious purposed. Threats can be insiders or outsiders. Therefore it is important to maintain the security of the patient as well as the hospital. Big data here can be used to derive a pattern of treatment involving its attributes to check whether the treatment is performed as required. It ensures the diagnostics is correct and there is no alteration to the actual medical record data. It keeps a track if any medical staff is indulging in any unfair means by performing regular audits, compliances and checks on each department and at an individual level too. the tools and other security devices installed generate security data on a continuous basis that is regularly checked to maintain fair practices.

Image Processing

There [7] are a lot of medical devices that generate reports in the form of images. To name a few we have Computer Tomography scan (CT scan) reports, X-Ray and MRI reports, ultrasound, fluoroscopy, mammography, etc generating reports all in the form of images. These images are sometimes hard to study. It is very important to read and identify what the images have to offer in order to make correct predictions about the diseases. The image data falls in a large size range and since it is generated on a daily basis, the data collected in the form of images is very large and requires a large database capable of storing hundreds of megabytes of not textual data. This also demands algorithms that are accurate in studying and predicting the diseases by analysing the images. Big data ecosystem offers a variety of tools to perform analysis and for storage.

4. Future Scope

Since [8] big data generation and analytics is an essential part that is used in healthcare system we can in future expand its functionalities. Big data is generated from a variety of sources both textual and non textual and also involves digital and electronic data also. This also demands the use of integrating various devices and tools generating and storing big data. Since big data tools and platforms are not used in majority of healthcare centres, in future, more people must be trained in order to utilise these platforms. Hadoop and Map Reduce are data processing platforms that involve commodity hardware and perform parallel processing and hence, can be made use of. The platforms should be easily available and must not be of higher cost. These platforms must be modified in order to decrease cost and processing time.

5. Conclusion

In the paper we saw how tremendous amount of data is generated and utilised in healthcare sectors to assist the medical experts and benefit the patients. Handling and processing large data has always been an issue which is solved using big data tools and applications. There are a number of applications involving big data and challenges that are encountered while we work with big data all of which has been discussed.

References

[1] Dr. P.Kamakshi, "Importance of Big data in Healthcare System-A Survey", *International Journal of Applied Engineering Research*,vol.13, issue 15, pp.12184-12187

[2] Revanth Sonnati, "Improving Healthcare Using Big Data Analytics", "*International Journal of Scientific & Technology Research*", vol. 6, issue 03, March 2017

[3] Quek Kia Fatt, Amutha Ramadas, "The Usefulness and Challenges of Big Data in Healthcare", *Journal of Healthcare Communications,* vol. 3, issue 2:21, February 2018

[4] Healthcare Big Data and the Promis of Value-Based Care. [Online]. Availabe- https://catalyst.nejm.org/doi/full/10.1056/CAT.18 .0290#

[5] Sabyachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, Sandeep Kaushik, "Big Data in Healthcare: management, analysis and future prospects", *Journal of Big Data,* Article no. 54, 2019

[6] Md Saiful Islam, Md Mahmudul Hasan, Xiaoyi Wang, Hayley D. Germack, Md Noor-E-Alam, "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining", *Healthcare (Basel)*, 6(2):54, May 2018

[7] Ashwin Belle, Raghuram Thiagarajan, S.M.Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, Kavyan Najarian, "Big Data Analytics in Healthcare", *Journal of Biomedicine and Biotechnology,* January 2015

[8] Blagoj Ristevski, Ming Chen, "Big Data Analytics in Medicine and Healthcare", *Journal of Integrative Bioinformatics",* 15(3):2017003, September 2018