

‘Applications of Big Data Analytics in Healthcare’

Mr.Sameer Baban Kamble

Lecturer, Dept. of Information technology, Patkar - Varde College, Mumbai, India

Abstract—‘Big Data’ as the name suggests, it is a huge amount of data that is yottabytes of data which is structured, unstructured or semi-structured in nature. This data has a huge volume, as well as requires huge velocity to transmit over the network; also the data has a lot of variety for e.g. numeric data, audio/video data, social media data etc. This paper contains the information about healthcare industry, the fields in healthcare industry using the concept of big data analytics. Also the paper tells about how the data is managed using big data with Hadoop as well. Big data is used in healthcare industry to predict the outcome of diseases and also to find out the rate of disease spread among people. It has many more applications which are mentioned in this paper.

Keywords—Bigdata, healthcare, Hadoop, map-reduce, data, predictive, prescriptive, etc.

I. INTRODUCTION

Big data is now-a-days used in different sectors such as telecom, airways, education, banking, healthcare, government sectors etc. Big data as the name suggests is the set of data that is huge in size(volume), transferred or generated with a very huge velocity and has a lot of variety in it.

Generally big data is a large set of information that is generated on daily basis. This data being generated, keeps on increasing every time. This data is collected by various organizations or even data scientists and then analysed to have a better outcome.

There are many sectors that collect large amount of data generated or provided by their users. This data is then analysed to have a better outcome of their business. Not only in business, but also in other sectors such as space where scientists keep a track of every movement taking place in the space by tracking the co-ordinates of their satellites, capturing images of other planets with the help of satellite which acts as data for the scientists which is then analysed by them to reach a particular conclusion as well as carry out their research.

Many more applications of big data are: banking and security, communication, media and entertainment, education, manufacturing and natural resources, government, healthcare, etc.

Healthcare industry uses big data to analyse the patients, collected ta about various symptoms and then find the disease they are infected with, tracking the spread of disease and so on.

Following are some of the technologies used in big data analytics:

A. *Analytics for prediction* [1]:

A technique that is used by companies to analyse, predict and take decisions based on the data available with the help of models used for prediction.

B. *No Sql databases*[1]:

A database used to store data which does not have a fixed form.

C. *knowledge discovery*[1]:

A method which searches and extracts knowledge and information from huge repositories of data to use for analysis and predictions.

D. *Stream Analytics*[1]:

This is a process where analysis of data coming from live sources is done. E.g. YouTube data is the source of data which is a live stream for analysing the pattern, type etc. of the channels or choices of the user and based on that, the user is sent recommendations to watch more similar channels.

E. *Distributed file system*[1]:

A method used to store data on multiple places so that it is easily accessible from anywhere. In case of data loss from one database, the data can be accessed from another database which is located at other location.

Following is the overview of areas of big data analysis in healthcare:

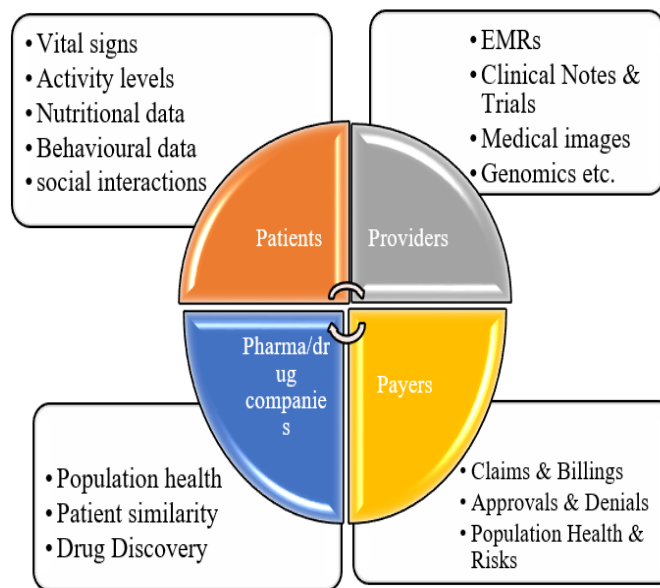


Fig.1.Areas of big data analytics in healthcare [10]

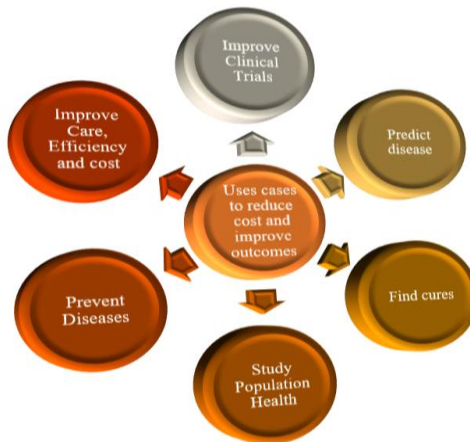


Fig.2.Improving Outcomes and a few more use cases[10]

II. STEPS FOR DATA ANALYSIS

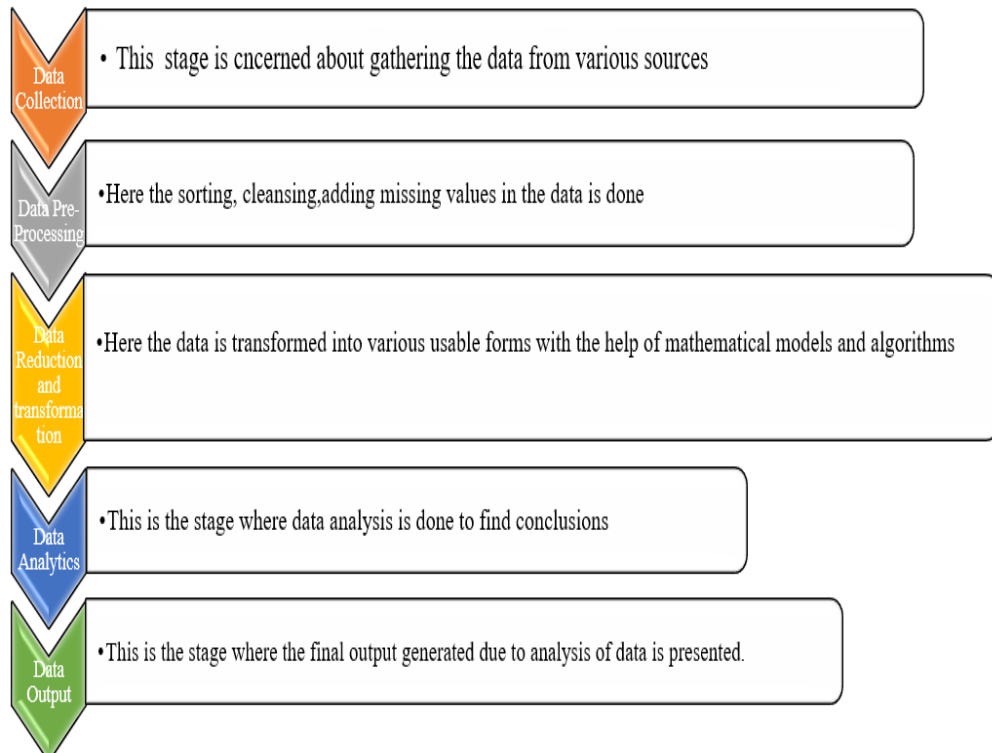


Fig.3. Data analysis steps[11]

III. APPLICATIONS

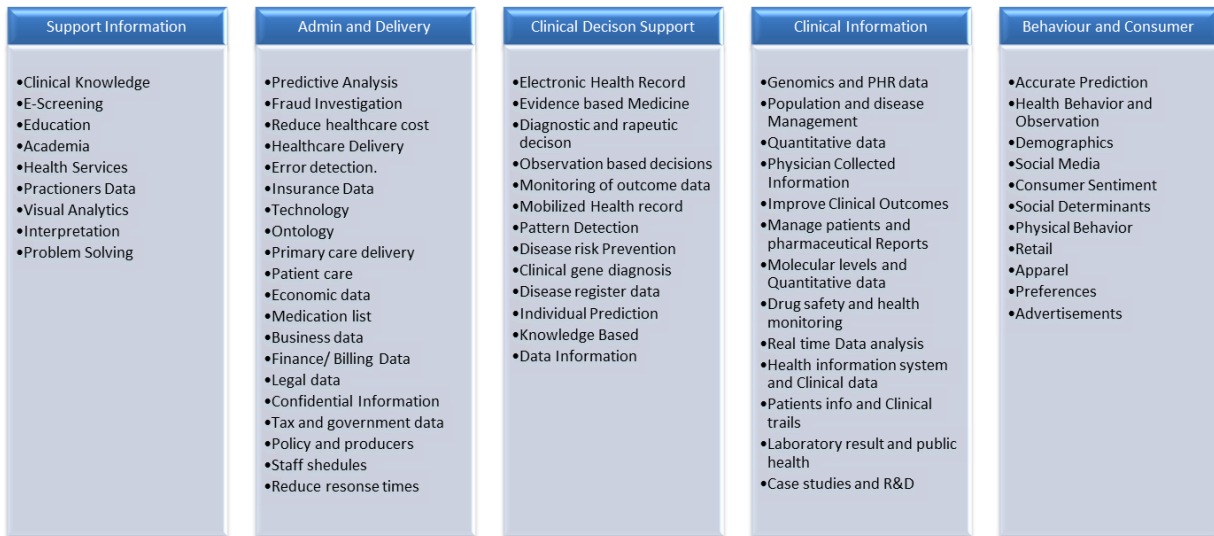


Fig.4.Categorization of uses in healthcare[6]

Following are the few primary areas to improve healthcare: Predictive modelling for risk and resource use, Population management, drug and medical device safety, disease and treatment, public health etc.

A. *Clinical decision support* [2,9,11]:

Here the decisions are taken using descriptive and predictive analysis. The data that is used here comes from the hospitals, clinics etc. Following are some of the clinical areas that use big data analytics:

1) *Cardiovascular Disease*[2]:

This is basically the heart disease. ECG is done to gather information regarding the heartbeat which is commonly used in this case. This ECG technique helps to gather data about blockages in the heart and then accordingly the doctors can treat the patients.

2) *Diabetes and Cancer*[2]:

Prediction of diabetes and cancer can be done with the help of big data analytics where the data about the increase and decrease in the rate of sugar level in a diabetic patient can be monitored by collecting data on frequent basis. Machine learning is used in case of cancer patients to predict the time period of their survival.

3) *Emergency Care*[2]:

This department keeps a track of emergency cases that arrive in the hospitals to analyse the disease if there is an outbreak as well as predict the number of patients that can arise and make arrangements accordingly.



B. Electronic health records[4]:

This is the place where all the records of the patient such as medical reports, symptoms etc. are stored and shared digitally. Also, modification of these records can be done easily without the need of paper.

C. Hadoop based applications[4]:

1) Monitoring of patient vitals:

Here the different hospitals worldwide keep on monitoring their patients' data and convert the unstructured data into a structured form to analyse the data for further results. This data is stored using HDFS(Hadoop Distributed File System)

2) Prevention and detection of frauds:

HDFS is used to avoid frauds such as illegally changing or deleting records from the databases, medical claims etc.

D. Image processing[5]:

Image processing is the technique used to gather data in the form of images which is done with the help of MRI (Medical Resonance Imaging),CAT(Computerized Axial Tomography),CT Scan etc.

E. Infection prevention, prediction and control[7]:

As the name suggests, this is the way where large amount of data is collected about the patients, this data is analysed to know what kind of disease is spreading, how it can be stopped or controlled and to what extent can the disease spread and through what mediums it can spread.

F. Disease management[11]:

The management of a particular disease i.e. stopping the spread of a disease and finding solutions to cure the patients with that particular disease with the analysis of available data comes under the concept of disease management.

G. Patient Matching[11]:

The data regarding existing diseases is stored prior. Whenever a patient arises, the patients' symptoms are matched with those in the database and accordingly the treatment is given to the patient.

H. Real-time alerting[13,15]:

This is based on collecting data at real time. In healthcare, a software known as the clinical decision support system is used to take right decision by collecting data at real-time. This is a cloud based system that collects the data about the patients' health and sends it to the cloud immediately. This data can be accessed by the doctors. The doctors of the particular patients receive an alert about the patient which helps them to treat the patient quickly.

I. Telemedicine[13,15]:

This is the technology which is used to monitor patients who are far away meaning that this technology is used for distance monitoring of the patients. As a result patients can get advice from their doctors anywhere. This is also helpful in saving time to treat the patients. Also patients who cannot go to their doctors or even the

doctors cannot visit the patient due to their busy schedule or due to large number of patients can use this to monitor their patients.

J. More accurate treatment[14]:

Big data helps collecting data on a large scale and analysing them. This helps in better understanding the data gathered and hence helps to take the right decision. In case of healthcare, the data gathered on a large scale helps the doctors to analyse and take the right decision about what treatment is to be provided to the patients. This treatment is more accurate as it is based on the analysis of large amount of data.

K. Strategic planning using health data[15]:

Here the data about demographics, population count, check-up details, increase in number of patients of a specific disease etc. is collected and analysed. Based on this the decisions such as building new hospitals, finding newer medicines to cure diseases faster and in a better way etc. are taken.

L. Heart attack prediction[15]:

Big data is also used for prediction of heart attack. The data regarding patients' living habits, what the patient eats, heart rate, blood pressure count ;all these sources are useful to gather the data to predict the chances of heart attack.

M. Nutrition management[15]:

Here the data about heart rates, weight, eating habits are observed and then the suggestion for proper nutrition is given to the patient based on some of these factors. One of the examples is the diet plan which is followed by most of the people. Here doctors or dieticians collect the data about patient's eating habits, what kind of food they eat and at what time and accordingly suggest a diet plan or a nutrition plan.

N. Ophthalmology[15]:

Here AI is used to generate diagnosis report where big data is used to evoke Artificial intelligence to do this job. Artificial intelligence makes use of an image by capturing and processing the image to analyse it and take the right decision about the medication or treatment can be easily done.

O. Tracking arthritis[15]:

Big data is now-a-days used to track the arthritis in a patient. Different data such as demographics, when the patient visited the hospital, vital signs, treatment history etc. Based on this data, the respective type of arthritis is predicted i.e. rheumatoid or periodontal.

IV. ADVANTAGES

Following are some of the advantages of Big data analytics [3]:

- Disease Prevention
- Identification of risks
- Designing of solutions to overcome the disease with the analysis of available data



- Delivery of medicines
- Health analysis of population
- Fraud detection and prevention and many more.

Following is the impact of big data on healthcare [4]:

- 1) *Living*: Concerned with the better living of a patient.
- 2) *Care*: Taking proper care in order to overcome the disease
- 3) *Provider*: Collecting data from various sources and analysing the data to provide right treatment to the patients.
- 4) *Innovation*: New medicines and treatment need to be invented in order to overcome newer disease arising in day to day life.
- 5) *Value*: Good quality of treatment and services need to be given to the patients to avoid the spread or increase of disease among people.

V. CHALLENGES

Following are some of the challenges faced by big data analytics in healthcare:

1) *Security and privacy*[8,9,12]

This technique is useful for a small set of data, but when it comes to big data which a huge volume of data, security and privacy needs to be taken care of because as the size of data increases, the risk with the data increases and there are larger chances of data being tampered. It is necessary to keep patient data secure and private to avoid any consequences.

2) *Data quality*[8]:

Data quality needs to be maintained in order to take proper decisions and predict the right disease from the symptoms.

3) *Insufficient real-time processing*[8]:

Real time data needs to be processed at a faster rate with more accuracy in order to give proper treatment to the patients.

4) *No fixed standards for healthcare data*[8]:

Data about healthcare is at times collected from different agents such as notes from medical students, images, sensors. There is a fixed standard which makes it difficult to process the available data.

5) *Heterogeneity*[9]:

Data collected in healthcare is of various forms such as structured, unstructured, semi-structured and arises from various sources which makes it difficult to process the data as there are limited tools to process them.

6) *Incompleteness*[9]:

In case of healthcare data cannot be stored constantly because of huge amount of data being generated. At times these records are not in a fixed format and therefore storing the data in same format is impossible and so the data is at times not stored in the record which leads to incomplete data.

7) *Timeliness and longevity*[9]:

The time period for storing data related to MRI, ECG etc. is very low. This time period should be increased to make data available in future as well for prediction and analysis.

8) *Ownership*[9]:

Patients' data is stored generally in hospitals, clinics, labs, doctors, etc. But Patients' accessibility towards their own data is restricted many times due to which patients themselves are not able to keep track of their health.

9) *Data mining and storage*[9]:

Data obtained from the clinics is in the form of hand written documents, unstructured form and is available on a large scale which makes it difficult to gather, analyse and process the data. Also data storage becomes difficult because of high cost required. Also the size of data is huge due to which huge storage space is required may it be physically or digitally which is very costly in case of medical data.

10) *Data sharing*[9]:

This is another challenge as sharing medical data all over is difficult as the data is stored over a wide range in various hospitals, clinics etc. and the type of data being stored is of various types such as image form of data, medical cost records, diseases that patients are affected with etc.

11) *Data classification and modelling*[12]:

Classification of data is necessary as data is in various forms such as structured, unstructured and semi-structured. This is necessary for proper processing of data in order to have proper outcome to take proper decision as well as to use the right model to solve the problems about diagnosis of diseases.

VI. CONCLUSION

Big data analytics has spread over a wide area not only in healthcare but also in other sectors such as education, healthcare, government etc. All these sectors have evolved a lot due to big data analytics. Health-care industry has also benefited a lot by the use of big data. Prediction of diseases, taking decision regarding treatments, treating patients from a long distance, collecting a large amount of data about the spread of a disease based on which predicting the spread of the disease or predicting to what extent the disease might spread further etc. can be easily done.

Also with the help of data collected, it is also possible to prevent a disease or stop the disease from spreading further at the right time can be done in order to save the lives of the patients. As per the saying 'prevention is better than cure' big data analytics helps in making this saying come true in real life may it be in case of healthcare or any kind of business or even in education and other sector.



REFERENCES

- [1] "A Detailed study of Big Data in Healthcare: Case study of Brenda and IBM Watson", Dipesh Jain, Vivek Kumar, Darpan Khanduja, Kamlesh Sharma, Ritika Bateja, International Journal of Recent Technology and Engineering (IJRTE), Volume-7 Issue-5, January 2019.
- [2] "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining", Md Saiful Islam, Md Mahmudul Hasan, Xiaoyi Wang, Hayley D. Germack and Md Noor-E-Alam, 23 May 2018.
- [3] "Application of Big Data Analytics: An Innovation in Health Care", Swarna Latha Garapati, Dr. Sridevi Garapati, International Journal of Computational Intelligence Research, Volume 14, Number 1, 2018.
- [4] "Big Data Analytics for Healthcare Industry: Impact, Applications and Tools", Sunil Kumar, Maninder Singh, BIG DATA MINING AND ANALYTICS, Volume 2, Number 1, March 2019.
- [5] "Big Data Analytics in Healthcare", Ashwin Belle, Raghuram Thiagarajan, S.M.Reza Soroushmehr, Fatemeh Navidi, Journal of Biomedicine and Biotechnology, January 2015.
- [6] "Big data in healthcare: What is it used for?", Rebecca Hermon, Patricia A H Williams.
- [7] "Big Data Technologies in Healthcare- Needs, opportunities and challenges", TF7 Healthcare subgroup, 21/12/2016.
- [8] "Big Data Analytics in Healthcare Systems", Lidong Wang, Cheryl Ann Alexander, International Journal of Mathematical, Engineering and Management Sciences, Vol.4, No.1, 2019.
- [9] "Big Data in Health Care: Applications and Challenges", Liang Hong, Mengqi Luo, Ruixue Wang, Peixin Lu, Wei Lu, Long Lu, Data and Information Management, 2018.
- [10] "BIG DATA IN HEALTHCARE", Prashant Dhamdhere, Jeremiah Harmsen, Raaghav Hebbar, Srinath Mandalapu, Ashish Mehra, Suju Rajan, Berkeley, university of California.
- [11] "Improving Healthcare Using Big Data Analytics", Revanth Sonnati, international journal of scientific & technology research volume 6, issue 03, march 2017.
- [12] "The Usefulness and Challenges of Big Data in Healthcare", Quek Kia Fatt and Amutha Ramadas, Journal of Healthcare Communications, iMedPub Journals, Vol.3, No.2:21, February 13, 2018.
- [13] Examples of big data analytics in healthcare that can save people available at <https://www.datapine.com/blog/big-data-examples-in-healthcare>
- [14] Areas where big data is transforming right now available at <https://hitconsultant.net/2019/09/09/4-areas-big-data-transforming-healthcare/#.X13KrVPhWyU>
- [15] Examples and applications of big data in healthcare available at <https://www.ubuntupit.com/examples-and-applications-of-big-data-in-healthcare/>