ISSN No:-2456-2165

The Technical Advancements of Big Data Analytics in Healthcare: a Survey Paper

Sumit Jain
Department of Information Science and Engineering
School of Engineering and Technology, Jain University
Bengaluru, India

Sahil Mehta
Department of Information Science and Engineering
School of Engineering and Technology, Jain University
Bengaluru, India

Soumya K N
Department of Information Science and Engineering
School of Engineering and Technology, Jain University
Bengaluru, India

Abstract:- The availability of cheap technology has resulted in millions of people possessing handheld and other devices. When millions of people use these devices trillions of Petabytes of data be it structured, semi-structured or unstructured is regenerated which in today's generation is called Big Data. A huge chunk of this data belongs to the healthcare center which is a very imperative aspect in today's life. A wide variety of challenges are faced by the healthcare industry on a daily basis which includes locating , analyzing and supervising vital information about patients to lead them a healthier life by considering diseases , therapies and to an extent predicting outcomes at an earlier stage and make real time decisions Our main agenda here is to highlight the potential Big Data Analytics offers in healthcare and explore how it makes advancement to treatments and provides an edge to patients, doctors and scientists. Our agenda also does describe about various key technologies of big data and what it has to offer in gathering huge chunks of data to have a deeper understanding of people's routine like their daily diet or the amount of exercise they do to predict outcomes. The following allows doctors and patients to draw solutions in preventing and curing life threatening diseases at an earlier stage thus promoting the patient's health and simultaneously reduce costs by getting immediate and best treatment.

Keywords:- *Big Data; Health Care; Analytics; EHR.*

I. INTRODUCTION

It is very vital for every human being to remain healthy to sustain and be energetic throughout the day. Healthcare is an immediate answer for diseases, injuries and other physical and mental deficiencies in the human body via diagnosis, prevention and treatment. Healthcare is an important practice around the globe and is highly looked upon as an important prospect. Doctors and their associates are a part of these health professionals. Along with them are dentists, nurses, pharmacists who too add on to the cause of healthcare.

The increase in variety of diseases and number of injuries alongside the ever increasing population has caused a desperate need for healthcare services. The population wish to live longer and healthier which has caused an increment in healthcare demand which in turn has affected the country's economy. For example, in 2012 the healthcare industry consumed an average of 9.3% of the GDP per capita across the 34 members of OECD countries. A few of the countries include United States, United Kingdom, France, Germany, Canada and Switzerland. As you can see in the figure below are most of the countries part of OECD along with the global leader in healthcare spending being none other than the United States spending 17% of its GDP.

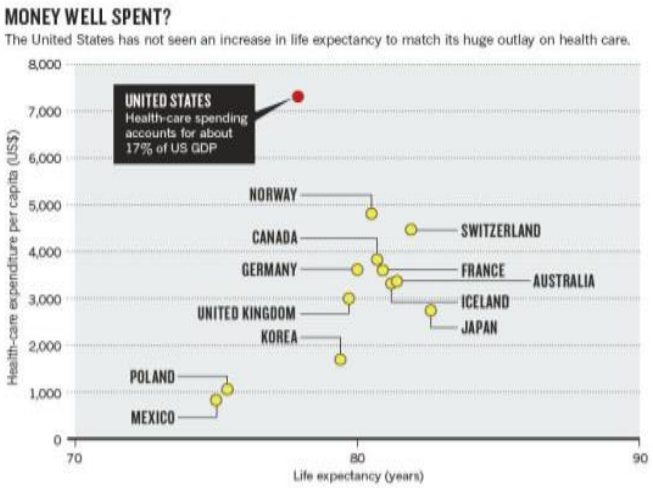

Fig 1:- Money Spent on Healthcare

As the need for healthcare doesn't seem to diminish, what we generally tend to ignore is the amount of irrelevant data that is generated. A single meeting with a doctor generates massive amounts of data which we tend to ignore because our sole purpose to visit a doctor is to tackle the problem at hand which causes us to neglect any upcoming disease which we might incur. All this data, collectively is called as big data. The relevant content is termed as the structured data and the irrelevant is called the un-structured data.

So, big data are data sets that are compound and capacious which cannot be analyzed and handled by the traditional application software. The four dimensions or V's of Big Data as shown below include volume, velocity, variety and the lately added veracity. Big data exists in various forms depending upon the corporations it is used by. For example, a small company with low processing power could consider about 100 GB as big data while a massive company with specialized equipment and high processing speed would not even consider 100 TB as big data hence the definition of big data varies upon its existence.

ISSN No:-2456-2165

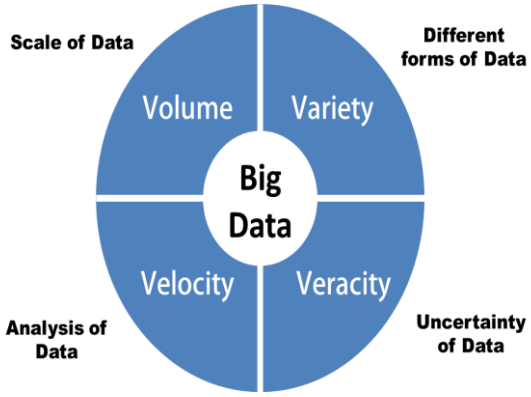

Fig 2:- Elements of big data

The massive amounts of data including unstructured data have significant importance if analyzed properly. A few of the challenges that big data provide are collecting data, storing data , analyzing data , querying and updating data. Lately, big data offers predictive analysis, behavioral analytics and other advanced analytics hence the term big data analytics has come to rise. Analytics of Big data allows doctors to make informed decisions as it analyses current and previous records to identify hidden patterns, trends and other information which could help in making accurate decisions.

The stress levels are booming in healthcare as data generation are off the charts. Under such situations the data must be immediately analyzed for quick and correct decisions. Big Data Analytics along with their advanced analytical tools provide an edge to the healthcare industry by providing real-time decisions which not only tackle the problem at hand but also analyze all the other vitals of a patient (Sugar level, RBC/WBC/platelet count etc.) for any upcoming problems the patient might face. Big data analytics, with their major application being predictive analysis prepare the patient for any upcoming mishap and allow the doctors to prepare a regime to prevent any diseases faced in the future.

II. BACKGROUND

With an ever increasing rate of data generation and reduction in time consumption, big data analytics is in need of high processing speeds and efficient resources. The initial stages of Big data analytics involves the generation and collection of data from various data resources like heart rate monitors, blood work etc. Which is then stored and processed using various tools to provide relevant data which could be used for predictive analysis.

As the popularity of big data analytics has been increasing terms like digital health, m-health, e-health, e-medicine, telemedicine etc. which are associated with collection analysis and applications of Big Data in health have become very common around. At first, big data was only used in terms of medical and clinical research by scientists but as time progressed it was introduced into practice of public health. As technology prevailed, data generation increased causing the once used resources not being sufficient enough which called out for technical advances to improve healthcare. Stakeholders, investors now prefer financing for better and reliable analytical tools thus opening up to opportunities big data analytics has to offer not to just an individual but to a greater population.

The European Council concluded their open, data-intensive and network research by signifying the importance of Big Data in Health if we truly harness the potential it has to offer and also subjected it to be a key to faster and wider innovation. Not only will big data improve the health of an individual, it will also change the performance and outcomes of the existing healthcare system.

The existing system involves data generation in the form of diagnosis and monitoring of patients vitals. Not all the information that is collected is in the same format thus making it necessary to rely on a different platform for recollection of all data in similar format which could cause inaccuracy. For production of personal medications the health record of only that individual is not sufficient enough, health information of a bigger population gives a greater perspective and allows relevant estimations. A combination between population data and an individual's vital measurements provides exciting openings for the implementation of personalized medications.

There are various sources which are referred for data generation in healthcare. A few of them include:

- EHR (Electronic Health Record)
- Social media (Facebook, Twitter)
- Genomic Data (Hereditary diseases)
- Pharmaceutical data (medicines that are prescribed)
- Insurance claims
- Telemedicine, mobile apps and sensors (Heart rate monitors)
- Other sources (income statistics, environmental databases etc.)

Possible data sources for Big Data are constantly changing and evolving, therefore, the list above cannot be regarded as complete. Also the combination of data sets from various sources increase complexity as data is not homogeneous but also gives rise to new possibilities for analytics.

These issues have led to the concept of data fusion becoming popular and widely used. Also research in big data and big data analytical tools have become even more significant in the last few years. This requires powerful computational techniques to recognize hidden patterns within massive data that is generated in the healthcare domain. It is a vital part to identify and create ways of systematic approach to manage, integrate, analyze, and interpret such large complex data sets.

Big data use in healthcare has already gained importance but is still increasing at a great velocity. Applications vary from a prospective data monitoring point of view to a retrospective data analysis point of view and may contribute to:

- Improving quality of treatments by e.g.:
- Earlier disease diagnosis,

ISSN No:-2456-2165

- Verify prescribed medications to allergies to eliminate negative reactions from medicines,
- Less medical errors,
- Determination of casualties, understanding of co-morbidity,
- Cross-linkage of health care providers and professionals,
- Intensification of research networks, and
- Fusion of different networks such as social networks, disease networks or medicine networks.

- Understanding the factors causing the disease thus helping in prevention at a population, subpopulation and individual level by providing solutions for betterment of the patient like a change in diet , weather etc.
- Information generated from patients allowing to create patient safety drills and take informed decisions
- Prediction of outcomes, e.g. Containment and improvement of chronic-diseases, global infectious disease surveillance through evolving risk maps and better understanding of demographic challenges and trends as well as disease transmission pathways,

Knowledge dissemination, e.g. Help physicians to stay current with the latest evidence guiding clinical practice

With an ever increasing growth in population on our planet it has resulted in an exponential spur of data generation in the healthcare sector. Along with that healthcare techniques face a transition period where in performance and result have become a critical aspect in today's healthcare environment. A primary motivation for each healthcare institute as we know it is to protect its fellow patient's life although we cannot deny it that profit is not and should not be a primary motivation. It is equally important for healthcare institutes to attain the accessible tools, infrastructure, and techniques to utilize Big Data effectively or else face a devastating loss of millions.

Healthcare, presently analyzes patient characteristics and calculates cost and outcomes of care using Big Data Analytics to identify the most clinical and cost effective treatment for the betterment of the patient. It utilizes its advanced analytic tools on existing patient profiles to categorize individuals who would benefit from preventive care or lifestyle changes. Big Data Analytics have also made a major impact on tackling large scale or terminal diseases by preempting them through studying the patient profiles using advanced analytic tools. It identifies the disease and provides prevention initiatives thus supporting the patient in shaping the care protocols that offer the best value of protection and prevention. Many sponsors are have deployed mobile applications that help patients manage their care, locate health centers for better treatment to improve their health. With the help of Big Data Analytics these sponsors have achieved the ability to monitor adherence to drug, treatment programs and identify movements that lead to individual and population wellness benefits.

Analytics of Big data is applied for three of the major aspects in healthcare that is diagnosis of diseases, treatment of those diseases and ensuring less or no cases of re-admission for the patient. A few examples to depict the following along with the software's used are mentioned below:

A. Diagnosis

IBM's PureData Solution is an integrated software that aids medical organizations all around the globe to use healthcare analytics. It provides rapid results while minimizing risks thus making it the most common software being utilised. Seattle's Children's Hospital, with the help of IBM's PureData Solution diagnose about 350,000 patients. This system detects similarities between patients, then locates which treatments have been effective on the patient who had suffered from the same, takes into consideration the existing patients vitals and maps a treatment which would provide great results. The system refers to data from various sources like the hospitals database, social networks etc. also verifies the data-base history of the hospital to locate additional resources which could help in providing faster and accurate results.

B. Treatment

The DNA of a patient is an important piece of information for big data analytics in healthcare. Along with the patient's vitals, the analytical tools use the patients DNA to chart various treatments which would prove effective against terminal diseases like cancer. Intel and Oregon Health and Science University developed a platform called Collaborative Cancer Cloud that allows various hospitals to upload and share patient records which can be utilized by any doctor to improve treatment with cancer. It also giver liberty to doctors to compare a particular cancer patients DNA with another patient to take an informed decision and improve the patients chances of survival. The Analytical tools locate trends and patterns between various patients and then provide the doctors with solutions for better treatments.

C. Readmission

Intel and Cloudera have utilized predictive analytics to revolutionize healthcare by producing booming results where-in they have reduced readmissions by 6,000 people which saves around $4 million in Medicare penalties and $72 million in medical service costs. Piedmont Healthcare's Patient First has led to a 50% improvement in the length of stay of cardiovascular patients, a 10% reduction in readmission rates of heart failure patients and a 12% reduction in readmission rates in heart attack victims, while saving more than $2 million in consulting services

III. TECHNOLOGY SUPPORT FOR BIG DATA ANALYTICS IN HEALTHCARE

D. Terminology

The big data analytics technology is a combination of several techniques and processing methods which collectively are effective to obtain useful results for strategic management and implementation. Here are some key technologies that enable big data for businesses.

ISSN No:-2456-2165

Predictive Analytics	It processes Big Data to discover and evaluate diseases which helps doctors and hospitals to predict what is to come and to come up with solutions to the problem
Stream Analytics	Data of a particular patient is stored in various hospitals and in multiple formats. Stream analytics software's filter, aggregate and analyze this big data. It also allows connection to external data sources and their integration into the application flow.
Distributed Storage	Distributed file stores contain replicated data to counter loss or corruption of data.
Data Integration	Most Hospitals face a challenge in processing terabytes of data in a way that can be useful. Data integration tools allow hospitals to streamline data across a number of big data solutions like Apache Hive, Apache Pig, Apache Spark, Hadoop etc.

Table 1. Key Terminologies

E. Example

An electronic health record (EHR) is a digital version of a patient's paper chart. EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users. EHRs can:

- Contains all the records of a particular patient which includes details about all his visits to the hospital, the medications given to the patient, what diseases has the patient been diagnosed with, treatments given, immunization dates (polio, chicken-pox), X-rays, MRI's etc.
- Provides access to doctors examining the patient to make accurate decisions for the betterment if the patient's health.
- Systematize and update provider workflow

One of the major applications of an EHR is that all of the patient's information is stored in a digitized format thus being accessible to anyone across the globe. It can accessible to more than one person at a particular time. EHR's are built to store information which can be shared between healthcare providers and organizations like laboratories, hospitals, pharmacies etc. and also they can updated on a day to day basis.

As of today, EHR's are significant in terms of healthcare. They contain the smallest of details of a patient ranging from birth to death, from allergies to terminal diseases. A secured information system is created between healthcare providers and organizations of both the public and private sector which allows them to share patient records. From time to time, patients wish to see other doctors then these doctors can make changes to the existing record of that patient thus keeping track of the patient's health at an up-to-date level which also eliminates the worry of data replication.

As the EHR contains an entire record of the patient, with the help of predictive analytics it can issue warnings for a particular patient to get a few tests done to track the patient's health and determine whether it has improved or deteriorated. A US based research collaborative, Optum Labs has collected EHRs of over 30 million patients and created a database that uses predictive analytic tools for better treatment of patients.

Their goal is to provide doctors with tools that allow them to make big data-informed decisions within seconds to improve patient's treatment. This particularly gives an edge in case of patients with complex medical histories, suffering from multiple diseases. New analytical tools have allowed doctors to predict, for example, who is at risk of cancer, and thereby create new diet and exercise plans to minimize the risk.

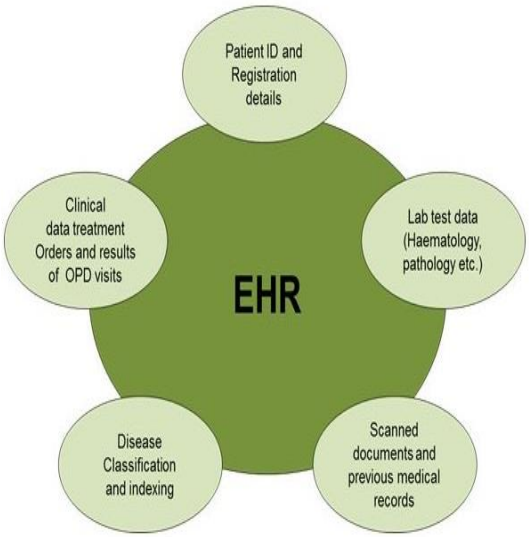

Fig 3:- Composition of HER

IV. BENEFITS AND CHALLENGES TO HEALTHCARE

F. Benefits

Major healthcare such as single-physician, multi-provider groups, some of the large scale hospitals and non-profit organizations may get significant benefits if they efficiently start using big data.

Some of these important benefits can be:

- *Cheaper Healthcare*

Big data Analytics has a great impact on the cost of medical treatment. The analysis of patient data provides an awareness to health care providers about spreading of dangerous diseases. The analytics allows the doctors to take initial steps in the prevention of initial stages of the diseases.

- *Increases Awareness*

Data analysis of a patient provides an insight to be more aware of their health and take control over it. The information sharing mechanism increases productivity and reducing overlapping of data. By thus, it is enhancing the coordination of care.

- *Special Treatment for each individual*

By analyzing the genetic blueprints of a particular patient, Big data analytics determines the tests and treatments needed for each individual. The provision of earlier treatment can reduce the health costs and can eliminate the risk of chronic diseases.

- *Prevent diseases from growing*

As our elders say Prevention is always better than cure. Following this notion with the help of Big Data analytics, it is easy to capture, analyze and compare patient symptoms earlier to offer a preventive care in a better way.

G. Challenges

Big data and its analytics does improve the performance of medical organizations but it also brings rise to new problems which are mentioned below:

- The sources of data vary from hospital to hospital thus causing each database to be of a different format. To utilize all the data from each hospital a single data warehouse must be created where all that data must be stored and managed in a single format thus being very costly.
- A major issue that we usually come across is the quality of data. Earlier, data was stored in records but even after digitization not all data is structured and standardized. It causes industries to put in extra effort in the conversion of data into meaningful data.
- Not all hospitals are keen on sharing their data with companies who wish to use big data analytics. Also, companies need to invest on reliable data analytic tools and other resources a data scientist would require.
- Big data analytics is still a growing profession thus causing very few analysts to have a great knowledge about analytical tools which makes it an expensive affair for companies who wish to hire them.
- It is difficult to check doctor performances as doctors prefer to send their high risk patients to their colleagues thus ensuring their success rate is not affected.

V. CONCLUSION

Big Data and Big Data Analytics have made a major impact on healthcare and caused it to be more effective and efficient. They can be used for a range of operations, from disease management, identification and prevention to medical research, and lead to insights which support healthcare providers in making more timely and informed decisions about the population they are managing. It has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. In the future we'll see the rapid, widespread implementation and use of big data analytics across the healthcare organization and the healthcare industry. As big data analytics becomes more mainstream, issues such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies will garner attention. Big data analytics and applications in healthcare are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process.

REFERENCES

[1]. Sanskruti Patel and Atul Patel "Big Data Revolution In Health Care Sector: Opportunities, Challenges And Technological Advancements", International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016.

[2]. Shankar Krishnan "Application of Analytics to Big Data in Healthcare", 2016 32nd Southern Biomedical Engineering Conference.

[3]. Shaila H Koppad and Dr. Anupama Kumar "Application of Big Data Analytics in Healthcare System to predict COPD", 2016 International Conference on Circuit, Power and Computing Technologies .

[4]. Manpreet Singh, Vandan Bhatia and Rhythm Bhatia "Big Data Analytics Solution to Healthcare", 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT).

[5]. Van-Dai Ta, Chuan-Ming Liu and Goodwill Wandile Nkabinde "Big Data Stream Computing in Healthcare Real-Time Analytics", 2016 IEEE International Conference on Cloud Computing and Big Data Analysis.

[6]. Raghunath Nambiar, Adhiraaj Sethi, Ruchie Bhardwaj and Rajesh Vargheese "A Look at Challenges and Opportunities of Big Data Analytics in Healthcare" 2013 IEEE International Conference on Big Data.

[7]. A.Rishika Reddy and P. Suresh Kumar "Predictive Big Data Analytics in Healthcare", 2016 Second International Conference on Computational Intelligence & Communication Technology.

[8]. Mimoh Ojha and Dr. Kirti Mathur "Proposed Application of Big Data Analytics in Healthcare at Maharaja Yeshwantrao Hospital", 2016 3rd MEC International Conference on Big Data and Smart City.