International Journal of Advanced Scientific Technologies,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

An Introduction to Data analytics in Healthcare industry

J. Sarada[1]and Dr. N. V. Muthu Lakshmi[2]

[1]Research Scholar,Dept of Computer Science, S.P.M.V.V, Tirupati, AP, INDIA
sarada.jada@gmail.com
[2]Assistant Professor: Dept of Computer Science S.P.M.V.V, Tirupati, AP, INDIA
nvmuthulakshmi@yahoo.co.in

Abstract – Data analytics is the science of examining bigdata to derive conclusions from the computed analytical results using some standard techniques. It is also referred to as qualitative and quantitative techniques which can be used to enhance productivity, business gain or any service of the organization. Data analytics is used in many industries to allow companies and organization to make good decisions to improve the performance. For the sciences, data analytics are used to verify or disprove existing models or theories.The techniques used in data analytics vary from one organization to another organization depending on their requirements. Data analytics is also known as data analysis where analysis of data is a process of inspecting, cleansing, transformingand modelling data with the goal of discovering useful hidden or uncoverinformation or knowledge.This useful information can be used to make conclusions or good decisions. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different domains like in sciences, business, healthcare, insurance industries, social science, political and so on. Among many industries, one of the most promising areas where big Data can be applied to make a change is healthcare and this industry utilizes data analytics to make good decisions to improve their services. Healthcare analytics has the potential to reduce costs of treatment, avoid preventable diseases, predict outbreaks of epidemics,avoid preventable deaths and improve the quality of life in general.In this paper, an overview of the data analytics,significance of big data in data analytics in health care industry is addressed. The data are collected for analysis purposes in various sources in health care industry, and these sources and also the devices through which data are collected are presented in this paper. There are various tools in different platforms are available for analysing big data and some of the most popularly used tools are discussed here. Many big data applications in healthcare using data analytics methods are successfully used for improving their services are presented in this paper. Even though data analytics play a vital role in health care industry but there are some challenges existing to face and which are addressed in this paper. Among many industries, health care industry is using data analytics for various purposes successfully by having efficient tools &analytical techniques but faces some challenges because of collecting large data from various sources in different formats and also privacy is the main concern for protecting patient personal data.

Keywords: Data Analytics, Big Data, Data Analysis, Health care

I. INTRODUCTION

Data analytics is widely used to create business value in expensive advantage. Compared with many other industries, healthcare has been a late adopter of analytics. Many health systems have plenty of chances to develop a clinical quality, patient financial performance. Healthcare has become one of India's largest sectors - both in terms of revenue and employment. Healthcare comprises hospitals, medical devices, clinical trials, outsourcing, telemedicine, medical tourism, health insurance and medical equipments. The Indian healthcare sector is growing at a brisk pace due to its strengthening coverage, services and increasing expenditure by public as well private players. Healthcare delivery, which includes hospitals, nursing homes and diagnostics centres, and pharmaceuticals constitutes 65 per cent of the overall market. Big data typically refers to three types of data. Structured, Semi-Structured and
Unstructured. Structured data is data that can be easily stored, queried, recalled, analyzed and manipulated by machine. structured and semi-structured data includes instrument readings and data generated by the ongoing conversion of paper records to electronic health and medical records. Historically, the point of care generated unstructured data office medical records, handwritten nurse and doctor notes, hospital admission and discharge records, paper prescriptions, radiograph films, MRI, CT and other images.

II. HEALTHCARE DATA SOURCES AND BASIC ANALYTICS

The various data sources and their impact on analytical algorithms will be discussed. The heterogeneity of the sources for medical data mining is rather broad, and this createsthe need for a wide variety of techniques drawn from different domains of data analytics. The nature of health data has evolved, so too have analytics techniques scaled up to the complex and sophisticated analytics necessary to accommodate volume, velocity and variety. Gone are the days of data collected exclusively in electronic health records and other structured formats. Increasingly, the data is in multimedia format and unstructured.

A.Electronic Health Records

Electronic health records (EHRs) contain a digitized version of a patient's medical history. Itencompasses a full range of data relevant to a patient's care such as demographics, problems, medications, physician's observations, vital signs, medical history, laboratory data, radiology reports, progress notes, and billing data. Many EHRs go beyond a patient's medical or treatment history and may contain additional broader perspectives of a patient's care. An important property of EHRs isthat they provide an effective and efficient way for healthcare providers and organizations to sharewith one another. The storage and retrieval of health-related data is more efficient using EHRs. It helps to improve quality and convenience of patient care, increase patient participation in the healthcare process, improve accuracy of diagnoses and health outcomes, and also improve care coordination [1].

B.Biomedical Image Analysis

Medical imaging plays an important role in modern-day healthcare due to its immense capability in providing high-quality images of anatomical structures in human beings. Effectively analysing such images can be useful for clinicians and medical researchers since it can aid disease monitoring, treatment planning, and prognosis [2]. The most popular imaging modalities used to acquire a biomedical image are magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and ultrasound (U/S).The final goal of biomedical image analysis is to be able to generate quantitative information and make inferences from the images that can provide far more insights into a medical condition. Such analysis has major societal significance since it is the key to understanding biological systems and solving health problems. The number of general problems that arise in analysing images are object detection, image segmentation, image registration, and feature extraction. All these challenges when resolved will enable the generation of meaningful analytic measurements that can serve as inputs to other areas of healthcare data analytics.

C.Biomedical Signal Analysis

Biomedical Signal Analysis consists of measuring signals from biological sources, the origin of which lies in various physiological processes. Examples of such signals include the Electroneurogram(ENG), Electromyogram (EMG), Electrocardiogram (ECG), Electroencephalogram (EEG),Electrogastrogram(EGG), Phonocardiogram (PCG).The measurement of physiological signals gives some form of quantitative or relative assessment of the state of the human body. The processing and interpretation of physiological signals is challenging due to the low signal-to-noise ratio (SNR) and the interdependency of the physiological systems. The signal data obtained from the corresponding medical instruments can be copiously noisy, and may sometimes require a significant amount of pre-processing. Several signal processing algorithms have been developed that have significantly enhanced the understanding of the physiological processes. A wide variety of methods are used for filtering, noise removal, and compact methods [3]. More sophisticated analysis methods including dimensionality reduction techniques such as Principal Component Analysis (PCA), Singular Value Decomposition (SVD), and wavelet transformation have also been widely investigated in the literature.

D.Genomic Data Analysis

A significant number of diseases are genetic in nature, but the nature of the causality between the genetic markers and the diseases has not been fully established. For example, diabetes is well In some other cases, such as the blindness caused by Stargardt disease, the relevant genes are known but all the possible mutations have not been exhaustively isolated. Clearly a broader understanding of the relationships between various genetic markers, mutations, and disease conditions has significant potential in assisting the development of various gene therapies to cure these conditions. One will be mostly interested in understanding what kind of health-related questions can be addressed through in-silico analysis of the genomic data through typical data-driven studies. Moreover, translating genetic discoveries into personalized medicine practice is a highly non-trivial task with a lot of unresolved challenges. For example, the genomic landscapes in complex diseases such as cancers are overwhelming complicated, revealing a high order of heterogeneity among different individuals.

E.Clinical Text Mining

Most of the information about patients is encoded in the form of clinical notes. These notesare typically stored in an unstructured data format and is the backbone of much of healthcare data. These contain the clinical information from the transcription of dictations, direct entry by providers, or use of speech recognition applications. It is needless to say that the manual encoding of this free-text form on a broad range of clinical information is too costly and time consuming, though it is limited to primary and secondary diagnoses, and procedures for billing purposes. Such notes are notoriously challenging to analyse automatically due to the complexity involved in converting clinical text that is available in free-text to a structured format. It becomes hard mainly because of their unstructured nature, heterogeneity, diverse formats, and varying context across different patients and practitioners. Natural language processing (NLP) and entity extraction play an important part in inferring useful knowledge from large volumes of clinical text to automatically encoding clinical information in a timely manner [4].The processing of clinical text using NLP methods is more challenging when compared to the processing of other texts due to the ungrammatical nature of short and telegraphic phrases, dictations,

shorthand lexicons such as abbreviations and acronyms, and often misspelled clinical terms.

F.Social networks

The analysis of data is unstructured social media text allows you to uncover the sentiments of your customers among those in different geographical locations or even some different demographic groups.Facebook, Twitter, and other social media platforms to generate a variety of data around the clock, giving a view into the locations, health behaviours, emotions, and social interactions of users.

G.Devices:

The devices are used for direct data entry to the computer system. There are some electronic devices that acts as a source data-entry such as scanner, cameras, bar-code reader, electronic chips, and audio.Healthcare providers and payers are increasingly turning to big data and analytics, to help them understand their patients and the frameworks of their illnesses in more detail.Massive data of health-related information are also collected and stored on mobile and home devices.

A.Smart phones:

Many no of Health apps capture information on the user's physical activity, nutritional intake, sleep patterns, emotions, and other parameters. Native cell phone apps (e.g. GPS, email, texting) can also give ideas about an individual's health status.

B.Wearable monitors and devices:

In health care systems related technologies are presenting new opportunities and risks for businesses. Pedometers, accelerometers, glasses, watches, and chips embedded under the skin also gather health-related information.

C.Telemedicine devices

Healthcare providers to monitor patients' parameters such as blood pressure, heart rate, respiratory rate, oxygenation, temperature, ECG tracings, and weight.

III.ADVANCED DATA ANALYTICS FOR HEALTHCARE

In this healthcare system, mentioned some of the advanced data analytics methods for healthcare providers and payers are increasingly turning to bigdata and analytics,These techniques include various data mining and machine learning models that need to be modified to the healthcare domain.

A.Clinical Prediction Models

Several predictionmodels have been extensively investigated and successfully deployed in clinical practice[5]. Such models have made a tremendous impact in terms of diagnosis and treatment of diseases. Most successful supervised learning methods that have been employed for clinical prediction tasksfall into three categories: (i) Statistical methods such as linear regression, logistic regression, andBayesian models; (ii) Sophisticated methods in machine learning and data mining such as decisiontrees and artificial neural networks; and (iii) Survival models that aim to predict survival outcomes. The goal is to predict the time of occurrence of a particular event of interest. These survival models are also widely studied in the context of clinical data analysis in terms of predicting the patient's survival time. There are different ways of evaluating and validating the performance of these prediction models.

B.Temporal Data Mining

Healthcare data almost always contain time information and it is inconceivable to reason and mine these data without incorporating the temporal dimension. There are two major sources of temporal data generated in the healthcare domain. The first is the electronic health records (EHR) data and the second is the sensor data. Mining the temporal dimension of EHR data is extremely promising as it may reveal patterns that enable a more precise understanding of disease manifestation, progression and response to therapy. Some of the unique characteristics of EHR data (such as of heterogeneous, sparse, high-dimensional, irregular time intervals) makes conventional methodsinadequate to handle them. Unlike EHR data, sensor data are usually represented as numeric time series that are regularly measured in time at a high frequency. Examples of these data are physiological data obtained by monitoring the patients on a regular basis and other electrical activity recordings such as electrocardiogram (ECG), electroencephalogram (EEG), etc. EHR data are usually mined using temporal pattern mining methods, which represent data instances (e.g., patients' records) as sequences of discrete events (e.g., diagnosis codes, procedures, etc.) and then try to find and enumerate statistically relevant patterns that are embedded in the data.

C.Visual Analytics

Visual analytics provides a way to combine the strengths of human cognition with interactive interfaces and data analytics that can facilitate the exploration of complex datasets. Visual analytics is a science that involves the integration of interactive visual interfaces with analytical techniques to develop systems that facilitate reasoning over, and interpretation of, complex data [6]. Visual analytics is popular in many aspects of healthcare data analysis because of the wide variety of insights that such an analysis provides.The rapid increase of health-related information, becomes critical to build effective ways of analysing large amounts of data by human–computer interaction and graphical interfaces. The ability to analyse and identify meaningful patterns in multimodal clinical data must be addressed in order to provide a better understanding of diseases and to identify patterns that could be affecting the clinical workflow. The multimodal, noisy, heterogeneous, and temporal characteristics ofthe clinical data pose significant challenges to the users while synthesizing the information and obtaining insights from the data [7]. The amount of

information being produced by healthcare organizations opens up opportunities to design new interactive interfaces to explore large-scale databases, to validate clinical data and coding techniques, and to increase transparency within different departments, hospitals, and organizations. While many of the visual methods can be directly adopted from the data mining literature [8], a number of methods, which are specific to the healthcare domain, and also have been designed.

IV. BIG DATA IN HEALTHCARE

Big data analytics can be applied to improve patient care and healthcarecentre operations are used to predict epidemics, cure disease, improve quality of life and avoid preventable deaths. With the world's population increasing and everyone living longer, models of treatment delivery are rapidly changing, and many of the decisions behind those changes are being driven by data. The drive now is to understand as much about a patient as possible, as early in their life as possible.

V. Tools and platforms for Analysing Big Data

Many platforms and tools are used for big data analytics in healthcare .When ever collected in theform of individual electronic health records, this data not only improves continuity of care for the individual, but it can be used to create huge datasets with which treatments and outcomes can be treated as efficient as well as cost effective manner.

A. Decision Management

This type of analysis enables individual recommendations across multiple channels, maximizing the value of every customer interaction. Oracle Advanced Analytics scores can be integrated to operationalize complex predictive analyticmodels and create real-time decision processes

B.Discovery tools

Discovery are useful throughout the information lifecycle for rapid, intuitive exploration and analysis of information from any combination of structured and unstructured sources. These tools permit analysis alongside traditional BI source systems. users can draw new insights, come to meaningful conclusions, and informeddecisions quickly.

C.In-Database Analytics include a variety of techniques for finding patterns and relationships in your data. Because these techniques are applied directly within the database, you eliminate data movement to and from other analytical servers, which accelerates information cycle times and reduces total cost of ownership.

D.Hadoop is useful for pre-processing data to identity macro trends, such as out of-range values. It enables businesses to unlock potential value from new data using inexpensive commodity servers. Organizations primarily use Hadoop as a precursor to advanced forms of analytics.

E.BI tools are important for reporting, analysis and performance management, primarily with transactional data from data warehouses and production information systems. BI Tools provide comprehensive capabilities for business intelligence and performance management, including enterprise reporting, dashboards, ad-hoc analysis, and scorecards.

VI.BIG DATA APPLICATIONS TO HEALTHCARE

The big data is in the most important areas in each field, including storage, retrieval, error identification, data security, data sharing and data analysis for electronic patient records, and also social media data. The era of big data in technology is being promptly applied to biomedical signal analysis,biomedical signal analysis, and also health-care fields.

A. Data Analytics for Pervasive Health

Pervasive health refers to the usage of advanced technologies like wearable sensors. A wide variety of sensor modalities can be used when developing intelligent health systems, including wearable and ambient sensors [9].These methods faces number of challenges like knowledge extraction from the large volumes of data. In the case of wearable sensors, sensors are attached to the body or woven into garments. Several practical healthcare systems have started using analytical solutions. Some examples include cognitive health monitoring systems based on activity recognition, persuasive systems for motivating users to change their health and wellness habits, and abnormal health condition detection systems.

B.Healthcare Fraud Detection

The complexity of the healthcare domain, which includes multiple sets of participants, including healthcare providers, beneficiaries (patients), and insurance companies, makes the problem of detecting healthcare fraud equally challenging and makes it different from other domains such as credit card fraud detection and auto insurance fraud detection.

C.Computer-Aided Diagnosis

Computer-aided diagnosis/detection (CAD) is a procedure in radiology that supports radiologists in reading medical images [10]. The three important stages in the CAD data processing are candidate generation (identifying suspicious regions of interest), feature extraction(computing descriptive morphological or texture features), and classification (differentiating candidates that are true lesions from the rest of the candidates based on candidate feature vectors).

VII. BIG DATA CHALLENGES TO HEALTHCARE

Applying Big Data analytics to the development aspects several challenges. Big data Challenges are faced by each of these are unique and arechances to capturing,

storing, searching, sharing, analysing and visualizing, improve service, reduce costs, and improve healthcare systems.Sensor data for a specific subject are measured in shorter period of time (usually several minutes to several days) compared to the longitudinal EHR data (usually collected across the entire lifespan of the patient).

A.Interpretation, Propensity, Correlations

Togenerate value from Big Data systems will depend on the statistically valid use of the information. The size and heterogeneity of the data being collected is a major challenge, particularly since the majority of statistical approaches to interpretation were developed in an era when "sample sizes" were relatively small, and when data acquisition technologies and computing power were limited [11]. Big Data encompasses a high level of messiness in the sense that the increase in the amount of information by orders of magnitude means giving up the preference for highly curated data for the sake of having a higher sample and effect size [12]. Structured data, such as tables of numbers, do not reveal everything that is known about a medication or biological process and much of what is known about living organisms exists in unstructured formats [13]. The analysis of unstructured, voluminous and disorganized data has brought significant discoveries [14] [15]

B.Standards and Interoperability

Standardization problems in the healthcare sector, as data is often fragmented, or generated in IT systems with incompatible formats [16]. Research, clinical activities, hospital services, education, and administrative services are siloed, and, in many organizations, each silo maintains its own separate organizational (and sometimes duplicated) data and information infrastructure. The lack of cross-border coordination and technology integration calls for standards to facilitate interoperability among the components of the Big Data value chain.

C.Expertise and Infrastructure

Big Data offers enormous possibilities for new insights, for understanding human systems at the systemic level, and for detecting interactions and nonlinearities in relations among variables.

VIII. CONCLUSION

The concepts like data analytics methods, advanced data analytics methods and some varied tools that perform data analysis ,cleaning and presentation. The concept of big data, its handling Privacy preserving data sharing, visual analytic techniques. Big Data analysis tools play a vital role in Healthcare system.Big Data Analytics has a crucial role as the government wants to increase the possibilities of developing new and innovative digital services in the society.

REFERENCES

[1] Catherine M. DesRoches et al. Electronic health records in ambulatory careanational survey of physicians. New England Journal of Medicine 359(1):50–60, 2008.

[2]Stanley R. Sternberg, Biomedical image processing. Computer 16(1):22–34, 1983.

[3] Athanasios Papoulis. Signal Analysis. McGraw-Hill: New York, 1978.

[4] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, and J. F. Hurdle.Extracting information from textual documents in the electronic health record: A review of recent research.Yearbook of Medical Informatics, pages 128–144, 2008

[5] E. W. Steyerberg. Clinical Prediction Models.Springer, 2009.

[6] DanielKeim et al. Visual Analytics: Definition, Process, and Challenges. Springer Berlin Heidelberg, 2008 Computer Graphics, 8(1):1–8, 2002.

[7] K. Wongsuphasawat, J. A. Guerra Gmez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B.Shneiderman.LifeFlow: Visualizing an overview of event sequences. In Proceedings of theSIGCHI Conference on Human Factors in Computing Systems,1747-1756.ACM, 2011.

[8] Daniel A. Keim. Information visualization and visual data mining.IEEE Transactions on Visualization.

[9]. Min Chen, Sergio Gonzalez, AthanasiosVasilakos, Huasong Cao, and Victor C. Leung. Body area networks: A survey. Mobile Networks and Applications, 16(2):171–193, April 2011.

[10]. KunioDoi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics, 31:2007.

[11]. OECD, Strengthening Health Information Infrastructure for Health Care Quality Governance.
2013: OECD Publishing. 180.

[12]. Mayer-Schönberger, V. and K. Cukier, Big data: A revolution that will transform how we live, work, and think. 2013: Houghton Mifflin Harcourt.

[13]. May, M., Life Science Technologies: Big biological impacts from big data. Science, 2014. 344(6189): p. 1298-1300.39. OECD, Strengthening Health Information Infrastructure for Health Care Quality Governance.2013: OECD Publishing. 180.

[14]. Baker, E.W., Relational Model Bases: A Technical Approach to Real-time Business Intelligence and Decision Making.Communications of the Association for Information Systems, 2013.33(1): p.23.

[15]. White, R.W., et al., Web-scale pharmaco vigilance: listening to signals from the crowd. Journal of the American Medical Informatics Association, 2013: p. amiajnl-2012-001482.

[16]. Roney, K. (2012) If Interoperability is the Future of Healthcare, What's the Delay? .Becker's HospitalReview.