Data Science: An Intelligent platform for healthcare Assessment and Future Diagnosis

S.Babu

Department of Computer Science and Applications, SCSVMV

Abstract - *Data science is the study of interdisciplinary field which extracts hidden knowledge and meaningful insights from many structured and unstructured data, using various scientific methods, machine-learning algorithms, data mining techniques and big data. The healthcare is an industry which generates huge amount of data based on Patient demography, Doctors profile, Planning of Treatments, Results obtained from medical examinations, insurance, etc. Such data snatch the attention of data scientists. The field of data science provides technical support to manage, process and to analyze the huge of amount of data generated by healthcare. The healthcare data needs effective analysis to obtain the accurate results. This article reviews about the process of data cleansing, data preprocessing and data analysis employed in healthcare applications. In addition, the article also presents the role of big data analytics in healthcare, the techniques used, current challenges and possible solutions. Data science and big data analytics provides the strategy for decision making in healthcare, which in turn reveals new scope to enhance the quality of healthcare.*
Keywords - *Big data, Data mining, Data analytics, Healthcare, Healthcare informatics.*

I. INTRODUCTION

The growth of digital platform leads the convergence of technology and healthcare, results in the new applications for data analytics [1]. In Healthcare, enormous amount of data like Patients Demography, Doctors Profile, Health Records of Patients, Prescriptions, Lab reports, Clinical reports, Sale and purchase records of Pharmacy, Investigation reports, Details of Medical Insurance, etc., were present. This provides a vast opportunity to study and analyze these data using the latest technologies. Machine Learning algorithms can be used on such data to analyze and understand various patterns that helps for fine decision-making which in turn provides good care on patient [2]. In addition, the analysis supports to understand the facts which help to improve the outcome of healthcare, early detection of disease and the treatment needed for the same. The Health information Exchange system should be established that extracts the data from various clinical repositories and put together in to one patient's record. The system also allows all physicians to access in secured manner [2, 3]. Hence, all the healthcare organizations should equip them to utilize big data, which will help them to establish the better network on healthcare which in turn provides a substantial benefit.

The emerging techniques of data mining will transfer the traditional medical database in to a one which was rich in knowledge in the forth coming years. Big data and data science in healthcare become more vital by the rise of social media and the mobile apps which will monitor the patient's health factors through analyzers and sensors. The task of data mining is to enhance the patient's information to present a fine care and treatment. This article focused on the methodologies and the advantages of the big data in the domain of health care. It also focused on the huge data generated in the healthcare system, quality of these data, issues related to security, handling of data and how the analysis of these data supports healthcare assessment and future diagnosis.

II. SEEKING APPROACH

A review of literature was conducted on the publications done during the last seven years on data science, big data in the domain of health care. The study focused on the articles which were published on the Scopus, Medline, Elsevier, Google Scholar etc., further more review also done on the articles mentioned in the references of the selected article. The understandings of the study are briefly explained in the following topics.

A. More about Big Data

Big data normally contains a huge volume of data which cannot be handled using the normal technologies. Big data can be defined and represented in many ways, in which 3 V's can be considered to define big data. i.e., 3 V's names as three dimensions - Volume, Velocity Big data normally contains a huge volume of data which cannot be handled using the normal technologies. Big data can be defined and represented in many ways, in which 3 V's can be considered to define big data. i.e., 3 V's names are three dimensions -

Volume, Velocity and Variety [4]. In big data, the word 'big' represents that the data Volume is huge. The speed at which data accessed is represented as Velocity. Variety represents the different form of raw data which was collected from various sources. In addition, 4th V- Veracity is also need to be added in the big data definition, which talks about the accuracy and reliability of the dataset. The 4th V ensures that the data in the dataset are relevant with high quality [4].

Big data needs the applications which use high-end computing and the algorithms based on AI in order to understand the data. Machine Learning, Fuzzy and Neural Network based approaches lead for automatic decision making. In addition to process, big data needs effective storage and visualization to know the insights and to make better decision.

B.   *Medicare repository for Big Data*
Healthcare is a domain designed to diagnose, prevent and to treat all kinds of diseases. Physicians, Clinics, drug centers, Laboratories etc., are the key part of the medical care domain [5]. Based on the disease severity, different levels of medical care are given. In all the level of medical care, Physicians need different kinds of data like medical history, previous treatment, drugs provided, lab diagnosis report, and other personal information. Traditionally these data are available in the form writings or in the form print.

The emergence of digital technology leads to digitize all the medical records. In the year 2003, The Institute of Medicine initiated "Electronic Health Record system", to represent portal that holds all the medical records related to the patients. Electronic Health Record system is an electronic system which is designed to store, retrieve and to connect the Physicians and patients for medical services.

C.   *Tools for Big data*
It is not an efficient way to access the big data from the centralized storage even with powerful processors. Hence to process the big data, it is efficient to distribute the data into several parallel systems [6]. Working with several processors involves with issues like parallel computing, data distribution and failure management. The open source tool like, Apache Hadoop, Apache Saprk, Apache Cassandra, Apache Srorm, Rapid Miner and Cloudera provides the platform to process big data in an effective way by addressing the above issues.

D.   *Health care Data Source for Big data*
In real world, many sources are available to obtain the data related to health care. The Figure – 1 depicts the same. The major part of health care data source was generated by means of Humans which includes Medical Records, Health Records and Patient Records [7]. The second largest source of data is online data. Usage of social media and emerging digital development in the medical sector lead to this. Sensor data, Big Transactional data and Biometric data are the other source of data for health care big data.
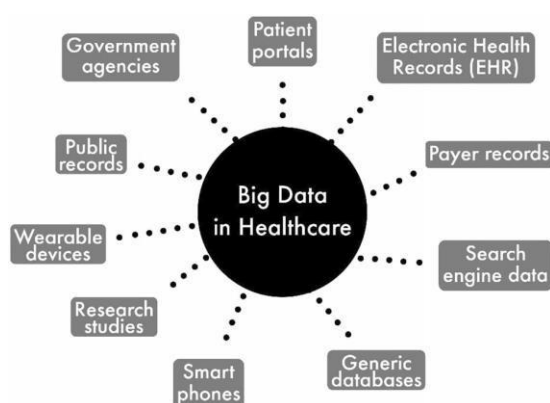

Fig-1: Healthcare data source for big data

E.   *Analytical applications in healthcare*
In healthcare, mainly there are six areas in which analysis needed [8]. The figure-2 depicts the same.
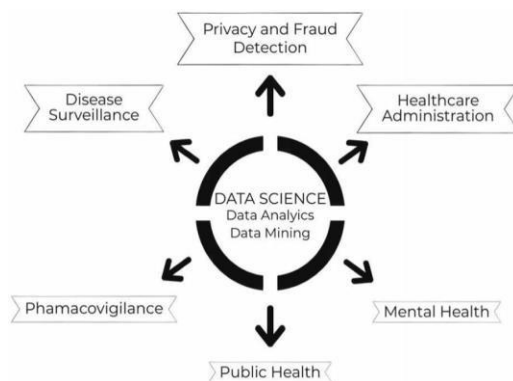
Fig-2: Applications of Data Science in Health care

The applications of healthcare include Surveillance of disease, Public health, Public mental health, Management and administration of healthcare, Phamacovigilance, Fraud detection and finally Fraud Detection. Researchers also employed on data storage and cloud computing, data quality optimization, cost cutting, influencing resources, patient handling and other fields [9].

F.   Disease surveillance
     It involves with the awareness of disease, knowing its condition, identifying its causes and prevention [10]. Figure-3 depicts the same. The data retrieved from Electronic Health Record and from the web has an enormous vision for the analysis of disease. Difference surveillance methodologies lead to plan good services, assessment of treatment, priority setting, health policy development and practice.
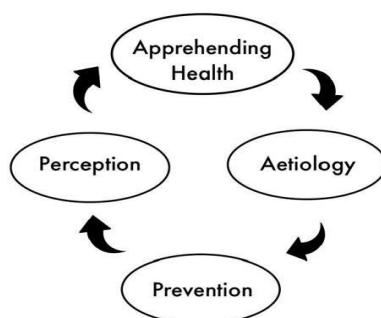

Fig-3: The disease analysis system

G.   Image processing on healthcare data
     Image processing on the data of healthcare provides precious knowledge about anatomy and organ functioning [11]. It also finds the disease stage and the health conditions of the patient. Currently in healthcare system, wavelet technology is mainly used for image processing. The usage of AI techniques on image processing improve the facet of healthcare which includes diagnosis, predictions and comparing medical images with other kinds of data [12]. In addition, AI techniques useful in Genomic data science, the study which enables researchers to decode the information hidden in DNA Sequence. This will support to provide proper treatment and also assist to diagnose the diseases in the early stage. The proportional increase in numbers on the medical facilities and the patient's aid to use better computer aided diagnosis and decision making.

H.   Healthcare administration
     The knowledge discovered from big data provides more insight to healthcare providers' information which helps them to provide proper medical care to their patients. Figure -4 depicts the role of big data in healthcare treatment process. Researchers implement various data mining techniques on medical data warehouse and cloud computing, which improves the quality of medical service, cost cutting, patient handling and other fields of healthcare.

I.   Medical data Privacy and fraud detection
It is very crucial to maintain privacy and fraud detection in medical data. This provides more demand for data scientists to preserve the medical data from hackers [13].Interoperability, Size and Privacy are the lacks of EHR data system. Introduction of mobile based cloud computing has solution to overcome these lacks. This cloud

system is designed to store medical data from various healthcare providers and provides restricted access to the physicians and patients. It uses encryption algorithm, OTP (one time password) or a two factor authentication to ensure data security.

The big data analysis performed by using the tool like big query tool and Map reduce. Such an approach reduces the cost, improves the efficiency and enhances data protection in better manner than the conventional algorithms.
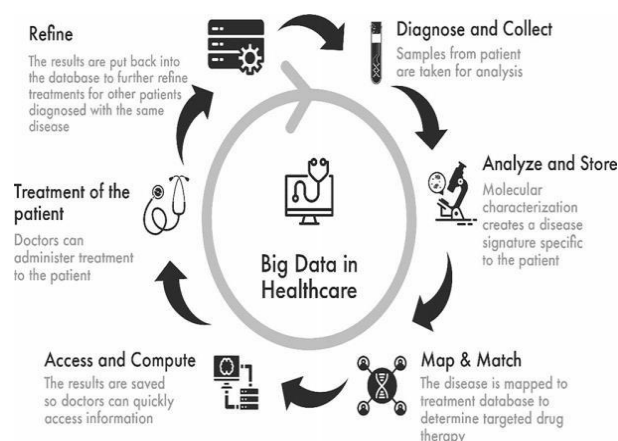

Fig-4: Role of Big Data in TreatmentProcess

J.   Mental Health and Public Health

Based on the national survey conducted on drug use, it was revealed that 52.2% of the total population in US wassuffered by either by drug addiction or mental problems [14]. In addition, it is also revealed that 30 million people were affected by panic attacks and anxiety disorders.

Kostkova et al. [15] performed the analysis on the online records collected based on the behavior pattern and the factors (Reported by Media) which affects the public and as well professionals. From the analysis, they found certain uniquefactors which affect the public health agencies, skilled and laypersons.

III.   CHALLENGES

Huge volume of data presents in a huge number of data points are sequencedata. Handling and controlling the entire data related to millions of patients would be a challenging one [16]. The main challenges includes the following

- The data which was collected willnot be in organized fashion and also inaccurate. This will raise the problem while gaining the insight in the data.
- Preserving patient information, guaranteeing the quality and accessing this data is very difficult to decide.
- Standardization of data, privacy maintenance, storage efficiency and data transfers needs more manpower to monitor constantly and to ensure that the requirements are met.
- Due to non availability of standards, it is very critical to integrate genomic data into medical for further studies.

IV.   FUTURE DIRECTION

Healthcare services needs to improve the quality of the treatment regularly. It is wondered that the innovative techniques lead to establish the better future. Big data is the revolt for the healthcare domain. The health behavior of patients, attitude of doctors and the service of healthcare providers had just begun to transform. The explosion of data science and emerge of different data driven applications plays a vital role in tuning the health sector as a leading service provider that provides better services to its customers. Data scientists canget more useful insights on medical data to improve the pharmaceutical productivity andto provide best medical services based on the huge volume of data that presents on the healthcare sector.

V.   CONCLUSION

Health care sectors exhibits huge amount of patient dataset. This includes surveillance data, laboratory data, genomics data, imaging data and electronic healthcare data records. Such a data needs proper management and analysis to acquire useful information. The utilization of big data supports for self-management long term visions, enhanced patient care and improved treatment. Data science employs with many predictive analytics

technique which is used to get deep knowledge about disease processes and enhanced patient centric treatment. It will also useful to develop the ability of researchers in the domain of data science. The accuracy on prediction is fully depending on the effective data integration which was collected from various sources. Modern health sectors can transform personalized medicine and medical therapy by combining health and biomedical data. Data sciences, an intelligent platform that can handle big data effectively, evaluate efficiently and interpret intelligently byestablishing the new trail in the comprehensive medical care.

REFERENCES

[1]. Sengupta PP (2013) Intelligent platforms for disease assessment: novel approaches in functional echocardiography. JACC: Cardiovascular Imaging 6(11):1206–1211.

[2]. Muni Kumar N, Manjula R (2014) Role of big data analytics in rural health care-a step towards svasth bharath. Int J Comp Sci Inform Technol 5(6):7172–7178

[3]. Ren Y, Werner R, Pazzi N, Boukerche A (2010) Monitoring patients via a secure and mobile healthcare system. IEEE Wirel Commun 17(1):59–65

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

[5]. Burghard C (2012) Big data and analytics key to accountable care success. IDC health insights :1–9

[6]. Bollen J, Mao H, Zeng X (2010) Twitter mood predicts the stock market. J Comp Sci 2(1):1–8.

[7]. Kuehn BM (2013) NIH recruits centers to lead effort to leverage "big data." JAMA 310(8):787–787

[8]. Castiglione A, Pizzolante R, De Santis A, Carpentieri B, Castiglione A, Palmieri F (2015) Cloud-based adaptive compression and secure management services for 3D healthcare data. Futur Gener Comput Syst 43:120–134

[9]. De Mauro A, Greco M, Grimaldi M (2016) A formal definition of big data based on its essential features. Library Review 65(3):122–135.

[10]. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of Things (IoT): a vision, architectural elements, and future

[11]. Sri Venkat Gunturi Subrahmanya · Dasharathraj K. Shetty · Vathsala Patil · B.M. Zeeshan Hameed (2022) The role of data science in healthcare advancements: applications, benefits, and future prospects, Springer, 191:1473–1483.

[12]. Atasoy H, Greenwood BN, McCullough JS (2019) The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 40:487– 500

[13]. Reisman M (2017) EHRs: the challenge of making electronic data usable and interoperable. Pharmacy and Therapeutics 42(9):572

[14]. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health information science and systems 2(1):3

[15]. Issa NT, Byers SW, Dakshanamurthy S (2014) Big data: the next frontier for innovation in therapeutics and healthcare. Expert Rev Clin Pharmacol 7(3):293–298

[16]. Baldwin T, Cook P, Lui M, MacKinlay A, Wang L (2013) How noisy social media text, how diffrnt social media sources? In Proceedings of the Sixth International Joint Conference on Natural Language Processing (pp. 356–364)