New Paradigm in Healthcare Industry Using Big Data Analytics

Hiya Luthra¹, T. Arun Sai Nihith¹, V. Sri Sai Pravallika¹, Raghuram Shree R¹, Ankur Chaurasia¹ and Hina Bansal¹

¹Amity Institute of Biotechnology, Amity University Uttar Pradesh, Sector 125, Noida-201303, U.P., India

E-mail: hinabansal@gmail.com

Abstract. New scientific methods, such as genome sequencing for the next decade, produce enormous amounts of biological data that leads us to scientific breakthroughs through careful study and interpretation. However, scholars fail to keep up with all the enormous data blocks. As the generation of salient Information is rising with each day, we can use it in a productive way. The medical data can be useful in fields of personalized medicine whose demand is growing exponentially. The term emerged in 2012, along with the "Industry 4.0" concept, Famous, and refers to big data, in the marketing of IT, in which industrial equipment-generated data could hold more Potential benefit for sector. Large Data Industrials refers to many Diverse produced time series by industrial equipment at a high speed. Preventive medicine for individual patients with chronic conditions, Big Data and healthcare are critical to address the risk of hospitalization. The current review discusses about application of big data in gene sequencing, healthcare, electronic health reports (EHRs), medical scans, genomic sequencing, reports of payers, pharmacy studies, wearables, and medical devices, gathered ample health data, to name a few. This paper mainly talks about the ways in which big data can be used and the tools for analyzing big data. Big data analytics aims to provide innovations that optimize patient care and generate value for healthcare institutions with improvements in medicine, infrastructure and funding.

1. Introduction
Big data analytics has emerged as a chief learning technology to handle with the massive amount of data. The collected data may have important evidences and may help to resolve several issues related to marketing, cyber security and healthcare [1]. In the present era individuals don’t work in a conventional way, so as the biologists of today. Everyone seeks for better results in terms of accuracy which evidently cannot be achieved by just doing experiments in a laboratory. Therefore, researchers are seeking different technologies to make the work easier and finer. Hence forth one have “Bioinformatics and Big data Analytics” as our advance guard in the field of biology. The operating procedure under bioinformatics is much more cost effective as well as economical. The results in terms of analytics are impeccable. For instance, network medicine, which is the new sunrise in the research and development sector. IT deals with many areas like wise Functional genomic interactions, system metabolic diseases, and research regarding clinical goals. As all of us know that the accumulation of significant biomedical data is escalating at a very high pace, which provides us with distinctive chances. These may include the evaluation of the result and
impression of on the known morphology, neuronal variability and environmental disruptions. By the support of amalgamation of the network medicine and biomedical data one can get impeccable outcomes. This combination can lead us to the structural framework to construct conceptual models and draw out impressive and poignant outcomes at network level. Due to the increase in the salient data, one can have the window to evaluate the upshot of cellular heterogeneity. The whole framework of biology is made up of systems like genes, proteins, and additional biological molecules, including the interrelation between these integrants. Biomedical data is extremely helpful when it comes to the interpretation of a certain phenotype, the operating system of a cell, or cause of a disease. By analyzing this biomedical – data, there will be an ease and ascendancy to understand the salient or fundamental characteristics of the interactions in the above-mentioned components. This process of analyzing biomedical data will lead us to the understanding of the physical process of the remarked phenotype. With the help of network medicine one can form a grip upon the biomedical data. It can be done through minimum optimization. The options while choosing suitable network method solely depends upon the fundamental biological inquiry, hypothesis, study design, and obtainable data [2].

1.1 The Disease-Based Network of the Human Disease Network

The interactome output map provided a cell-based contact map for certain types of protein input that may pass through the basal patho phenotype [3]. With the availability of accurate and relevant biomarkers, network-based disease control, single networks, better disease classification, specific drug development targets, and repeated drug enhancement, the disease can be catch in terms of biomedical data entering network systems. With the help of this PARADIGM one can say get the disease - certain signals in different ways. The thing that should always be remembered is the topological characters of the nodes and analyze the important role of their durability, i.e., the function of a node to have an air-filled network using "case - in association" - a feature based on direct evidence but a combination of additional genes, albeit with care. Cellular network networks can be used to identify subset - automated networks related to disease phenotype, as an addition to student genetic guidance. The network tool can also give the formation of a disease phase. Network-based techniques for detecting genetic disease and associated processes Two forms can be broken down: observational and observational methods [4][5]. One might examine biochemical patterns due to the corruption in analytical techniques. Chu et al., for example, built on existing methods of autophagy known to create an angiogenesis PPI network. In comparison, computational approaches are aimed at defining particular disease-related genes and pathologies. For example, Ogden and the team developed a gene sequencing method to classify a genetic system of genes impacted by de denovo CNVs that are rare in autism centered on a genomic framework [6]. Recently, Huang and his team systematically tested 21 networks of protein interactions in order to detect genetic sets [7]. After adjusting the size, it has been observed that the Database for Interacting Proteins (DIP) network had the highest potential for regenerating disease genes [8].

1.2 Identifying Key Genes using Co Patterns - Multiple Biomolecules

Analyzing the text quantity or genetic marking of a phenotype (case control) in multiple samples is one of the main or primary methods used to investigate the whole system as combined with the central belief in molecular biology. With the identification of the main genes created by the disease a separate genetic analysis was performed. However, it does not satisfy any clues as to how these genes are affected or how they affect other genes. Those genes that have similarity in their expression patterns have been identified as being part of structures, interactions, or fragments of similar patterns or pathways. This has been a new phenomenon in the formation of GCNS where the proliferation of text is known in relation to this disease. The main belief of this paradigm is to join the required genes through an organic network of signaling pathways that result from genetic data from a direct system.
There are many methods that can include integration or mass formation, including pearson integration, spearman level integration, individual details, Gaussian graphical model, background methods, Base methods, random matrix belief, and partial integration. Regarding retaliation for rehabilitation or externalization CGN'S recognizes the systematic functioning of genetic involvement. With the help of microarray or RNA data - seq, CGN'S can be signed or unsigned, weighted or non-weighted, and can be synthesized. Obtaining co-expression networks that do not carry a good amount of surveillance is required while using blockchain methods, as they are personalized and change network structure and topology; procedures regarding junctions, random matrix theory, or soft grips, which increase weights with a limited amount of force to punish angles or weak edges [9], have been used to designate this limit. Multiple isoform and different uses can be used during the construction of GCN'S, in line with complete genetic expression levels [10].

2. Big Data Parameters

Big data management processes with the collection and retrieval of multiple, confusing, dynamic and large datasets. Big data is typically Defined by five quantities.

a. Value: Quality is one of Big Data 's core characteristics. Saving enormous values in their libraries is critical for the IT infrastructure scheme.

b. Velocity: The phase at which velocity information is produced impacts the information output.

c. Variety: Wide selection states in different types of sources as well as the material is provided in both organized and unstructured ways.

d. Volume: Amount is the word "Big Data" that addresses incredibly vast data.

e. Variability: Variability is a concept which interacts with inconsistent data.

3. Equipment for Big Data Analytic

Table 1 is showing list of tools available for big data analytics. Semantria: Decision requires gives us a specific program that is obtained from different customers by gathering different data. And after that, the method of processing the data is faithfully implemented to generate the most precious and desirable insights. Open text: is a module for Emotion Analysis. It is a special kind of engine which is used to figure out different subjective trends of classification. It is often used in text type to determine the expressions of emotion that are present. Refer table 1.

| S. No | Data tools | Data storage framework | Data filtering and extraction tools | Data and cleaning validation tools |
|------|------------|------------------------|-----------------------------------|----------------------------------|
| 1    | Semantria  | Apache HBase (Hadoop database) | Scraper | DataCleaner |
| 2    | Opinion Crawl | CouchDB | OctoPrase | MapReduce |
| 3    | Open Text  | MangoDB | ParseHub | Rapidminder |
| 4    | Trackur    | Apache spark | Mozenda | OpenRefine |
| 5    | SAS Sentiment Analysis | Oracle, NoSQL Database | Content Grabber | Talend |

Trackur: is a method used during data collection. After decisions are being made, it uses its automated text review looked at the specific keywords tracked by users. The mood may well be good, bad, or it may be
indifferent to the related text. SAS Sentiment Analysis: is also a method for emotion analysis that collects emotions in real time automatically. With the aid of a mixture of different methods of statistical modelling, it performs this job. This method of production is based on the natural language theory. Refer table 2.

Table 2: Comparison among Spark, Hadoop and MongoDB

| Criteria               | Hadoop            | Spark              | MongoDB            |
|------------------------|-------------------|--------------------|--------------------|
| Model of Processing    | Batches           | Mini Batches       | Single Document    |
| Latency                | High latency      | Very low latency   | Low                |
| Fault tolerance        | Uses replications | Uses RDD           | Replica Set        |
| Supported languages    | Java              | Scala, Java, Python| C++                |
| Batch Framework        | HDFS              | Core Spark API     | BSON               |
| Security               | High Secure       | Less secure        | Secure             |
| Advantages             | processing and speed, low cost, secure, Storage | powerful caching, Scalable, high speed, | Expandable |
| Weakness               | ‘Name Node’, is the single point of failure | small files problems, Expensive, less no. of algorithm | Fault tolerance issues |
| Hardware Cost          | Cost is more      | Cost is more       | Cost effective     |

4. Big Data Analytics in Personalized Medicine

It is well known that nowadays the producers in the field of medicine are focusing upon cost effective and targeted drugs, which means that the future of drug industry puts forward the concept of targeted drugs for faster recovery with least amount of harm to the patients. With this innovatory vision we may achieve some significant results in healthcare domain. So, in healthcare the technologies that are used for big data analysis has given proven results when it comes to patient’s treatment cost. Now the question arises that what is this biological data and what is the need to analyze such data? One can say that this data can be in the form of diagnostic images, genetic test results and biometric information. Such data is accumulating day by day in the electronic health records, but by nature it has a very high pace, variety and volume so here arises the need to channelize, process or analyze such data in a productive manner. The need for big data analysis to generate “smart data” which in turn provide us with some pragmatic details that reinforce better decision for personalized medicine (Figure 1).
Figure 1. Big Data Analytics in Personalized Medicine: A Holistic Approach Towards the Human Welfare

To make it more successful there is a need of enormous amount of structured and unstructured data of individual patients which includes its datatypes respectively. To utilize the healthcare big data many researchers have administered frameworks and technique. One of the most acclaimed and esteemed frameworks is “Hadoop”. Hadoop contributes to the analysis of the big datasets. This is one of the frameworks which has been used widely when it comes to patient prognosis, cancer diagnosis, critical disease warning, disease decision-making rules, general medical data testing, and personalized recommendations programs. Who would have thought that a patient’s unique characteristic could be the sunrise in the future of medicine? Precision medicines are more elaborate than the usual treatment. These days some cardiologists are using a certain algorithm for patients which gives the contingency for myocardial infarction within 5 or 10 years [11]. The outcome of this algorithm is based on a number of factors such as weight of the body, blood pressure, smoking, results of blood lipid analysis, and personal and family history of the heart. The concept of a personalized medicine is not new, but the dawn of herculean analytical tools has opened new doors for predictable, preventative, participatory, and personalized drugs, known as P4 drugs [12]. According to the stats in 2015, personalized medicine has been included in more than 25% of the novel new drug validated by the US Food and Drug Administration (FDA) [13]. This shows the turning point for personalized medicine towards becoming a significant constituent of treatment products. There are many examples to show that researchers are focused on advanced genomic analysis to find customized management by building computational methods. For example. Baseline Study project by Google Inc., Cancer Genome Atlas, and the 100,000 Genomes Project (100KGP) [14]. As the analysis of data increases, one can determine a pattern in each patient’s data. With the help of machine learning methods Vidyasagar found some biomarkers, to a greater extent he also found out other than the importance of life that can predict the drug response [15]. Table 2 is showing the Comparison among Hadoop, Spark and MongoDB.

A very genuine question to ask is that how does one extract vital information from big data? We have a holistic data which can be discovered through several data-driven consolidative workflows that are normally in need of inference of associations among numerous entities. The task can be made easier by predicting the outcome or by describing the event using machine reading data for multiple viewing. This can be done by standard modeling models (GLMs). GLMs work in the context of a broader model in which the results are directly linked to objects and covariates with a linking function that allows for model limitations typically with multiple opportunities or Bayesian strategies. In addition, Bayesian models such as the Naïve Bayes
classifier and ensemble-learning models such as random forest along with neural networks and deep new learning can also analyze the data with multi-view details.

5. Big data analytics in Genome Sequencing

Novel biomedical methods for example next-generation sequencing, are creating meaning out of immense volumes of bio data which upon careful analysis and interpretation are encouraging scientific breakthroughs, although scientists are finding it difficult to maintain and update with the huge blocks of human biological data. As an example, the National Institutes of Health has started an initiative known as the Big Data to Knowledge (BD2K) and another one which is ‘Precision Medicine Initiative’, whose goal is to develop genetically guided treatment with personalized medicine (and the use of precision medicine for) enhanced predetection, prevention and right treatment of complicated medical problems [16]. Hemoglobinopathies refers to the disorders and diseases caused in red blood cells. Genetic blood disorders are proven to be excellent individuals for practicing gene therapy as, in an autologous hematopoietic stem cells (HSCs), the action of gene therapy can change the causative gene and make the necessary changes in the hematopoietic system. Two inherited blood diseases are beta thalassemia and another very prominent one is sickle cell disease. β-thalassemia is caused by a single-point modification or removal from the β-globin type, resulting in less production of β-globin in the body. Sickle cell disease occurs due to a mutation of Glutamine to Valine occurs in the lower β-globin fraction of hemoglobin, resulting in abnormal production of hemoglobin in the body. A universal method for curing all β-globin disorders is to re-express the paralogous γ-globin genes. The Bauer team used CRISPR / Cas-based cleavage technology for the GATA1 binding erythroid enhancement site which reduced erythroid expression of L-globin BCL11A suppressor and increased γ-globin expression. CRISPR Therapeutics has launched three clinical trials for one in 2018. Allife Medical Science and Technology Co., Ltd in 2019 also launched clinical trials for patients with β-thalassemia and sickle cell disease by attracting CRIPSR / Cas9-edited CD34 + human HSC transfusions (CTX001) for candidates [17].

Adrian Lee, a professor at the University of Pittsburgh has given his entire career for understanding and studying breast cancer. He along with some fellow researchers are trying to find the connection between a host of clinical data which includes demographic information which is the age, ethnicity and body weight of the patient and the molecular picture of patients with breast cancer. “We've got a big haystack and we're trying to find the needle,” says Lee. “But we're also trying to incriminate the needle, by linking it to lots of things” [18]. As big-data scientists explore through vast tumor databases searching for mutation pattern, big-data methods are introducing new categories/classifications of breast cancer that could potentially expose cellular pathways that were previously ignored. Pleiotropy occurs when several traits are affected by a single locus. Not only does the characterization of pleiotropy's molecular mechanisms help to clarify the relationship between diseases, but it can also contribute to new insights into the pathological mechanism of each particular disease, which helps in improving the prevention, diagnosis and correct treatment of the disease. Biomedical science has been pioneered by the emergence of Big Data [19]. Table 3 is showing the list of Conferences held revealing the application of big data in personalized medicine.

Table 3: Big Conferences Regarding Personalized Medicine

| Initiative                          | Research focus | Initiated in (year) |
|-------------------------------------|---------------|---------------------|
| Human Genome Diversity Project      | General       | 1990                |
| Global Network of Personal Genome Projects | General       | 2005                |
| The Encyclopedia of DNA Elements    | General       | 2003                |
6. Genetics and big data in multiple sclerosis

Multiple sclerosis is a complex disorder in which biology and environmental factors play a key role in its susceptibility. Family experiments have shown the contribution of genetics, but this point is still to be proved. Big data has become an amazing revolutionary scientific strategy, and multiple and efficient tools have been created for researchers to treat multiple sclerosis and genetics with the help of user-friendly graphical user interface tools. Multiple sclerosis (MS) is defined as a central nervous system disorder which gives rise to impairment in nervous system, particularly in young females. Multiple Sclerosis is a complicated disease that leads to the pathogenesis of both genetic and environmental factors [20].

For example, MS Base is dedicated to the combination, comparison and analysis of large repositories and a virtual data entry system. By using the statistical analysis and expanding the sample in the longitudinal direction, the timely increase of the female to male ratio in relapsing-remitting multiple sclerosis has been observed to find the effect of drug-modifying illness-discontinuation and the prediction of potential pathogens have been identified and evaluated (Figure 2) [21].

![Figure 2. Representation of Disease Susceptibility by GWAS (Genome-wide association study)](image-url)
7. Big Data Analytics in Preventive Medicine

Big data and healthcare are important for patients with chronic diseases who can’t bear the expenses of hospitalization. Optum Labs, which is a US based examination community, is gathering electronic wellbeing records of 30 million patients to make an information base for predictive analytics devices that will improve the medical care sector [22].

The goal of big data in preventive medicine is to help physicians make data-driven decisions in seconds and improve patient care. It is possible for health care facilities to provide reliable preventive care and, ultimately, reduce hospital admissions by understanding techniques such as drug type, symptoms, and frequency of patient visits. This is especially helpful in cases of patients with complex medical histories, suffering from a number of conditions.

This won’t just diminish the degree of risk which prompts less spending on in-house patient care, however it will additionally ensure that space and assets are accessible for those that require it most. This is clearly a sign of how huge information examination in medical services can improve and save people’s lives. Big data in medical services has the power to help in novel treatment and creative medication treatment. Medical experts may perceive potential qualities and shortcomings in preliminaries or cycles by utilizing a blend of chronicled, continuous, and prediction tools as an intelligible combination of information representation procedures [23].

Additionally, through information driven hereditary data examination forecasts patient’s historic data, huge information investigation in medical care can assume a significant function for the advancement of new medications and groundbreaking treatments. Medical care information investigation can develop, give security, and save many lives. This provides confidence and consistency, and the way forward. Big data has allowed healthcare institutions to take a full view of the health condition of a patient and contribute to new results, new treatment options, and more precise diagnosis. Data availability has drawn attention to previously overlooked factors associated with health conditions [23].

Some races, for instance, are genetically more predisposed than other races to heart diseases. Today, when heart disease is sustained by a patient representing one of these races, the time has come to investigate the information of patients having a place with a similar race who have griped of heart issues. It assists with studying the eating designs, way of life, hereditary cosmetics, family DNA, proteins, cell metabolites, tissues, organs, species, and biological systems of such patients.

7.1 Case Study - IBM Watson

Big data is useful in preventive medicine. There are several tools that have been developed for its analysis and application. One of these tools is IBM WATSON (figure 3). When a hospital is treating a patient, it makes sure to have all the humongous patient record details. This important data can be used to predict diseases with an amazing degree of precision. For example, if a patient has undergone a stroke, the hospital will have details about the time of the stroke, the distance between strokes in the past in the case of several strokes, affecting pre-stroke events such as a mentally traumatic event or heavy physical activities. Based on the patient’s knowledge, hospitals may have critical measures to avoid strokes.

The capacity to accurately predict the course of a patient's sickness will permit the guardian to more readily target patients who will have the advantage from costly and complex treatments, in such a way that it decreases the rate of infection in patients and cut off the expenses of the medical services framework. Immune system issues, for example, scleroderma, rheumatoid joint inflammation, and fundamental lupus erythematosus are instances of such conditions.

The information that IBM Watson uses for assessment can include therapy rules, electronic clinical record information, notes from specialists and chaperons, research materials, clinical examinations, diary articles,
and patient data. Regardless of being created and marketed as a "diagnosis and treatment advisor", Watson has been useful in helping with identifying alternative treatment for patients who have already been diagnosed with different diseases. Watson for Drug Discovery encourages researchers to perceive novel medication targets and new treatment for existing medications. The stage will permit scientists to rapidly find new ties, which can provoke novel thoughts and coherent accomplishments [24]. (figure 3)

Figure 3. IBM WATSON Functioning: for drug discovery encourages researchers to perceive novel medication targets and new treatment for existing medications.

8. Big Data in Omics Studies

Big data is playing a key role in the field of bioinformatics [25]. Table 4 discuss about various bioinformatics analysis tools such as: SPARKSEQ, SAMQA, ART, DISTMAP, SEQWARE, CLOUDBURST, HYDRA, BLUESNP, and MYRNA. These all tools are responsible for analysis, genome correction, finding relation and generating simpler data from genome sequences [26]. Refer table 4.

Table 4. Description of Bioinformatics Tools

| Bioinformatics Tools | Description |
|----------------------|-------------|
| Sparkseq             | Sparkseq is a database analyze genome sequence using “Hadoop” and “apache-spark”. |
| SAMQA                | It identifies errors caused by mutations or defects in a genome. |
| ART                  | Finds no of errors per a gene extracted using “SOLiD”&”illumina” platforms. |
| DistMap              | It is based on “Hadoop-cluster” for short-read mapping of a gene/genome. |
| SeqWare              | Seqware is also using “Hadoop” to generate data of genomes by integrating genome tools & browsers. |
| CloudBurst           | Used in gene mapping procedure to improve feasibility of reading large sequences. |
| Hydra                | Based on “Hadoop library” for sequencing protein sequences, HYDRA is capable of performing 27 billion peptide scorings in less than 1 hour. |
| BlueSNP              | based on “Hadoop-library” which finds relation b/w genotype & phenotype. |
Myrna provide data on differences in expression of genes and statistical modeling.

9. Platforms for Healthcare Analytics

9.1 AYASDI
AYASDI is based on machine learning and AI algorithms and is capable of analyzing and managing EHR’S of patients and financial data and predict the quality and cost of treatment and also assessment and management of population health.

Figure 4 explains features of AYASDI such as: clinical variation management and fraud, waste & abuse management. Clinical variation management analyses patient EHR and generates clinical pathways or solutions which cost patients at low price. Whereas fraud, waste & abuse management analyze and validates existing detection models to predict fraud people in management of patient outcomes.

Figure 4. Features of AYASDI related to healthcare analysis & fraud, abuse and waste management

9.2 LINGUAMATICS
It is a natural language processing-based algorithm that depends on interactive text mining algorithms and it explains information on genetic relationships. Some of its goals are “Boost Innovation”, “Speed R&D and Clinical Processes”, “Optimize Quality & Improve Efficiency”, “Reduce Risk & Costs” and to “Improve Patient Outcomes”.

Figure 5 explains features of LINGUAMATICS such as: drug discovery, precision medicine, patient quality and supports clinical research. Whereas professional services: support pharma, healthcare, NLP, solutions delivery.

NLP technology platform covers ontologies, chemistry, web-services, EASL and multi-lingual NLP. Infrastructure discuss about: private cloud technology, content storage, and federated architecture.
Figure 5. Features of LINGUAMATICS such as: various services, solution areas and technology to extract genetic relationship

9.3 Applications of Big Data in Healthcare

a. “Infectious disease management”: Young et. al. [27] on survey through social media, collected 553K tweets and concluded more than 9000 tweets with HIV related words. Studies said that there is a positive relation between HIV-tweets and HIV-cases.

b. “Population health management”: Lamarche-Vidal et al. [28] analysed the data from more than 400K patients from 2008-09 and in conclusion show that around eight percent were in-hospital fatalities and around 19% were out-hospital fatalities.

c. “Mental-health management”: Mental health management tools analysed the tweets posted and declare the mental behaviour and emotional patterns. Dabek and Caban presented a neural network model that can predict psychological disorders, like nervousness, personality disorders, depression, and shellshock.

d. “Disease control”: “CANHEART” team aimed at improving cardiovascular health and cardiovascular care provided in Ontario, Canada. Kupersmith-et-al introduced IT in health care to detect diabetes & chronic kidney diseases via vha [US veterans’ health administration].

e. “Virtual physiological human (VPH)”: normally health professionals such as nephrologist diagnose kidneys and dermatologist looks at skin but there is no doctor who diagnose whole body. so VPH is introduced to make collaboration of all parts of a human body. VPH will be able to identify relation between different parts of body and gives a solution to patients suffering with co-morbidity.

f. “In tacking osteoporosis”: FRAX is a database which detects the mineral decomposition rate in bones of osteoporosis patients, so that we can know when fracture occurs exactly. This can allow us to take proper medication before fracture.

9.4 National Population Based Cancer Databases

Databases SEER [surveillance, epidemiology and end results] extracts and uploads cases related to cancer from registries collected from population [29, 30], NPCR: is National Program for Cancer Registries, was established by congress in 1992 and controlled by CPCD. It collects data related to cancer prediction, and treatment methods and resulting outcome.
10. Big Data in Industry

A. Predictive Analytics: It predicts future outcomes based on current data. Predictive analytics based on statistical methods. It predicts the failure of jet engines from sensors. Also predicts customers' next move based on what they buy and what they say on social media [31].

B. Text Analytics/Mining: it can be done in two steps:
   a. Information Extraction: It extracts structured data from unstructured data, like drug name, dosage, and frequency from medical prescriptions. IE has two types: ER [Entity Recognition] + RE [Relation Extraction] ER categorizes text such as person’s info, location, organization, etc. ER finds relations between entities like drug, genes, person, organization etc. [32].
   b. Text Summarization: It produce a summary of extracted documents. It is of two types: extractive and abstractive. Extractive summarization does not require understanding of a text, whereas abstractive summarization require semantic information or full understanding of a text via NLP techniques. Abstractive summary has good summary than an extractive, while extractive summary is used for big data widely [33].

C. Audio Analytics: In healthcare audio analysis helpful in the diagnosis and treatment of persons suffering from depression, schizophrenia, and cancer. Also useful in customer call centres and in healthcare. It is useful in improving customer experience, improve sales turnover rates and monitor privacy and security policies. In healthcare also useful in analysing infant cries and determines infant health and emotional status.

D. In Industry-4.0: System infrastructure: It is based on data extraction, storage, uploading and data analysis. Data analytics summarize the data captured and stored via system infrastructure. Some of the data analytic methods are: “Descriptive analytics”: it predicts what was happened in the past. “Predictive analytics”: it will predict what can happen in future based on descriptive analysis. “Prescriptive analytics”: it will let us know how to get prepared for future based on predictive analysis.

E. Big Data analytics in Computational Intelligence: Instead of manually writing the data generated by drilling, automated big data tools are also used. Big data tools like ELECTRONIC DRILLING RECORDER and WEIGHT-ON-BIT are important to improve drilling performance [34][35].

F. Big Data in Refining: To refine a cracked gas a compressor is used called “cracked gas compressor (CGC)”. Using big data analysis, the production, quality, contaminants, and minerals extracted in a compressed gas are analysed.

G. Big Data in Oil and Gas Transportation: research found that how to predict the performance of ships through propulsion power and predicted how to lower greenhouse gas emissions [GHG] using sensors like XGboost, MLP were used for big data analysis [36][37].

11. Conclusion

As the generation of salient Information is rising with each day, one can use it in a productive way. The medical data can be useful in fields of personalized medicine whose demand is growing exponentially. Biological and medical study and exploration have been significantly improved by the advent of rapid DNA sequencing methods. Prophylaxis, or Preventive treatment consists of interventions that are taken to eliminate diseases. Environmental variables, genetic predisposition, viral infections, and lifestyle preferences impact illness and disability and are complex pathways that occur before people know that they will be affected. Disease reduction relies on anticipatory measures that can be described as main, basic, secondary or tertiary preventive measures. In conclusion, we can state that big data is serving several applications in healthcare including risk and disease management, self – harm prevention, medical imaging, augment cancer treatments, telemedicine, improved strategic planning, developing new therapies predictive analytics, enhancing patient engagement and many more. Big data is a launch to transfigure Universal health care on many levels to drive the company forward. The shifts in medicine, infrastructure, and support that
data analytics promises in health provide solutions that maximize clinical care and increase medical professionals' value.

12. References

[1]. Amalina F, Hashem IA, Azizul ZH, Fong AT, Firdaus A, Imran M, Anuar NB. Blending big data analytics: Review on challenges and a recent study. IEEE Access. 2019 Jun;178:3629-45.

[2]. Ratra R, Gulia P. Big Data Tools and Techniques: A Roadmap for Predictive Analytics. International Journal of Engineering and Advanced Technology (IJET), Volume-9 Issue-2, December 2019: 4986-4992.

[3]. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB—a database for integrating human functional interaction networks. Nucleic acids research. 2009 Jan 1;37(suppl_1): D623-8.

[4]. Carter H, Hofree M, Ideker T. Genotype to phenotype via network analysis. Current opinion in genetics & development. 2013 Dec 1;23(6):611-21.

[5]. Chu C, Hsu AL, Chou KH, Bandettini P, Lin C, Alzheimer's Disease Neuroimaging Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage. 2012 Mar 1;60(1):59-70.

[6]. Gilman SR, Fossiiov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron. 2011 Jun;70(5):898-907.

[7]. Huang JK, Carlin DE, Yu MK, Zhang W, Kreisberg JF, Tamayo P, Ideker T. Systematic evaluation of molecular networks for discovery of disease genes. Cell systems. 2018 Apr 25;6(4):484-95.

[8]. Marcotte EM, Xenarios I, Eisenberg D. Mining literature for protein–protein interactions. Bioinformatics. 2001 Apr 1;17(4):359-63.

[9]. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC bioinformatics. 2008 Dec 1;9(1):559.

[10]. Hulsen T, Jamuar SS, Moody AR, Kames JH, Varga O, Hedensted S, Spreafico R, Hafler DA, McKinney EF. From big data to precision medicine. Frontiers in Medicine. 2019 Mar;1:634.

[11]. Kensler TW, Spira A, Garber JE, Szabo E, Lee JJ, Dong Z, Dannenberg AJ, Hait WN, Blackburn E, Davidson NE, Foti M. Transforming cancer prevention through precision medicine and immun-oncology. Cancer Prevention Research. 2016 Jan 1;9(1):2-10.

[12]. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nature reviews Clinical oncology. 2011 Mar;8(3):184-7.

[13]. Nice EC. From proteomics to personalized medicine: the road ahead. Expert Rev Proteomics. 2016;13(4):341–343

[14]. Ibrahim, Rania, Maria Pasic, and George M. Yousif. "Omics for personalized medicine: defining the current we swim in." (2016): 719-722.

[15]. Vidyasagar M. Identifying predictive features in drug response using machine learning: opportunities and challenges. Annual review of pharmacology and toxicology. 2015 Jan 6; 55: 15-34.

[16]. Cirillo D, Valencia A. Big data analytics for personalized medicine. Current opinion in biotechnology. 2019 Aug 1;58:161-7.

[17]. Tom C. Studying genetics in the age of big data. Spectrum Autism Research News. (2018) (https://www.spectrumnews.org/news/studying-genetics-age-big-data/)

[18]. How Big Data is Changing Genetic Research. Getsmarter (2019) (https://www.getsmarter.com/blog/market-trends/how-big-data-is-changing-genetic-research/)

[19]. Adams JU. Genetics: big hopes for big data. Nature. 2015 Nov;527(7578): S108-9.

[20]. Didonna A, Oksenberg JR. The genetics of multiple sclerosis. Exon Publications. 2017 Nov 8:3-16.
[21]. He KY, Ge D, He MM. Big data analytics for genomic medicine. *International journal of molecular sciences*. 2017 Feb;18(2):412.

[22]. Isobe N. Genetics in multiple sclerosis: Updates in the era of big data. Clinical and Experimental Neuroimmunology. 2018 Feb;9(1):19-24.

[23]. Razzak MI, Imran M, Xu G. Big data analytics for preventive medicine. *Neural Computing and Applications*. 2020 May;32(9):4417-51.

[24]. Sandra D. 18 Examples of Big Data Analytics In Healthcare That Can Save People. 2020. (https://www.datapine.com/blog/big-data-examples-in-healthcare/)

[25]. Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. Methods. 2019 Aug 15; 166:4-21.

[26]. Guidi G, Miniati R, Mazzola M, Iadanza E. Case study: IBM Watson Analytics cloud platform as analytics-as-a-service system for heart failure early detection. *Future Internet*. 2016 Sep;8(3):32.

[27]. Young HS, Wood CL, Kilpatrick AM, Lafferty KD, Nunn CL, Vincent JR. Conservation, biodiversity and infectious disease: scientific evidence and policy implications.

[28]. Vidal N, León-Garcia M, Jiménez M, Bermúdez K, De Vos P. Community and health staff perceptions on non-communicable disease management in El Salvador’s health system: a qualitative study. *BMC Health Services Research*. 2020 Dec;20(1):1-4.

[29]. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*. 2019 Dec 1;6(1):54.

[30]. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights*. 2016 Jan; 8: BII-S1559.

[31]. Tsai CJ, Riaz N, Gomez SL. Big Data in Cancer Research: Real-World Resources for Precision Oncology to Improve Cancer Care Delivery. In *Seminars in radiation oncology* 2019 Oct 1 (Vol. 29, No. 4, pp. 306-310). WB Saunders.

[32]. Xu LD, Duan L. Big data for cyber physical systems in industry 4.0: a survey. *Enterprise Information Systems*. 2019 Feb 7;13(2):148-69.

[33]. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*. 2015 Apr 1;35(2):137-44.

[34]. Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalized healthcare. *IEEE journal of biomedical and health informatics*. 2015 Feb 24;19(4):1209-15.

[35]. Iqbal R, Doctor F, More B, Mahmud S, Youssuf U. Big Data analytics and Computational Intelligence for Cyber–Physical Systems: Recent trends and state of the art applications. *Future Generation Computer Systems*. 2020 Apr 1;105:766-78.

[36]. Mohammadpour M, Torabi F. Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*. 2018 Dec 1.

[37]. Nguyen T, Gosine RG, Warrian P. A Systematic Review of Big Data Analytics for Oil and Gas Industry 4.0. *IEEE Access*. 2020 Mar 9;8:61183-201.