Big Data Analytics in Healthcare: Challenges and Future Research Directions

Padmini Mansingh[1], Binod Kumar Pattanayak[2], Bibudhendu Pati[3]
[1,2]Department of Computer Science and Engineering,Institute of Technical Education and Research (ITER), Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, INDIA
[3]Department of Computer Science ,Rama Devi Women's University,Bhubaneswar, Odisha, INDIA
Email:[1]padminimansingh@gmail.com, [2,]binodpattanayak@soa.ac.in, [3]patibibudhendu@gmail.com

ABSTRACT

Due to recent development of new technologies, "Big Data" is no more a perpetual crucial word for the current humanity. It has tremendous contribution in the health care sector including advanced solutions for precarious health hazards in various dimensions. The current research article portrays big data analytics as a benefit to modern civilization in terms of handling increased growth of data from different sources. In the field of healthcare industry, it is expected to discover hidden pattern for real-time prediction of critical diseases through big data analytics. Further this article focuses on significant applications of big data analytics with respect to bioinformatics for advanced health care. Our investigations facilitate different benefits in recent progress in real-time through the study of imaging, predicting, detecting and formulating solutions for critical diseases at early stage is the future research challenge. However, we highlight an assortment of exigent concern for advanced healthcare services with respect to biological variables

Keywords
Big Data Analytics, Cloud Services, Healthcare, Disease Prediction, Hadoop, Data Privacy, Data Processing

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

A Sharp intellect behind Big Data is explosion of data that comes with different contradicts characteristics that need various recent improved technological methodologies for a real-time improved decision making to address this situation, researchers derived different tools used to deal and extract valuable knowledge from different areas like biologic, genomics, business, biochemistry etc. and it can only be possible by proper in-memory big data storage management and processing needs to be considered for management of automatically increasing state of huge amount of data (Zhang et al., 2015). But in this review, we present overall top down description of healthcare application for an improved performance in personalized treatment mode. Furthermore, perspective sentient monitoring to build pointed highlights, some limitations of healthcare industry motivate researchers for better healthcare management.
Now days, various emerging diseases have been appearing due to different unknown viruses. For that reason, we need some better healthcare analytics to overcome this situation. Smarter healthcare systems stimulate the requirement of observation study for emergency medicine that is accomplished by health centre (Gkoulalas-Divanis et al., 2014) which is one of the important techniques to reduce the gaps observed in previous healthcare systems. Due to increasingly usage of sensors in medical field, large volume of unstructured medical data is generated on daily basis. By the new infographics from oracle (*Oracle,* 2019), the predicted acceleration rate of data generated by healthcare industry will supposed to touch 25,000 terabytes by 2020. How to improve the level of care to each patient by analysing large amount of data and getting them a reliable analytics results, is the main topic behind this research.
In this paper, in order to analyse the behavioural role of big data analytics in the field of healthcare industry, we have ornately explained different opinions and also intend to apply them in big data analytics for storage and processing of heterogeneous medical data.
In summary, our goal in this paper is to briefly describe the following further sections:
2) The inclusive overview of big data in terms of storage and processing management;
3) Role of big data analytics in the field of healthcare industry.
4) Some of healthcare difficulties that can be overcome by using big data analytics.
5) Future research directions in terms of BDA (Big Data Analytics) in healthcare industry.
The remaining part of the paper is organized as follows. An overall overview of big analytics and healthcare analytics is mentioned in section 2 and need of big data analytics in the field of healthcare in section 3.

Overviews Of Big Data In Terms Of Storage And Processing Management

Big Data

"Big data is concerned with a huge amount exponentially growing data that couldn't be controlled by traditional database management system (Andreu-Perez et al., 2015)" and the related term "big data analytics" have different opinions on its definition. However, the very popular opinion is "big data analytics is the process of examining big data to uncover hidden patterns unknown correlation, market trends, customer preferences and other useful information's that can help organization make more informed business decisions (SAS, 2019)".

Now we will discuss the different features of big data as follows:

2.1.1 Characterization of big data;
2.1.2 Data types involved in big data;
2.1.3 Sources of big data resources;
2.1.4 Big data analytics in terms of storage

Big data analytics in terms of processing;

Characterization of Big Data

The characterization of big data is mainly based on huge amount, high quality, high speed, exactness and types of data (*Marr, 2014*). But due to certain requirement, we proposed a new 12V characteristic of big data analytical model when dealing with complex big data structure. As shown in Figure 1, we model 12V as velocity, variety, volume, veracity, validity, vulnerability, verily, volatility, variability, value, visualization and vagary as per the requirement to beat the challenges of huge amount of data. The long-term process between 3V (Laney, 2001) to 12V is going to become a historical model for overcoming from semantic gap between an assortment of assorted data resources.

Figure 1. Different characteristics of Big Data

Datatypes Involved in Big Data

Due to the increasingly usage of biomedical as well as lots of different types of data that are generated from different qualitative and quantitative sources, are generated in an accelerating manner. These are structured data, semi structured data and unstructured data which are illustrated in Figure 2. But especially, Big Data is more related to unstructured type because of daily updated millions of biomedical images, videos and documents. The amount of data that we gather from various data sources is divided into: structured, semi structured and unstructured respectively.

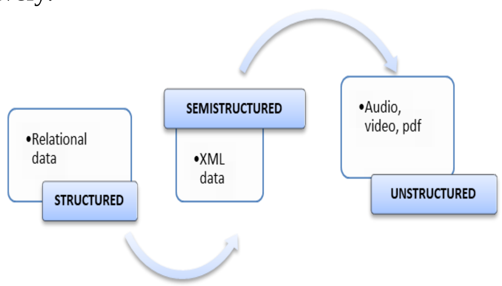

Figure 2. Types of data

Sources of Big Data

Uncertain and uncontrolled data obtained from different sources cause difficulties in processing. Various areas which have attracted the attention of researchers are detailed below.

Healthcare

Due to advancement of medical sectors to make various research projects, large volume of unstructured data like patient information, insurance, health plans and others is being generated in this field at every second. So, it is impossible to maintain the storage by traditional database management system. Big data analytics in life care comes with many new challenges. This is our major point in this ongoing research. The biomedical data which are analysed by different methodologies of big data applications carry through some regulations of different data gathering areas like health informatics, translational informatics, sensor informatics and imaging informatics (Andreu-Perez et al., 2015).

Transportation

The earliest means of transportation data included different vehicles' data, road traffic data, a well-developed transportation service's data, and distance data. Due to many changes brought in the global transport system, a huge amount of data is frequently gathered and the maintenance is a headache for the data scientist. Basically, monitoring of vehicular telemetry related to road safety and security of heterogeneous vehicles is to provide innovative network selection techniques for better road tracker (Tian et al., 2015) which is the future of research for the analysts. Also, for tracking vehicle movement for selection of minimum path for pointing to destination by using an overlapping and hierarchical social clustering model (OHSC) result model, researchers proposed social-based localization algorithm (SBL) algorithm to abetment in localization (Lin et al., 2017).

Social Marketing

Adapting to new changing technology Big data Analytics, health entrepreneurs are not only having in-depth understanding of their clients(patients) by giving them good healthcare business management but also have an endless amount of data searching for unlocking the opportunities of medical development. Marketing can be used to design and promote social beneficial behaviour. The knowledge of social marketing is incomplete for public health professionals. So, by describing varieties of case studies and discussions, the proper use of social marketing in the field of public health intervention is to be improved for the future development (Grier & Bryant, 2005).

Telecommunication

The large volume of data gathered by the biggest revolution of telecommunication companies needs proper analysis. That is the basis for the telecommunication industries to come with some new features which can necessarily improve customer and network services. Every second, billions of data are being collected from different sources of call detail records and network services. Mobile cloud computing system is only to intensify performance of cloud mobile media (Yin et al., 2015). The exceeding usage of mobile devices, the cloud-based mobile media help to manage the large-scale data (Wen et al., 2014). Rapid use of smart phones in healthcare field creates a huge amount of data that need to be analysed by CMA (cloud based mobile augmentation) performance which depends on two factors like mobile cloud distance and intermediate hops (Abolfazli et al., 2014).

Social Media

The amount of user-generated media by using more and more people for sharing of information such as Facebook, twitter and more gather huge amount of information by uploading millions of healthcare pictures and tweets. Around 250 billion health related photo uploads and 342,222 tweets in twitter are the biggest challenges to handle and especially for privacy implications of emerging trend of geo-tagged social media are fully based on analysis (Smith et al., 2012). People share their experiences and opinions on personal health issues, symptoms, expected treatments, medicine side-effects and make openly available social media data as a valuable resource for mining remarkable and actionable healthcare insights. For making a valuable decision by both doctor and patient with the real time disease surveillance system for monitoring disease types and symptoms, spatial, temporal and text mining on twitter data can be useful (Lee et al., 2013).

Banking

Presently, bouquets of services are available at the client's demand in today's repository system. Due to rapid growth of medical services, data generated from different stockpile of medical investment through banking areas of different online sectors create a biggest problem for data scientist. Authorized auditing and fine-grained update request supported for a scheme related to data security and also for other several protection methods are now proposed by new researchers (Liu et al., 2013). The misuse of patient cards for fraud detection is also overcome by big data analytical tools.

Search Engine Data

Today, we all are being active on internet at every second. We retrieve lots of valuable wide range of information from different medical databases which wouldn't be controlled by relational database management system. This growing range of medical data creates a peculiar behaviour. Evidence-based medicine is intended for predicting risk rate for readmission and providing a right care by analysing structured and unstructured (S&T) EMR (Electronic Medical Record) data, financial data, operational data, clinical data, and genomic data to match treatments with outcomes (Raghupathi & Raghupathi, 2014).

Government

Large volume of data is gathered by lot of sectors which are included in the government healthcare programs such as planning of early disease detection programs, health welfare agencies, poverty reduction agencies, survival development programs and so on. Managing and analysing this huge volume of data present now a headache for all countries in the world. The data which are gathered from different government and non-government agencies are properly analysed and managed by Big data analytic tools, researchers overview the ethical issues associated with big data, which extends into an examination of the practice issues for big data applied to the social welfare sector (Gillingham & Graham, 2017). The scenario of global health which is provided through online announcements by government agencies, informal channels, and press reports to blogs to chat rooms to analyses of Web searches (see Digital Resources for Disease Detection). This is completely different from disease reporting of traditional public health infrastructure (Brownstein et al., 2009). That is the only reason behind big data analytics for making proper valuable decision making.

Video Surveillance

The fast-growing data collected by medical surveillance cameras allow patients-doctors to use and annotate related surveillance events. The interrelationship between object and surveillance event could be measured by one of the case study of big data analytics. In addition, there is implication that how video resources are based on their links (Xu et al., 2015). For analysing huge amount of spatial climate data by big data-based surveillance system and which is uninterrupted monitoring of climate data changing with respect to deadly disease Dengue (Manogaran & Lopez, 2018) is one of the case studies solved by big data disease surveillance system. Today, cameras are everywhere for security concern. It is capable of providing a lot of information that can be used for long term analysis in order to get the attention of big volume of streaming data that may not be captured trough old techniques. According to IDC (International *Data* Corporation), the expected growth of big data business is $ 187 billion in 2019(IDC by press release) (IDC, 2018).

Big Data Analytics in Terms of Storage

The main aim of HACE theorem is to model the heterogeneous complex huge number of datasets with amendment of data mining techniques (Wu et al., 2013) and for preventive maintenance of Big Data according to different attributes of data by implementing traditional methods (Wan et al., 2017). So, Hadoop is an open source framework which is invented by Doug Cutting Mike Caferella in 2005 that plays a major role in this situation (Apache, 2015). With the help of Hadoop, we can store the amount of rapidly growing data which are gathered from different sources. HDFS (Hadoop Distributed File System) is a module of Hadoop architecture that provides a cluster of commodity hardware which is very much suitable for storage (**HDFS, 2019**).

Mainly HDFS has five services: name node, secondary name node, job tracker, data node and task tracker as shown in Figure 3 and explained one by one in different categories which is depicted below.

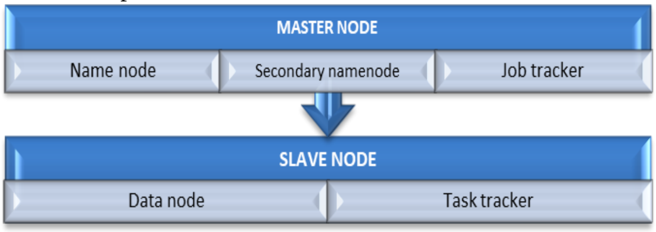

Figure 3. Layers of HDFS

Name Node

Name node is a software which can be broadly defined as the metadata holder that occupies the information about data nodes and the entire extraction of files. In general, it regulates and specifies where to write data and when to retrieve. Name space image and edit log are two varieties of the processed data which are stored on local disk permanently (Yang, 2011). HDFS has one name node and number of data nodes; the single name node manages file system (**HDFS, 2019**) and is the intermediary and warehouse for all HDFS metadata. All the machine whether it is name node or data nodes, run on a GPU (*Graphics Processing Unit)* OR LINUX operating system (**HDFS, 2019**).

Job Tracker

According to Wikipedia, job tracker is one of the master services and it communicates with name node for deciding which data are stored, in which location and gives the status updated by job tracker.

Secondary Name Node

According to Wikipedia, the secondary name node combines the image, edits log files from time to time, stores and edits log size within an edge. Name node also is a single point of failure for Hadoop distributed system. So, there is optional secondary name node for creating checkpoint and doesn't allow for any redundancy of name and also stores multiple copies of metadata (Apache, 2015).

Data Node

In master/slave HDFS architecture, data node is a software on a separate commodity hardware which is primarily used for file storage inside the total file given by the Client as an input that is split into one or more blocks and all are scattered into different Data nodes. as shown in fig.4, all read and write operations requested by Client are conducted through data node. One dot text (a block) file is replicated by default 3 times in three dissimilar data nodes. Both the name nodes as well as data node are allowed to run on commodity hardware. These are run a GNU/Linux operating system (**HDFS, 2019**). If heartbeat messages couldn't be sent by data node to name node, then Name Node will take redirect to replicated-copy of Block that mislaid on that node (Yang, 2011).

Task Tracker

The stepwise operation of task tracker is illustrated in below figure (Figure 4). Task tracker repeatedly send the heart beat to job tracker and as a node in the cluster that accepts tasks map, reduce and shuffle operations from the job tracker.

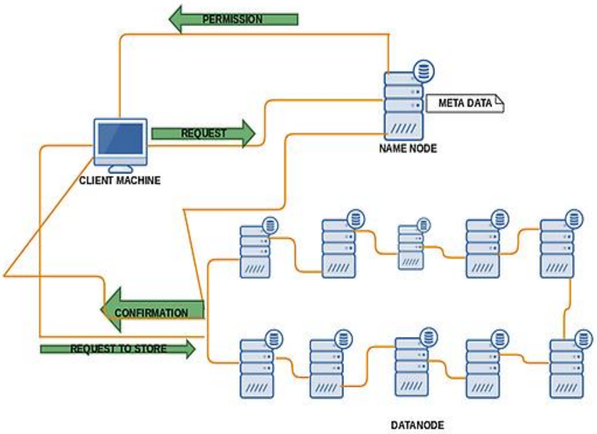

Figure 4. The HDFS Architecture

Big Data in Terms of Processing

After storage phase of rapidly growing online real time data, let us move to second phase i.e. processing. In this review, we present an overview of big data processing. A bottleneck inference can be obtained in real time from large and high dimensional observations. For instance, high dimensional spaces may arise from an extensive set of biomarkers, heath attributes, and sensor fusion. Several researchers have used big data bioinformatics approach to provide risk assessment level for prevention of bipolar disorder (BD) by assimilating and processing massive amount of compound phenotypic, anamnesis, behavioral, family and personal 'omics' profiling (McIntyre et al., 2014). Another upcoming research for big data processing is the progress by complex data integration and determining the level of problems faced by different health care communities related (Dong & Srivastava, 2013).
Figure 5 shows the major steps of big data processing. First, dataset for HDFS goes through map reduce algorithm for overcoming a mashie situation giving a valuable point wise result to researchers.

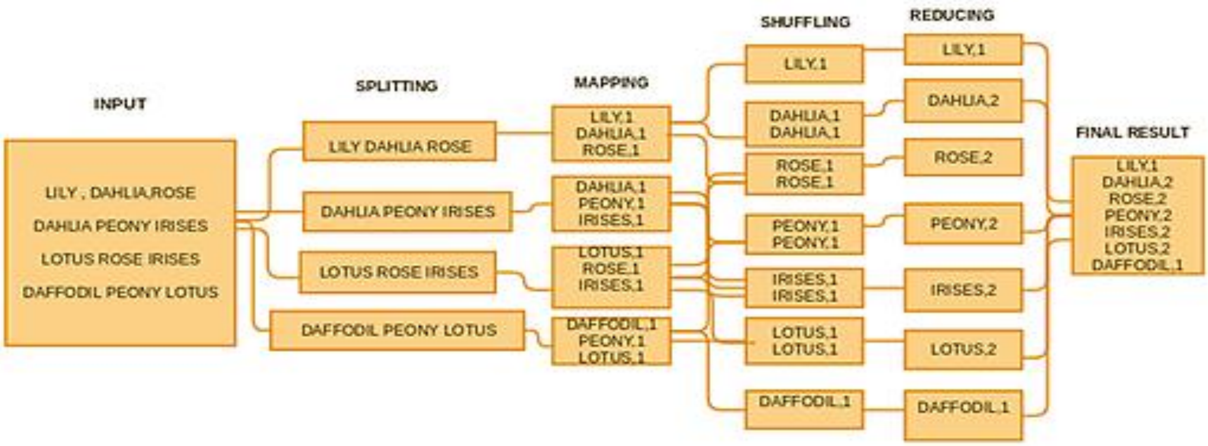

Figure 5. The Map Reduce Architecture

Mapper Phase

It is very much confusing whenever we discuss the map reduce phase because it includes different combinations of phases like mapper, shuffler and reducer. Map function splits the entire set of files into some input split which are in the form of segments that are stored in HDFS. The set of sentences from HDFS is split into some InputSplit generated by the InputFormat and then, Mapping by some intermediate input key and value pairs.
For each key/value pair in the input split
protected void map(KEYIN key,
VALUEIN value,
org.apache.hadoop.mapreduce.Mapper.Context context)
throws IOException,
InterruptedException
Throws:
IOException
InterruptedException
public class Mapper<KEYIN, VALUEIN, KEYOUT, VALUEOUT> {
102
103 /**
104 * The <code>Context</code> passed on to the {@link Mapper} implementations.

105 */
106 public abstract class Context
107 implements MapContext<KEYIN,VALUEIN,KEYOUT,VALUEOUT> {
108 }
public class Mapper<KEYIN,VALUEIN,KEYOUT,VALUEOUT>
extends Object
This above syntax referred from (Online Source, 2019).
By virtue of the above processing phase, many healthcare related huge amount of data comes under simpler phase for a better solution and gives a big achievement for healthcare industry to classify hyperlipemia with real time dataset by a simultaneous aided diagnosis model thereby achieving 90% of accuracy for diagnosis of outpatient doctors with the help of big data analytics (Hu et al., 2018).

Shuffle Phase

In shuffle phase, the values are listed with respect to keys. It is an intermediate phase between map function and reduce function.

Reduce Phase

In reduce phase, all the values are merged with respect to keys which are listed by shuffle phase using HTTP across the network. In the sorted inputs, reduce method is called by reduce(Object,Iterable,org.apache.hadoop.mapreduce.Reducer.Context) for each
<key, (collection of values)> pair.
The output of reducer is mentioned by a RecordWriter via TaskInputOutputContext.write(Object, Object) The above syntax has been referred from (Hadoop, 2019).
As shown in Figure 6, the ecosystem of Hadoop maintains its configuration with different tools in order to develop distributed coordination among jobs.

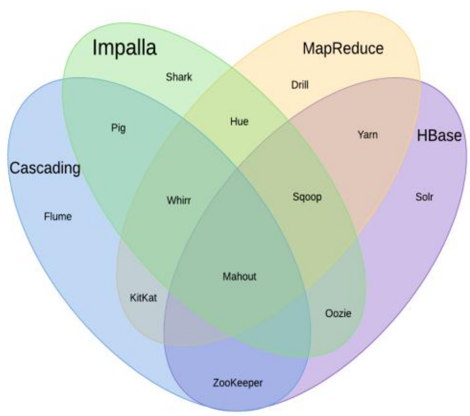

Figure 6. Ecosystem of Big Data

Spark

Open source Apache Spark is another tool of big data analytics which is much faster than Hadoop and is developed by Berkeley AMP Lab US in 2009. It runs on every platform like Hadoop, Apache Mesos, Kubernetes, standalone and for solving the biggest issues of spark, it maintains both batch and streaming data within memory cluster computing. The primary application programming interface RDD is the resilient distributed dataset (RDD) of SPARK (***Zaharia et al., 2016***). The frequency of words from a set of files is calculated by the following RDD-centric Scala programming. For streaming analytics, spark streaming performs RDD transformations and supports Kafka, Twitter, and Flume. Distributed machine learning memory-based spark based MLib (***Machine Learning Library***) is much faster than that of Apache Mahout. Distributed graph processing framework Graphx supports two parallel algorithms PageRank and MapReduce. Figure 7 gives a complete overview of spark architecture with different components to label the propagation.

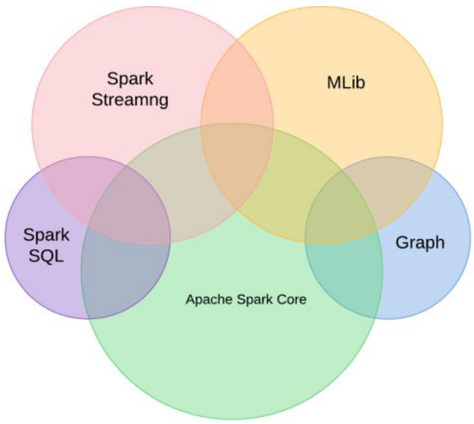

Figure 7. Components of Spark

RDD-centric scala programming:

```
val conf =newSparkConf().setAppName("file1_test")
valsc=newSparkContext(conf)
val data =sc.textFile("/path/to/dir1")
val tokens =data.flatMap(_.split(" "))
valwordFreq=tokens.map((_, 1)).reduceByKey(_ + _)
wordFreq.sortBy(s => -s._2).map(x => (x._2, x._1)).top(10)
```

Spark SQL

The primary application programming of spark SQL is written below. DataFrames introduced by Spark SQL are calculated in Scala, java, python and also support both structured as well as unstructured dataset.

```
import org.apache.spark.sql.SQLContext

valurl = "jdbc:mysql://yourIP:yourPort/test?user=yourUsername;password=yourPassword"
valsqlContext = new SQLContext(sc)

val df = sqlContext.read.format("jdbc").option("url", url) .option("dbtable", "employee").load()

df.printSchema()
valcountsBySalary = df.groupBy("salary").count()
```

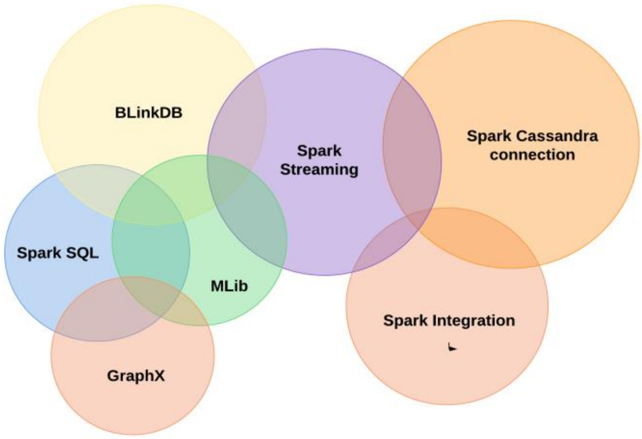

Figure 8. Ecosystem of Spark

Table1. Comparative study of Hadoop and Spark

	Spark	**Hadoop**
Year	2009	2006
Developed by	Apache	Apache
Type	Open source	Open source
Storage	HDFS, DBMS and Kafka	HDFS
Supporting language	Java, Scala, python, R	Java
I/O	In memory (cache)	Disk
Data structure	(key, value), RDD	(key, value)
Mode	Batch and stream	Batch
Machine learning framework	Spark Mlib	Mahout
Website	Spark.apache.org	Hadoop.apache.org
Services	• Driver program • Cluster manager • Worker nodes	• Namenode • Secondary name node • Job tracker • Data node • Task tracker

As shown in Table 1, the comparison between two important tools of big data is synthesized by us to improve the point of view of researchers for getting a valuable data processing tool. In Figure 8, the ecosystem of spark has been properly described for a better understanding of correlation between Spark SQL+Data frame, streaming, MLib, Graphx, Spark Cassandra connector and spark integration and, offers a currently needed tool for an appropriate research.

Role Of Big Data Analytics In The Field Of Healthcare Industry

Challenges in healthcare industry include monitoring, measuring and analysing vast amount of data which are gathered from upcoming areas of medical research to improve health outcomes. We need expertise to find value from big data analytics and improve care process. Basically, it changes the way of clinical observations.
There are several applications for big data in healthcare industry:

3.1 Personalized based treatment (Dilsizian & Siegel, 2014)

3.2 Analyzing EHRS (Electronic Healthcare Records) (Ghani et al., 2014)

3.3 Data privacy for medical research (Patil & Seshadri, 2014)

3.4 New viruses/ chronic disease management;

3.5 Cost reduction for better organization management (Bates et al., 2014)

3.6 Remote consultations (Wan et al., 1999)

Personalized Based Treatment

For analysing this huge amount of data, several difficulties come through the middle neck of research like move from traditional to personalized medicine by high throughput data storage and processing (Alyass et al., 2015). So, in the personalized based treatment phase, to reach to the right medicine at right time by accessing HPC(High-performance computing) and healthcare database is really an abecedarian step (Dilsizian & Siegel, 2014), that is why for a semantic data driven setting, different challenges of healthcare industry are solved by big data analytics with authorizing personalised medicine (Panahiazar et al., 2014). And healthcare gets better with convention/ proceeding through different big data analytical tools and the key to a personalized healthcare treatment is to consider and display the outcomes of applicability to a patient and to narrow focus on reducing the re-admission rates (Chawla & Davis, 2013). Big data analytics plays a vital role to analyse the heterogeneity of data and to find out the hidden part of the research. That is why chemotherapy is a benefit for one population and harmful for other and to assessing patient-specific risk of chemotherapy related hospitalization is analysed by clinical data that may be helpful for clinical experts for an improved analysis of disease network (Brooks et al., 2015). Big data is also saving many more lives in the whole world by giving them an appropriate medicine for specific disease which is at a rate of high risk. By implementing big data analytical methods, researchers classify patients according to their risk rate and also reduce the cost of high cost patient readmissions, triage and decompensation (Bates et al., 2014). Beyond different tools of big data analytics, machine learning provides an opportunity for personalized treatment. Researchers contemplated convolution auto encoder for feature learning of small growths on the lung for providing intelligent healthcare services to patient (Chen et al., 2017). Another use case of an abnormal feature detected in the abnormalities of chest by some automatic method is actually useful in fighting against TB (Tuberculosis) and brain tumor too may be cancerous and dangerous (Van Ginneken et al., 2002). Thus, CNN (*Convolutional* Neural Network) may detect the deformity change and resolved different methods like Gaussian –pyramid multi scale input feature –fusion technologies for early clinical trial diagnosis (Liu et al., 2018).

Analyzing Ehrs (Electronic Healthcare Records) (Ghani et al., 2014)

Large amount of data come from electronic health records and PAC (Picture Archiving and Communication System) from radiologic images that are frequently verified and analyzed by healthcare exporters to improve patient care (Ghani et al., 2014). Generally speaking, EMR (Electronic Medical Record) and sensor data are two categories of electronic health records. So, from a statistical point of view, unstructured data are available in a fluctuating manner with different sensor devices. Due to this situation, the first step is, storage have to maintain by one of the tools of big data analytics because it's beyond the control of traditional database management system. And also, predictive model which is created by machine learning methods is also help to analyzing heterogeneous healthcare data. According to medical treatment for an improved approach of cancer patients is a valuable step about applying right medicine at right time: precision medicine. If it moves forward to genome sequencing, it will get a platform for big data analytics in biology. Distributed and paralyzed data processing data of huge amount of bioinformatics and EHR data is need to generate key meta data, the pattern of series event should be included and now researchers begin to speed up the process of discovering new variant diseases with the help of big data analytics (Frey et al., 2014). The integration of both patient genotype and phenotype data to provide methodology using big data technology (Frey et al., 2014). The deep observable characteristics of organism increases with increasing in huge storage of genomic information and EHR in medicine, one of the energetic areas of research is intermix all the characteristics of an organism with total inheritance. For the classification of phenotype and genotype relationship is supported by semantic and sequential similarity analysis through use of big data analytic concept (Frey et al., 2014). Therefore, for advancing clinical and biomedical research, EHRs data comes from various personal devices. Biomedical research data and clinical data are typically captured for improved quality of research and also creating new methodologies for analyzing clinical data for patient consent and risk-based assessment through big data analytics (Margolis et al., 2014).

Data Privacy for Medical Research (Patil & Seshadri, 2014)

Patient and healthcare organization privacy issues in big data are the most complex challenges which can be properly solved by big data analytics. Patient privacy depends upon data security of sensitive information that nobody should know and healthcare organization privacy implies that organization secrets need to be protected. To increase the global health security, it is important for any healthcare organization to choose security tools of BDA as a service to setup security system which is beyond the capabilities of unauthorized individuals and organizations. New technologies like fog computing are helpful in solving the storage complexities of EMR data which are stored in the healthcare cloud (Dubey et al., 2015) and it also creates better algorithms to combine with map reduce and improve the privacy (Hashem et al., 2015; Patil & Seshadri, 2014). The collected biomedical data maintain the privacy of a patient that is related to social and technical challenges leading to influence healthcare (Weber et al., 2014). The security and privacy of a patient's data is much more difficult task. Hence, according to HIPAA (Health Insurance Portability and Accountability Act of 1996) regulation Act, data privacy and security measures are developed to maintain the health-related information of both patient and hospital with the help of BDA (Gahi et al., 2016). To improve the quality of data, researchers take a forward step towards analysis of data types, storage model and privacy to outline the different challenges of big data as per future trends (Lv et al., 2017). The data privacy and confidentiality mainly focus at if it is related to healthcare industry. In most cases, structured and unstructured clinical data originate from different sources whether it is from hospital or from social media (Hansen et al., 2014).

New Viruses/Chronic Disease Management

From latest survey of DONS (Disease Outbreak News) by world health organization, it is observed that a number of unknown viruses appear at an accelerating rate (WHO, 2018).

a. Yellow fever – Brazil, 27thfeb 2018;
b. Chikungunya –Mombase, Kenya;
c. Rift valley fever –Gambia 26thfeb 2018;
d. Human infection with avian influenza A(H7N4) virus –China 22th feb,2018;
e. Lassa fever-Liberia 22thfeb 2018;

Some of the chronic diseases such as heart diseases, cancer, diabetes, stroke and arthritis which is lasting for three months or more by the definition of the U.S. National center for health statistics (Wu et al., 2013)and these are not prevented by vaccines, neither cured by medications which is one of the major problems to be solved by healthcare leaders. As several reports reveal, challenging diseases create a proper solution for better understanding of serious disease states.
Now, the advanced data analytics can deliver evidence-based improvements in quality of care and identify the optimal performance effectively and efficiently (Bentzen, 1998). Major evidence –based medicine (EBM) is the integration of certain factors like patient values and expectation, individual clinical expertise and best external evidence (Sackett, 1997). Evidence-based radiation oncology is also a recent research for clinical development (Bentzen, 1998).

Cost Reduction for Better Organization Management (Bates et al., 2014)

To reduce readmission cost (Zolfaghar et al., 2013) and provide them with proper healthcare facilities is also one of the applications of big data analytics. How to identify at-risk patients by proper investigation of past history and reduce hospitalization costs by organization is also a management procedure of BDA. Faster treatment for at-risk chronic disease affected patients need to reduce health care cost because it exceeds the spending per capital for economic operation and development (OECD) (Viceconti et al., 2015). The great way to reduce cost of readmissions, decompensation of critical patients and improve treatment optimization is to solve through big data analytics (Bates et al., 2014). Apart from the current scenario of CMC-I Plus, big data provides best business support to hospital and medical for identifying fraud, waste and errors using predictive –modeling techniques for improving insurance claims (Srinivasan & Arunasalam, 2013). Using integrated dataset, risk of readmission for congestive heart failure (CHF) patient can be predicted with a scalable data mining model and NIS dataset (Zolfaghar et al., 2013). It is discovered that each year millions of dollars are wasted for readmission in healthcare. That is why researchers identify the high-risk patient and low risk patient to guide them at right time (Priyanka & Kulennavar, 2014). Patient protection and affordable care act or Obama care, is related to hospital readmission. The act is made to accept all the conditions to reduce extra cost in Medicare payment (Orszag & Emanuel, 2010). In 2010, the approximate cost of osteoporosis is €37 billion and in OECD (Organization for Economic Co-operation and Development) countries, USD $3395 per year spend in healthcare (Viceconti et al., 2015).

Remote Consultations (Wan et al., 1999)

The unstructured mass of data generated by micro level audio and video images shared between hospital and patient during remote consultation is now very efficiently handled by big data analytics and also diagnosis of patients based on remote consultation making a proper valuable decision by physicians. Real time remote consultation in the outpatient clinic experience at a teaching hospital (Wan et al., 1999), the patient’s activities produced by rich multimedia are processed and executed using real time data which is one of the benefits of big data analytics and Hadoop. Every acceptance of healthcare industry, electronic health and mobile health plays a major role for personalized treatment including mobile phones and other wireless devices. With this value-based care, online psychological therapy and remote video consultation can give precise treatment for mental health problem (Hollis et al., 2015). The medical co-ordination is maintained by rural patients, health workers and urban city specialists with targeted IOT for improving quality of care (Xu et al., 2014a). It is important to note that the real-time observation of ECG (Electrocardiogram) waveform and other vital signs keep doctors informed to take proper treatment. A cloud-based teleradiograph recently spotted either for medical or educational research opportunities, in the mean future teleconsultation to help and allow patients to track disease and avoid preventable emergency situation (Hsieh et al., 2013). The innovative process of BDA (Big Data Analytics) is more resourceful for Onsight doctor assortment (Hoens et al., 2013). Researchers proposed Multiple Kernel Learning (MKL) and Adaptive Neuro-Fuzzy Inference System (ANFIS) methods for classification of heart disease affected patients among normal healthy patients (Manogaran et al., 2018). For evaluating level of high-risk diabetic patients by soft computing prediction model with real time clinical data for preventing heart attack or stroke has been proposed in (Eswari et al., 2015) and for risk assessment process of observing the level of risk for diabetic patient by analyzing medical data using genetic algorithm is detailed in (Sabibullah et al., 2013). To reduce stress test in ED (Emergency Department) patients, one out of 50 patients needs to be treated for heart attack within 14 days (Hess et al., 2012). Growing cost of care, patient volumes with passage of the affordable care act, an aging population with high risk of disease, storage of medical experts and reducing in reimbursement are different major challenges of big data analytics (Nambiar et al., 2013). Different use cases of big data analytics in healthcare industry help in improving treatment of chronic diseases like cancer, effectiveness of treatment strategy and reducing readmission cost (Barrett et al., 2013). The set of circumstances needed for a better identification of both acute and chronic diseases represents a combination of traditional medical informatics, mobile health and social health records. For detecting and capturing voluminous streaming of healthcare data through implantable devices (within the body) used for controlling the life-threatening diseases can be achieved with the help of big data analytics (Andreu-Perez et al., 2015). Various major parameters such as physical activities, diet, tobacco consumption and exposure to pollution undergo some major actions by applying big data analytics for decreasing the risk levels of individuals/population for prevention of diseases (Barrett et al., 2013). Now, some of the leading chronic diseases like heart attacks and cancers negatively affect the growth of global economy. In 23 developing countries in 2005, 50% of diseases' burden were generated by chronic diseases and it has cost them more than $84 billion by 2015 (Nugent, 2008). Remote patient monitoring tends to become a leading chronic disease detector for easily detecting and monitoring the diseases at right time. The real time information generated from remote sensing devices are straight away deposited into HDFC (Wang et al., 2018). In a journey of development, most of critical element restrictions are solved by big data analytics and the reason to be concerned is to improve the diagnosis of clinical trial and most importantly, monitoring of patients (Diaz et al., 2012). A recent evolution of mobile telemedicine has taken advantage of traditional healthcare technology for patient monitoring without deprecation EMD (electro medic devices) using the RS232 interface for patient monitoring and utilizing through internet (Figueredo & Dias, 2004). In short, the importance of remote monitoring by analyzing the physiological factor of a disease is to alert the physician for immediate care (Schultz, 2013). The internet of things in healthcare industry is growing rapidly so that generated data are properly analyzed and stored into the cloud by big data analytics for an improvement of lives (Hossain & Muhammad, 2016). For monitoring cardiovascular patients, the way is open for prediction of risk by analyzing PAF (Patient Assement Form) data without coding of medical rules (Tseng et al., 2008). Puers projected traditional patient monitoring system in 1989 (Puers et al., 1989).

Healthcare Difficulties That Can Be Overcome Using Big Data Analytics

Overcoming difficult situations like newly added diseases is a leading challenge for society. Researchers focus on personalized treatment for rare diseases which is beyond the capacity of traditional healthcare system. Currently, Big Data Analytics improves health outcomes by identifying and developing a data model. The centers for disease control and prevention (CDC) (Harford, 2014), for example, has been using Big Data as a defense against Ebola (Bates et al., 2014) and other pandemics, which merges population migration in real time in order to track epidemics (Mitchell, 2014). Surveillance of disease progression and prevention is the best way to control Ebola like serious diseases. Centers for disease control and prevention (CDC) controls a tool called Bio-Mosaic that is indifferent to predict the spread of diseases (Mitchell, 2014). The detection of 2014 Ebola virus outbreak in West Africa: Guinea, Liberia and Sierra Leone by utilizing real-time data is one of the case studies of improved analytics. To stop these infectious diseases, WHO (WHO, 2018) has published advices. Brutal disease like Ebola is one of the world's deadest diseases (Anema et al., 2014). To quickly track the spread of influenza across the US, the team of researchers from google and CDC found what people tend to search related to flu symptoms (Harford, 2014).

Saving Life by Big Data Analytics (Marr, 2018)

According to Forbes article, with the help of combination of machine learning with Big Data Analytics, an algorithm has been derived to find out predicted future admission trends and the daily /hourly expectation rate of patient on each hospital and solve the problem of admission rate of assistance publique-Hopitaux de Paris. The combination of both IT and medical science forms a new technology i.e., medical informatics, which can improve the quality of healthcare system (Dimitrov, 2016). For tackling outbreak of serious diseases like influenza attack by analyzing patterns of diseases captured by huge amount of data from different population, various tools of big data analytics can be useful. In lieu of this fact, the healthcare sector can grow much faster with different tools of big data analytics for a unbelievable result (Feldman et al., 2012). According to a Forbes article, Big Data Analytics evaluates the numbers of expected patients at each hospital of assistance publique – hopitaux de Paris (AP-HP) for quality of care at right time (Marr, 2018; Datapine, 2019). For providing better medical resources to each patient, Hospitals of Paris has played a lead role in fighting with cancer and creating personalized medicine with time series analysis technique.

Big Data Is Helping to Prevent Opioid Abuse (Datapine, 2019)

The accidental deaths in the US are mostly caused by overdoses of opioids (by datapine.com). According to Forbes article (Datapine, 2019): Canada has declared opioid abuse to be "national health crises". Improving accuracy and determining the risk of opioids overdose can be achieved by capturing different source of analyzed data (Karanges et al., 2016). Prescribed use of opioid in Australia for non-cancer pain is increasing as compared to US and Canada (Karanges et al., 2016). Much more efforts have been put into for tackling of this prescribed use of opioid misuse by young adults. Now, it becomes a global health concern for *each and* everyindividual (Lankenau et al., 2012). Opioid risk factor for opioid misuse patient in the medical record of US is low as compared to real factor. Opioid drug dependency is much higher than mentioned factor written in medical record of US. Because of this, outpatient individuals mainly depend upon prescribed opioid for pain relief (Boscarino et al., 2010). In the study of Birnbaum's article, the societal opioid abuse burden for US govt has been detailed (Birnbaum et al., 2011). Jennifer Bresnick mentioned in her article about the funding assigned to states of Columbia to track and monitor drug overdose and explained the overview of CDC about the impact of big data analytics for right against opioid misuse (Bresnick, 2017). According to Forbes, major problem of down flow of social economic status for last 5 years is due to opioid overdoes crisis. John Kelley Community Voice Forbes Technology Council mentioned in its article that using Big Data Medical Analytics to address the opioid crisis, $78 billion were misused per year in the form of social and economic welfare in the US and as a result of that, 64,000 people died of opioid overdose. To decrease the role of nonfatal opioid misuse, US government takes some major challenges regarding prescription & illicit drug by less addictive formulation (Kelley, 2017). one of the analytics expert Bernard Marr writes in Forbes article about opioid abuse, the former president Obama earmarked $1.1 billion dollar for overcoming from drastic situation of opioid abuse (Marr, 2018). The prescribed opioid creates a big-difficulty for economic and behavioral health of US in 2015 (Han et al., 2017). According Robert Glatter, MD Contributor, Pharmacy & Healthcare writes in an article of Forbes magazine about opioid crises that "can vending machines that Dispense prescription opioids Help addicts survive the opioid crisis?"(Glatter, 2018).

Prevent Unnecessary ER Visits (Wells et al., 2016)

Using big data analytics, unnecessary emergency room visits for non-urgent patients decreases and increases the emergency room visits for serious patients, according to central goal of the affordable care Act (Mechanic et al., 2012). Big data analytics gives the platform for physician and patient a quality of care at a clinic or at home. By virtue of this, one serious patient gets total satisfaction. Patient health condition tends to change along with point of alert / or by notification from oncologist and data scientist. Unfortunately, the complexities of overdose prevention in the last two decades have become headache for various countries. But now due to Big Data Analytics, countries expect to overcome this opioid related death by analysing different health related programs, collaborative and evidence –based approach, reporting and evaluation, promoting safer drug use policy process to save lives (Wells et al., 2016). "Big Data can help in ER visit" by Jennifer Goforth Gregory Technical Writer, writes about reducing emergency room visits and to $2billion of needless medical cost. Actually, visiting emergency room is a stressful experience for everyone if someone has a chronic disease and requires frequent care, then that patient needs some extra suggestion to avoid unnecessary pay-of-pocket cost. According to CDC (disease control and prevention), 136 million visits to emergency room have been reported in last few years. The unnecessary trips to hospital are 1.3 million recently recorded by Minnesota department of Health (Gregory, 2016).To prevent unnecessary ER (emergency room) visiting by sharing patient to ER staff, it is necessary to provide medical health facilities by identifying high risk patient and that can be managed by programs like pre manage ED that is one of the best way of big data analytics to the help the doctors to be notified in real time. So, the recovery rate becomes closer (TechplusHealth, 2016).

However, all the policy modifications in the US healthcare are meant only for the cost control (Mechanic et al., 2012). CMS (Centre for Medicare and Medicaid services) claims that quality of data is available for analysis by Affordable Care Act for a valuable result using big data analytics (Roski et al., 2014). The shortened name of Patient Protection and Affordable Care Act (PPACA) or the Affordable Care Act (ACA) or Obamacare is a multiple level policy of US laws by 11th US congress that marked by Ex-former president Barack Obama, and main goal behind PPACA is to trim down cost, enhanced patient care and to augment the insurer (Bunolna, 2015). According to indiscriminate study in Oregon, a greater number of patients are going to ED due to Affordable Care Act and ED is used by new insured patient compared to uninsured up to 40% (Nave, 2014).

Hadoop Technology in Cancer Treatments and Genomics

Massive amount of data by genome studies causes problem for professionals and researchers to be organized in a useful way in eliminating cancer. Patients require proper individual treatment by early detection of disease. Thus, the diagnosis and prognosis of cancer by doctors/researchers for an incredible huge amount of pet bytes of genomic datasets couldn't be organized by traditional technology. Hadoop is one which will help efficiently to manage the genomic datasets and get a change to increase the survival rate of cancer. According to SEER (surveillance, epidemiology and end result)-Medicare health outcome survey (MHOS), the death rate of cancer patient properly reduced through quality of analysis process of unstructured data. (Kale & Patil, 2016) Now Hadoop and MapReduce, the two main tool of Big data could helping for penalization and mapping for analyzing huge amount of genome data, and giving personalized medication each cancer patients is main objective of big data in healthcare to develop(Sotiriou & Pusztai, 2009).The main motivation of Bioscience is to help patient for fighting with cancer by identification the genetic mutation (Positive Bioscience, 2019), meanwhile genomic sequencing with Big Data Analytics could change the expected personalized care process in a positive manner to reduce death rate by monitoring clinical trial because particularly some patient may respond to treatment and giving a precious result for researchers(Bhardwaj et al., 2014).High -speed analysis of multiple gene is a source of big data because of its density, variety and relevance. The repercussion of DNA (Deoxyribonucleic acid) sequencing for healthcare system and policy issues: more patient, developing coverage and reimbursement policy and cost (Phillips et al., 2014). The 5V (*Marr*, *2014*) characteristics of big data also helping chemotherapy treatment for a positive respond of a cancer patient because of heterogeneous data generated from demography, chemo dosimetry, multimodality of image feature related to 5V. These different draws near to progress for mining of big data in oncology (El Naqa, 2016).Let's explore the field of biomedical and molecular oncology to providing large amount of unstructured data for various research program and different software platform like LAS (laboratory assist suite) to integrate vast quantity of biomedical data to present a graphical tool for improvement of data quality(Baralis et al., 2012).For reducing the probability of birth defects with respect to new drug measure the prediction of cancer which is helpless for traditional computing method, so; Spark is one of the tools of big data analytics for analyzing human genome (Ding et al., 2016).By consistently analyzing Big Data related to acquisition, storage, distribution and analysis to meet the different limitation for touching the genomic poses to overcome from different challenges because Genomics is a Big Data science(Stephens et al., 2015).To find cancer-causing variation , and uncovering trends by analyzing large scale genomic data with distributed and parallelized data processing techniques of Apache Hadoop(O'Driscoll et al., 2013).Behind the search for cancer alleviate, OBAMA'S precision medicine initiative with intensifying clinical trial with analyzing 3 billion of DNA base pair and determine why the cancer mutated in different through Hadoop technology by taking preventive care, the pin point treatment option are categorized for mapping of 100 genome through map reduce. Disease Prevention or treatment by identifying cancer gene and identify useful drug through genome sequencing using Big Data Analytics through analyzing 20 billion of raw complex healthcare data By CASI (Complex Adaptive Systems Initiative) and developing genome data lake for the complex adaptive system initiative at the Arizona state university through big data analysis (Dezyre, 2017).

The Role of Hadoop Technology in Hospitalization (Dezyre, 2017)

Now, technology of big data analytics changes the strategy of hospital industry. The very unknown Hadoop open source framework could help the unorganized and unstructured data into organized and structured data by map reduce which is a part of Hadoop. Despite the next important point to mapping and storing a large volume of data by Hadoop, physicians/researchers also manage big amount of data gathered from hospitals which helps them to reduce and faster the time of treatment. Fighting the deadly diseases, the total expenditure amounts up to $1.2 billion for generating 150 Exabyte of data. To reduce total health management expenses about $300 billion - $500 billion by using big data in healthcare projected by McKinsey (Dezyre, 2017). Challenges like data knowledge representation, database design, data querying and clinical support are managed and getting a maximum rate of accuracy through big data analytics(Dezyre, 2017).Most of the data are stored in printed form but in case of analyzing real data from electronics health record to find out the top treatment procedure for different types of patients(Marr, 2015).It is pretty much different for conventional hardware and software to managing like huge amount of 79% unstructured with high volume of data and totally depends upon both the patient and their wellbeing with respect to refund models by posing vital challenges (Golub et al., 1999; Kale & Patil, 2016).

Hadoop Technology in the Field of Healthcare Intelligence (Preuveneers et al., 2016)

There are several challenges faced by healthcare industry. As Hadoop comes into middle of the stage to store the tremendous generated data, process and save more life. Up till now, due to rapid digitization of data, many recent technologies like a multi-disciplinary field Ambient assisted living (AAL) in personal health care and tele healthcare system, remote monitoring system in health, can address the main barriers of life long illness (Preuveneers et al., 2016). The large-scale integration of rapid digitization of data, researchers present SAMURAI, a batch streaming architecture to expose the big data scalability and positive feedback of users (Preuveneers et al., 2016). More than 75% of data are unstructured from disparate sources that have been successfully used by big data analytics for tracking diseases outbreak and improve patient care. Researchers are still discovering different techniques for improving patient satisfaction with the help of newly advanced technologies like mobile e-health application that comes to the use of real-time data and verifying different parts by explaining different work steps of cloud components including deep learning, indexing food images, segmentation, monitoring calorie-intake by cloud broken model which is used to process millions of healthcare images' parallelism (Peddi et al., 2017). Checking on high risk patient wearable medical devices with sensors actually over exceed the generation/gathering of data at a continuous manner called Big Data, sensors implanted on body for measuring respiratory rate, heart rate, BP, body temp and blood sugar. If the normal value exceeds a threshold, then message is sent to doctor through wellness network with security mechanism to protect data (Manogaran et al., 2017). Effective management of healthcare insurance can be achieved by using Hadoop technology for processing of huge data set for haul out meaningful information. Sunil Kakre, director of IT, Dignity Health, spoke in a summit about Hadoop relating with Healthcare analytics in recent Hadoop. Healthcare investment can be optimized when the recovery rate of critical time sensitive diseases like cancer, AIDS is combined faster with Hadoop's distributed approach (Dezyre, 2017).

Future Research Directions Of Big Data Analytics In Healthcare Industry:

To address the future of big data analytics, it is necessary to classify different characteristics of huge amount of data i.e. Big Data. Managing sensor data generated from heart rate, BP rate, pulse oximeters (Han et al., 2017), distance walked, and glucose monitor etc. can be eased by adding new IOT devices that gives a future scope to researchers in medical sector for increasing the survival rate (Xu et al., 2014b). Another major issue is wrong medicine prescribed by professionals by mistake. Big data analytics gives the platform to analyse the user data and the prescribed medicine (Saratchandran, 2018; Ullah et al., 2017). For high risk patient, big data analytics can tend to increase the quality of care by avoiding extra time and to analyse physiological data as well (Stivoric et al., 2003). Doctors determine the risk data. The numerical factors of newly added rare diseases become more and more critical due to limitations in effective treatment planning and lack of smarter healthcare system that reduces the rate of death during the follow-up period which is a major future problem for researchers. Now, Alzheimer's (AD) and Parkinson's (PD) diseases are also under control frequently finding neurodegenerative diseases (Mesholam et al., 1998) and it affects around 80 million people worldwide. How to control these diseases by making predictions and determine the rate of risk is also a future research topic around the researchers by providing a solid big data visualization tool (Nalls et al., 2014; Haas et al., 2016). Some of the studies, which are now in progress are like to identify and manage the high risk and cost patient by David W. Bates (Bates et al., 2014) and another one gives stress upon the diagnosis model for outpatient departs via healthcare big data analytics (Hu et al., 2018). In addition, the ongoing future research is the healthcare informatics defining the logistic technique for decision making by physicians and patients. Basically, the main aim of the machine learning is to learn machine to predict clinical decision by physicians.

Creating a vital personalized care for each and every patient becomes a local mode of concern in researcher's mind. Offering all details on incidents of cancer diseases, researchers discover new drug using diversified types of colossal volume of data. Verification of responses of personalized medicine can be made using Big Data (Xu et al., 2014b). Changing hazard of CKD (CHRONIC KIDNEY DISEASE) by using Bayesian multi resolution hazard modeling for survival analysis, the HER system is required to be analyzed which becomes a challenge related to health care organization (Hagar et al., 2014). Survival scrutiny is a statistical technique of survival time with respect to certain treatment inward bound. This could be obtained through triumphant research using Big Data method for different risky patient. It's an enhanced cleverness. The overall advancement of Hadoop and Apache Hive is being used for data summarization, query and analysis for traumatic brain injury (TBI)patients. By analysis of modal operational data sets, the endurance transience and morbidity rates built-in into all TBI clinical trial can necessarily reduce the misclassify rate and misdiagnosis rate (Rodger, 2015; Han et al., 2017). For identifying potential health problem, both Alzheimer's disease (AD) and Parkinson's disease (PD), olfactory dysfunction was found including order identification, recognition, and detection of threshold by analyzing large online dataset through non parametric analysis (Mesholam et al., 1998). The outpatient engagement in China increased due to newly added Health insurance schemes and diminution diagnostic timing. Classification using SVM (Support Vector Machine) and NN (neural network) for featured extraction by referring to several key attributes of Big Data analytics, technique like diagnosis is employed to increase outpatient survival rate (Hu et al., 2018). With its diversity formats, now different sensors are used for dynamically monitoring the physiological behaviors (Stivoric et al., 2003). Big data related to different wireless sensor technologies can be helpful for providing a prediction of disease at early stage (Raghupathi & Raghupathi, 2014) where wrong medicine or medication taken by mistakenly is a consequence of continuing error created by professionals. Keeping in mind both physical and patient concrete goals, big data analytic enters into healthcare industry in order to reduce mistakes and save lives (Saratchandran, 2018). Introducing influence tools in medicinal practice in protective decisive circumstance and ambiguity due to measurement errors, missing data, or errors in coding, the information pertaining to the medical Big Data are exaggerated (Lee & Yoon, 2017).

Symptoms of diseases are traced by assorted IOT devices thereby analyzing them by Big Data Analytics for a precious treatment in healthcare industry, care in real time should be a major clinical need to prescribed medicine at right time that is now a focusing clinical movement (Ullah et al., 2017). Large scale meta-analysis of Parkinson's diseases has been carried out across 24 loci (locus/ location of gene) and identifying risk loci have been combined with DNA methylation (Nalls et al., 2014). That infers to a fact that clinical development programs related to chronic diseases can be organized for making healthcare more opportune. A lot of times, many technologies mutate the direction of treatment by entering the newly added diseases. For Alzheimer's diseases, researchers create statistical modelling and simulation for better medication through widen antibodies that is not in favour of beta – amyloidal (Haas et al., 2016). These are the working areas to develop using big data analytics and predicting healthcare frauds where no fraud can be spotted through classification (Berry & Linoff, 1997) but used as traditional statistical techniques. Researchers need new technologies like AI and ML or different tools like Hadoop or Spark for getting full benefits of personal healthcare, relative output, reducing readmission rates are measured by introducing new techniques of big data analytics (Chawla & Davis, 2013).

Researchers are also exploring through different tools related to big data analytics which creates opportunities to discover new drugs in the clinical trial.

Table 2. Comparative study of different ecosystem of Hadoop

Sl no	Tool	Developed by	Purpose	Year	Language
1	Hadoop	Apache	Collection of open source software	2011	Java
2	Sqoop	Apache	Command line interface application	2015	Java
3	Pig	Apache	High level platform for creating program	2008	Java
4	Hive	Apache	Data warehouse	2017	Java
5	Zookeeper	Apache	Centralized service for distributed system	2017	Java
6	Cassandra	Apache	Distributed NoSQL database mgt system	2008	Java
7	Avro	Apache	Data serialization system	2016	Idl (interface description language)
8	Oozie	Apache	Server based workflow scheduling system	2017	Java
10	Talend	Talend company	Software integration vendor	2005	JavaScript
11	Openrefine	Bsd	Open source desktop application	2010	Java
12	Solr	Apache	Open source enterprise search platform (wiki)	2017	Java
13	Mahout	Apache	Collaborative filtering, clustering, and classification	2017	Java/Scala
14	Statwing	Qualthics	Web-based statistical analysis	2012	
15	Storm	Apache	Distributed streaming processing	2017	Java
16	Kafka	Apache	Stream processing software platform	2011	Java/Scala
17	Flink	Apache	Distributed streaming dataflow engine	2018	Java/scala
18	Redis	Bsd	Open source in memory database project	2009	Database project
19	Impala	Apache	Parallel processing SQL query engine	2012	C++
20	Spark	Apache	Cluster computing framework	2014	Scala, java, python, R
21	Flume	Apache	Distributed, reliable service, based on streaming dataflow	2017	Java
22	Pentaho	Pentaho corp.	Business intelligence software	2016	Java
23	Mongo db	Mongo dbinc	Document oriented database program	2009	C++ & JavaScript
24	Whirr	Apache	For running cloud services	2007	Python
25	Weka	University of waikato	Machine learning	2017	Java
26	Hbase	Apache	Distributed database modelled	2018	Java
27	Chukwa	Apache	Open source data collection	2010	Java

In Table 2, we present a summary of different ecosystem of Big Data for a proactive platform to maintaining tools in the real time analyzing huge datasets.

Conclusion

The end point of this study is what techniques are used in this huge amount of unstructured big data which are generated from variety of different sources of healthcare industry. This topic/article nevertheless imposes complex challenges pertaining to healthcare industry and evaluates the risk of newly added diseases. It also analyses and determines the right analytical method making use of different tools of big data analytics. For the healthcare rich domain, this study summarizes the overall study of big data analytics including storage, processing, security and predicting patient risk for several chronic diseases for this massive variety of bioinformatics heterogeneous data.
The results of this whole study clearly mention the scenario of a new era of big data analytics for the healthcare domain and give the procedures how to provide a quality healthcare by predictive analytics tools. Despite the limitations, it's a very good ongoing process. This article refers to the researchers' effective meta version of big data analytics tools used for healthcare informatics for the better healthcare practices performance.

References

[1] Zhang, H., Chen, G., Ooi, B. C., Tan, K. L., & Zhang, M. (2015). In-memory big data management and processing: A survey. IEEE Transactions on Knowledge and Data Engineering, 27(7), 1920-1948.

[2] Gkoulalas-Divanis, A., Loukides, G., & Sun, J. (2014). Toward smarter healthcare: Anonymizing medical data to support research studies. IBM Journal of Research and Development, 58(1), 9-1.

[3] Oracle Health Sciences. Adopt Actionable Analytics Enabled by Data Aggregation and Integration, Risk Stratification and Visualization of Enterprise Data. Retrieved on 10 January 2019 from: http://www.oracle.com/us/industries/health-sciences/healthcare-analytics-info-2660933.pdf

[4] Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T., & Yang, G. Z. (2015). Big data for health. IEEE journal of biomedical and health informatics, 19(4), 1193-1208.

[5] Big Data Analytics, what it is and why it matters. SAS Institute Inc. Retrieved on 10 January 2019 from: https://www.sas.com/en_in/insights/analytics/big-data-analytics.html.

[6] Marr, B., 2014. Big data: The 5 vs everyone must know. LinkedIn Pulse, 6. Retrieved from: https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/

[7] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), 1.

[8] Tian, D., Zhou, J., Wang, Y., Lu, Y., Xia, H., & Yi, Z. (2015). A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics. IEEE transactions on intelligent transportation systems, 16(6), 3033-3049.

[9] Lin, K., Luo, J., Hu, L., Hossain, M. S., & Ghoneim, A. (2016). Localization based on social big data analysis in the vehicular networks. IEEE Transactions on Industrial Informatics, 13(4), 1932-1940.

[10] Grier, S., & Bryant, C. A. (2005). Social marketing in public health. Annu. Rev. Public Health, 26, 319-339.

[11] Yin, Z., Yu, F. R., Bu, S., & Han, Z. (2015). Joint cloud and wireless networks operations in mobile cloud computing environments with telecom operator cloud. IEEE Transactions on Wireless Communications, 14(7), 4020-4033.

[12] Wen, Y., Zhu, X., Chen, C., & Rodrigues, J. J. P. C. (2014). Mobile cloud media: Reflections and outlook. IEEE Transactions on Multimedia, 16(4), 885-902.

[13] Abolfazli, S., Sanaei, Z., Alizadeh, M., Gani, A., & Xia, F. (2014). An experimental analysis on cloud-based mobile augmentation in mobile cloud computing. IEEE Transactions on Consumer Electronics, 60(1), 146-154.

[14] Smith, M., Szongott, C., Henne, B., & Von Voigt, G. (2012, June). Big data privacy issues in public social media. In 2012 6th

IEEE International Conference on Digital Ecosystems and Technologies (DEST) (pp. 1-6). IEEE.

[15] Lee, K., Agrawal, A., & Choudhary, A. (2013, August). Real-time disease surveillance using twitter data: demonstration on flu and cancer. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1474-1477).

[16] Liu, C., Chen, J., Yang, L. T., Zhang, X., Yang, C., Ranjan, R., & Kotagiri, R. (2013). Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates. IEEE Transactions on Parallel and Distributed Systems, 25(9), 2234-2244.

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

[18] Gillingham, P., & Graham, T. (2017). Big data in social welfare: The development of a critical perspective on social work's latest “electronic turn”. Australian Social Work, 70(2), 135-147.

[19] Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—harnessing the Web for public health surveillance. The New England journal of medicine, 360(21), 2153.

[20] Xu, Z., Liu, Y., Mei, L., Hu, C., & Chen, L. (2015). Semantic based representing and organizing surveillance big data using video structural description technology. Journal of Systems and Software, 102, 217-225.

[21] Manogaran, G., & Lopez, D. (2018). Disease surveillance system for big climate data processing and dengue transmission. In Climate Change and Environmental Concerns: Breakthroughs in Research and Practice (pp. 427-446). IGI Global.

[22] IDC. (2018). Guide to citing internet sources [Online]. Available: Retrieved from: https://www.idc.com/getdoc.jsp?containerId=prUS42371417

[23] Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. IEEE transactions on knowledge and data engineering, 26(1), 97-107.

[24] Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., & Vasilakos, A. V. (2017). A manufacturing big data solution for active preventive maintenance. IEEE Transactions on Industrial Informatics, 13(4), 2039-2047.

[25] Apache Hadoop. 2015. Retrieved from: http://hadoop.apache.org/

[26] Hadoop Distributed File System (HDFS™). Retrieved on 15 January 2019 from: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

[27] Yang, G. (2011, October). The application of mapreduce in the cloud computing. In 2011 2nd International Symposium on Intelligence Information Processing and Trusted Computing (pp. 154-156). IEEE.

[28] McIntyre, R. S., Cha, D. S., Jerrell, J. M., Swardfager, W., Kim, R. D., Costa, L. G., ... & Brietzke, E. (2014). Advancing biomarker research: utilizing ‘Big Data’approaches for the characterization and prevention of bipolar disorder. Bipolar disorders, 16(5), 531-547.

[29] Dong, X. L., & Srivastava, D. (2013, April). Big data integration. In 2013 IEEE 29th international conference on data engineering (ICDE) (pp. 1245-1248). IEEE.

[30] Online Source, Retrieved on 15 January 2019 from: https://hadoop.apache.org/docs/r2.7.0/api/org/apache/hadoop/mapreduce/Mapper.html#run(org.apache.hadoop.mapreduce.Mapper.Context

[31] Hu, Y., Duan, K., Zhang, Y., Hossain, M. S., Rahman, S. M. M., & Alelaiwi, A. (2018). Simultaneously aided diagnosis model for outpatient departments via

healthcare big data analytics. Multimedia Tools and Applications, 77(3), 3729-3743.

[32] Hadoop Map Reduce, Retrieved on 15 January 2019 from: https://hadoop.apache.org/docs/r2.7.0/api/org/apache/hadoop/mapreduce/Reducer.html

[33] Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Ghodsi, A. (2016). Apache spark: a unified engine for big data processing. Communications of the ACM, 59(11), 56-65.

[34] Dilsizian, S. E., & Siegel, E. L. (2014). Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Current cardiology reports, 16(1), 441.

[35] Ghani, K. R., Zheng, K., Wei, J. T., & Friedman, C. P. (2014). Harnessing big data for health care and research: are urologists ready. European urology, 66(6), 975-977.

[36] Patil, H. K., & Seshadri, R. (2014, June). Big data security and privacy issues in healthcare. In 2014 IEEE international congress on big data (pp. 762-765). IEEE.

[37] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123-1131.

[38] Wan, A. C. T., Gul, Y., & Darzi, A. (1999). Realtime remote consultation in the outpatient clinic—experience at a teaching hospital. Journal of Telemedicine and Telecare, 5(1_suppl), 70-71.

[39] Alyass, A., Turcotte, M., & Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. BMC medical genomics, 8(1), 33.

[40] Panahiazar, M., Taslimitehrani, V., Jadhav, A., & Pathak, J. (2014, October). Empowering personalized medicine with big data and semantic web technology: promises, challenges, and use cases. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 790-795). IEEE.

[41] Chawla, N. V., & Davis, D. A. (2013). Bringing big data to personalized healthcare: a patient-centered framework. Journal of general internal medicine, 28(3), 660-665.

[42] Brooks, G. A., Kansagra, A. J., Rao, S. R., Weitzman, J. I., Linden, E. A., & Jacobson, J. O. (2015). A clinical prediction model to assess risk for chemotherapy-related hospitalization in patients initiating palliative chemotherapy. JAMA oncology, 1(4), 441-447.

[43] Chen, M., Shi, X., Zhang, Y., Wu, D., & Guizani, M. (2017). Deep features learning for medical image analysis with convolutional autoencoder neural network. IEEE Transactions on Big Data.

[44] Van Ginneken, B., Katsuragawa, S., terHaarRomeny, B. M., Doi, K., & Viergever, M. A. (2002). Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE transactions on medical imaging, 21(2), 139-149.

[45] Liu, J., Chen, F., Pan, C., Zhu, M., Zhang, X., Zhang, L., & Liao, H. (2018). A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas. IEEE Transactions on Biomedical Engineering, 65(9), 1943-1952.

[46] Frey, L. J., Lenert, L., & Lopez-Campos, G. (2014). EHR big data deep phenotyping. Yearbook of medical informatics, 23(01), 206-211.

[47] Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., ... & Green, E. D. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. Journal of the American Medical Informatics Association, 21(6), 957-958.

[48] Dubey, H., Yang, J., Constant, N., Amiri, A. M., Yang, Q., & Makodiya, K. (2015). Fog data: Enhancing telehealth big data

through fog computing. In Proceedings of the ASE bigdata & socialinformatics 2015 (pp. 1-6).

[49] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. Information systems, 47, 98-115.

[50] Weber, G. M., Mandl, K. D., & Kohane, I. S. (2014). Finding the missing link for big biomedical data. Jama, 311(24), 2479-2480.

[51] Gahi, Y., Guennoun, M., & Mouftah, H. T. (2016, June). Big data analytics: Security and privacy challenges. In 2016 IEEE Symposium on Computers and Communication (ISCC) (pp. 952-957). IEEE.

[52] Lv, Z., Song, H., Basanta-Val, P., Steed, A., & Jo, M. (2017). Next-generation big data analytics: State of the art, challenges, and future research topics. IEEE Transactions on Industrial Informatics, 13(4), 1891-1899.

[53] Hansen, M. M., Miron-Shatz, T., Lau, A. Y. S., & Paton, C. (2014). Big data in science and healthcare: a review of recent literature and perspectives. Yearbook of medical informatics, 23(01), 21-26.

[54] World Health Organization. (2018).” Disease Outbreak News” [online]. Available: http://who.int/csr/don/en/.

[55] Bentzen, S. M. (1998). Towards evidencebased radiation oncology: improving the design, analysis, and reporting of clinical outcome studies in radiotherapy. Radiotherapy and oncology, 46(1), 5-18.

[56] [Sackett, 1997] Sackett, D.L., 1997, February. Evidence-based medicine. In Seminars in perinatology (Vol. 21, No. 1, pp. 3-5). WB Saunders.

[57] http://doi.org/10.1016/s0146-005(97)evidence –based medicine David

[58] Zolfaghar, K., Meadem, N., Teredesai, A., Roy, S. B., Chin, S. C., & Muckian, B. (2013, October). Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In 2013 IEEE International Conference on Big Data (pp. 64-71). IEEE.

[59] Viceconti, M., Hunter, P., & Hose, R. (2015). Big data, big knowledge: big data for personalized healthcare. IEEE journal of biomedical and health informatics, 19(4), 1209-1215.

[60] Srinivasan, U., & Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs. IT professional, 15(6), 21-28.

[61] Priyanka, K., & Kulennavar, N. (2014). A survey on big data analytics in health care. International Journal of Computer Science and Information Technologies, 5(4), 5865-5868.

[62] Orszag, P. R., & Emanuel, E. J. (2010). Health care reform and cost control. N Engl J Med, 363(7), 601-603.

[63] Hollis, C., Morriss, R., Martin, J., Amani, S., Cotton, R., Denis, M., & Lewis, S. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. The British Journal of Psychiatry, 206(4), 263-265.

[64] Xu, B., Da Xu, L., Cai, H., Xie, C., Hu, J., & Bu, F. (2014). Ubiquitous data accessing method in IoT-based information system for emergency medical services. IEEE Transactions on Industrial informatics, 10(2), 1578-1586.

[65] Hsieh, J. C., Li, A. H., & Yang, C. C. (2013). Mobile, cloud, and big data computing: contributions, challenges, and new directions in telecardiology. International journal of environmental research and public health, 10(11), 6131-6153.

[66] Hoens, T. R., Blanton, M., Steele, A., & Chawla, N. V. (2013). Reliable medical recommendation systems with patient privacy. ACM Transactions on Intelligent Systems and Technology (TIST), 4(4), 1-31.

[67] Manogaran, G., Varatharajan, R., & Priyan, M. K. (2018). Hybrid

recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. Multimedia tools and applications, 77(4), 4379-4399.

[68] Eswari, T., Sampath, P., & Lavanya, S. (2015). Predictive methodology for diabetic data analysis in big data. Procedia Computer Science, 50, 203-208.

[69] Sabibullah, M., Shanmugasundaram, V., & Priya, R. (2013). Diabetes patient's risk through soft computing model. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2(6), 60-65.

[70] Hess, E. P., Knoedler, M. A., Shah, N. D., Kline, J. A., Breslin, M., Branda, M. E., ... & Stiell, I. G. (2012). The chest pain choice decision aid: a randomized trial. Circulation: Cardiovascular quality and outcomes, 5(3), 251-259.

[71] ambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013, October). A look at challenges and opportunities of big data analytics in healthcare. In 2013 IEEE international conference on Big Data (pp. 17-22). IEEE.

[72] Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013). Big data and disease prevention: from quantified self to quantified communities. Big data, 1(3), 168-175.

[73] Nugent, R. (2008). Chronic diseases in developing countries: health and economic burdens. Annals of the New York Academy of Sciences, 1136(1), 70-79.

[74] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, 3-13.

[75] Diaz, M., Juan, G., Lucas, O., & Ryuga, A. (2012, July). Big data on the internet of things: An example for the e-health. In 2012 sixth international conference on innovative mobile and internet services in ubiquitous computing (pp. 898-900). IEEE.

[76] Figueredo, M. V. M., & Dias, J. S. (2004, September). Mobile telemedicine system for home care and patient monitoring. In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Vol. 2, pp. 3387-3390). IEEE.

[77] Schultz, T. (2013). Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle. Bulletin of the American Society for Information Science and Technology, 39(5), 34-40.

[78] Hossain, M. S., & Muhammad, G. (2016). Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring. Computer Networks, 101, 192-202.

[79] Tseng, V. S., Chen, L. C., Lee, C. H., Wu, J. S., & Hsu, Y. C. (2008, March). Development of a vital sign data mining system for chronic patient monitoring. In 2008 International Conference on Complex, Intelligent and Software Intensive Systems (pp. 649-654). IEEE.

[80] Puers, B., Sansen, W., & Leuven, K. U. (1989, May). Patient monitoring systems. In Proceedings. VLSI and Computer Peripherals. COMPEURO 89 (pp. 3-152). IEEE.

[81] Harford, T. (2014). Big data: A big mistake?. Significance, 11(5), 14-19.

[82] Mitchell, B. (2014, Dec) "Testing big data defense against Ebola" [online]. Fedscoop (Available) https://www.fedscoop.com/cdcs-big-data-defense-ebola/.

[83] Anema, A., Kluberg, S., Wilson, K., Hogg, R. S., Khan, K., Hay, S. I., ... & Brownstein, J. S. (2014). Digital surveillance for enhanced detection and response to outbreaks. The Lancet Infectious Diseases, 14(11), 1035-1037.

[84] Marr, B. (2018), "Big Data in Healthcare: Paris Hospitals Predict Admission Rates Using Machine Learning" [online]. Available: https://www.forbes.com/sites/bernardmarr/

2016/12/13/ big-data-in-healthcare-paris-hospitals-predict-admission-rates-using-machine-learning/ #5828452979a2 Accessed May 1, 2018.

[85] Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. Healthcare informatics research, 22(3), 156-163.

[86] Feldman, B., Martin, E. M., & Skotnes, T. (2012). Big data in healthcare hype and hope. Dr. Bonnie, 360, 122-125.

[87] "Big data in healthcare" [online]. Retrieved on 5 February 2019 from: https://www.datapine.com/blog/big-data-examples-inhealthcare/.

[88] Karanges, E. A., Blanch, B., Buckley, N. A., & Pearson, S. A. (2016). Twenty-five years of prescription opioid use in Australia: a whole-of-population analysis using pharmaceutical claims. British journal of clinical pharmacology, 82(1), 255-267.

[89] Lankenau, S. E., Teti, M., Silva, K., Bloom, J. J., Harocopos, A., & Treese, M. (2012). Initiation into prescription opioid misuse amongst young injection drug users. International Journal of Drug Policy, 23(1), 37-44.

[90] Boscarino, J. A., Rukstalis, M., Hoffman, S. N., Han, J. J., Erlich, P. M., Gerhard, G. S., & Stewart, W. F. (2010). Risk factors for drug dependence among out-patients on opioid therapy in a large US health-care system. Addiction, 105(10), 1776-1782.

[91] Birnbaum, H. G., White, A. G., Schiller, M., Waldman, T., Cleveland, J. M., & Roland, C. L. (2011). Societal costs of prescription opioid abuse, dependence, and misuse in the United States. Pain medicine, 12(4), 657-667.

[92] Bresnick, J. (2017, Sept), "cdc awards $28.6m for big data analytics to track opioid abuse" [online]. Available: https://healthitanalytics.com/news/cdc-awards-28.6m-for-big-data-analytics-to-track-opioid-abuse.

[93] Kelley J. (2017, Oct). "Using Big Data Medical Analytics to Address the Opioid Crisis" [online]. Available:

[94] https://www.forbes.com/sites/forbestechcouncil/2017/10/02/using-big-data-medical-analytics-to-address-the-opioid-crisis/#56602313142c

[95] Han, B., Compton, W. M., Blanco, C., Crane, E., Lee, J., & Jones, C. M. (2017). Prescription opioid use, misuse, and use disorders in US adults: 2015 National Survey on Drug Use and Health. Annals of internal medicine, 167(5), 293-301.

[96] Glatter, R. (2018, Jan). "Can vending machines that dispense prescription opioids help addicts survive the opiate crisis?" [online]. Available: https://www.forbes.com/sites/robertglatter/2018/01/31/can-vending-machines-that-dispense-prescription-opioids-help-addicts-survive-the-opiate-crisis/

[97] Wells, T. S., Ozminkowski, R. J., Hawkins, K., Bhattarai, G. R., & Armstrong, D. G. (2016). Leveraging big data in population health management. Big Data Analytics, 1(1), 1.

[98] Mechanic, R. E., Altman, S. H., & McDonough, J. E. (2012). The new era of payment reform, spending targets, and cost containment in Massachusetts: early lessons for the nation. Health Affairs, 31(10), 2334-234.

[99] Gregory, J.G. (2016, Mar). "3 ways data analytics can reduce total healthcare costs" [online]. Available: http://www.ibmbigdatahub.com/blog/3-ways-data-analytics-can-reduce-total-healthcare-costs.

[99] TECHPLUSHEALTHTECHNOLOGY. (2016, Jul). "Rashid Al Maktoum Humanitarian & Charity Est. & Dubai Health Authority near completion of 130 free cardiac surgeries" TechplusHealth Technology[online].

[100] Available:http://healthtechnology.in/2018/07/16/fortis-healthcare-mumbai-mohammed-bin-rashid-al-maktoum-humanitarian-charity-est-dubai-health-

authority-near-completion-of-130-free-cardiac-surgeries/

[101] Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: opportunities and policy implications. Health affairs, 33(7), 1115-1122.

[102] Bunolna, R. A. (2015). Impact of Education Related to the 2010 Affordable Care Act and the Importance of Primary and Preventive Care on Health Care Access and Utilization of those Services among Homeless Women (Doctoral dissertation, Brandman University).

[103] Nave, R. L. (2014). Science Study Latest Blow to Patient Protection and Affordable Care Act Effect on Emergency Department Use: Newly Insured Visit Emergency Departments More Often. Annals of Emergency Medicine, 63(3), A13-A15.

[104] Kale, S., & Patil, P (2016). NEED OF ERA THE GROWING AVALANCHE AND BOON TO HEALTHCARE: BIG DATA.

[105] Sotiriou, C., & Pusztai, L. (2009). Gene-expression signatures in breast cancer. New England Journal of Medicine, 360(8), 790-800.

[106] Positive Bioscience. Retrieved on 30 March 2019 from: www.positivebioscience.com

[107] Bhardwaj, R., Sethi, A., & Nambiar, R. (2014, October). Big data in genomics: An overview. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 45-49). IEEE.

[108] Phillips, K. A., Trosman, J. R., Kelley, R. K., Pletcher, M. J., Douglas, M. P., & Weldon, C. B. (2014). Genomic sequencing: assessing the health care system, policy, and big-data implications. Health affairs, 33(7), 1246-1253.

[109] El Naqa, I. (2016). Perspectives on making big data analytics work for oncology. Methods, 111, 32-44.

[110] Baralis, E., Bertotti, A., Fiori, A., & Grand, A. (2012). LAS: a software platform to support oncological data management. Journal of medical systems, 36(1), 81-90.

[111] Ding, D., Wu, D., & Yu, F. (2016, August). An overview on cloud computing platform spark for Human Genome mining. In 2016 IEEE International Conference on Mechatronics and Automation (pp. 2605-2610). IEEE.

[112] Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... & Robinson, G. E. (2015). Big data: astronomical or genomical?. PLoS biology, 13(7), e1002195.

[113] O'Driscoll, A., Daugelaite, J., & Sleator, R. D. (2013). 'Big data', Hadoop and cloud computing in genomics. Journal of biomedical informatics, 46(5), 774-781.

[114] 5 Healthcare applications of Hadoop and Big data" DeZyre, 2017, [online] Available: https://www.dezyre.com/article/5-healthcare-applications-of-hadoop-and-big-data/85.

[115] Marr, B. "How Big Data Is Changing Healthcare. [0n-line] Forbes 2015." Available at WWW:< http://www. forbes. com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare.

[116] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., ... & Bloomfield, C. D. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. science, 286(5439), 531-537.

[117] Preuveneers, D., Berbers, Y., & Joosen, W. (2016). SAMURAI: A batch and streaming context architecture for large-scale intelligent applications and environments. Journal of Ambient Intelligence and Smart Environments, 8(1), 63-78.

[118] Peddi, S. V. B., Kuhad, P., Yassine, A., Pouladzadeh, P.,

Shirmohammadi, S., & Shirehjini, A. A. N. (2017). An intelligent cloud-based data processing broker for mobile e-health multimedia applications. Future Generation Computer Systems, 66, 71-86.

[119] Manogaran, G., Thota, C., Lopez, D., & Sundarasekar, R. (2017). Big data security intelligence for healthcare industry 4.0. In Cybersecurity for Industry 4.0 (pp. 103-126). Springer, Cham.

[120] Xu, H., Aldrich, M. C., Chen, Q., Liu, H., Peterson, N. B., Dai, Q., ... & Jiang, M. (2015). Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. Journal of the American Medical Informatics Association, 22(1), 179-191.

[121] Saratchandran, V. (2018). 5 Ways Big Data is Changing the Healthcare Industry. [Online] Available at: https://www.fingent.com/blog/5-ways-big-data-is-changing-the-healthcareindustry [Accessed 2 May 2018].

[122] Ullah, F., Habib, M. A., Farhan, M., Khalid, S., Durrani, M. Y., & Jabbar, S. (2017). Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare. Sustainable Cities and Society, 34, 90-96.

[123] Stivoric, J., Gemperle, F., & Kasabach, C. (2003). U.S. Patent No. 6,527,711. Washington, DC: U.S. Patent and Trademark Office.

[124] Mesholam, R. I., Moberg, P. J., Mahr, R. N., & Doty, R. L. (1998). Olfaction in neurodegenerative disease: a meta-analysis of olfactory functioning in Alzheimer's and Parkinson's diseases. Archives of neurology, 55(1), 84-90.

[125] Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G., Saad, M., ... & Schulte, C. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature genetics, 46(9), 989.

[126] Haas, M., Stephenson, D., Romero, K., Gordon, M. F., Zach, N., Geerts, H., & Brain Health Modeling Initiative. (2016). Big data to smart data in Alzheimer's disease: Real-world examples of advanced modeling and simulation. Alzheimer's & Dementia, 12(9), 1022-1030.

[127] Hagar, Y., Albers, D., Pivovarov, R., Chase, H., Dukic, V., & Elhadad, N. (2014). Survival analysis with electronic health record data: Experiments with chronic kidney disease. Statistical Analysis and Data Mining: The ASA Data Science Journal, 7(5), 385-403.

[128] Rodger, J. A. (2015). Discovery of medical Big Data analytics: Improving the prediction of traumatic brain injury survival rates by data mining Patient Informatics Processing Software Hybrid Hadoop Hive. Informatics in Medicine Unlocked, 1, 17-26.

[129] Lee, C. H., & Yoon, H. J. (2017). Medical big data: promise and challenges. Kidney research and clinical practice, 36(1), 3.

[130] Berry, M. J., & Linoff, G. (1997). Data Mining Techniques: For Marketing, Sales, and Customer Support.