A Review on "Role of Big Data Analytics in Health Care Sector and Pharmaceutical Industry"

[1]P. Shailaja,[2] D. Nikhila Pallavi,[3]B.Gunasree,[4]G.Snehalatha,[5]K.SatyaAshok.

[1]Associate Professor,[2&3]Students, [4&5] Research Scholars.
[1]Department of Pharmaceutical Regulatory Affairs,
[1]Andhra University College of Pharmaceutical Sciences, Visakhapatnam, India.

Abstract-
Health care and pharmaceutical industry is complex sector where a large amount of data from multiple sources is generated at an increased rate. Big Data Analytics is a current trend which plays a crucial role in pharmaceutical sector (drug discovery, precision medicine, clinical trials, research and development) and health care sector (case specific treatment, supporting preventive health, electronic health records, retinal scans, biomarkers data, consumer review etc.,). This article gives an insight about eventuality of big data in revolutionizing the health care & pharmaceutical sector, significant advantages and challenges faced in the development of data analytics; its influence in skill enhancement, furnishing patient specific treatments and aids in the development of technology used.

Keywords- **Big data analytics, Pharmaceutical and health care industries, Research and Development, Precision medicine, Preventive health, Clinical trials.**

I. INTRODUCTION

Health care and pharmaceutical industries extract large volumes of information from different sources with the main objective to utilize this big data to support, enhance patient care and to apply better treatments. Traditional system of data management such as maintaining records and reports in hard copies are outdated. In this new era, with increased rate of data generation each day, it is necessary for health care and pharmaceutical industries to adopt digitization in the form of big data to have an expansive view on public health. A report delivered to the U.S Congress in 2012 defines big data as "large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information.[1] "Big data analytics is the collection of massivequantities of unstructured data which is generated from multiple sources in pharmaceutical industry and health care sector such ashospitals or clinics, regulatory requirements, medical records, patient care, Electronic Health Records (EHR), demographics, social media, genomic data, medical history, clinical reports, doctor prescription, diagnostic reports where the data is analysed and interpreted to ensure public health. The collected data is integrated on a day-to-day basis with the help of advanced techniques. Effective use of big data in Pharmaceutical and health care sector helps in prostrating major challenges such as surge in healthcare costs. It also helps in in assuring public health, aids clinical improvement, determining and executing relevant treatments procedures for patients, ensures patient safety, efficiently monitors health care system.In health care industry, big data supports multiple kinds of data management functions which include disease surveillance, monitoring patient's health.

II. THE FIVE 'V's OF BIG DATA [1]

Big data can be described through five main characteristics i.e., the five V's which include: Volume, Value, Variety, Veracity and Velocity.

1. **Volume:** This is the primary traitor big data. It is the amount of data that pharmaceutical companies analysed and interpreted. Organizations induce terabytes or pet bytes of data while conducting their business and complying with government regulations
2. **Value:** The value of big data analytics comes in the form of better outcomes, smart strategic decision making, and improved business efficiencies
3. **Variety:** It is the principle characteristic of Big Data. Different types of data is uprooted from the health care organizations on a daily basis which include data generated from electronic health records, medical devices, unstructured data, structured data, clinical and non-clinical data, spread sheets, audio, video, images which is needed to be standardized. Structured data (traditional databases, like SQL); semi-structured data (data that have tags and markers, without a formal structure like database); and unstructured data (i.e., unorganized data)
4. **Veracity:** The data generated should be accurate and of high quality. It is determined as the trust on the collected data
5. **Velocity:** It is the rate at which the data is produced. The data should be integrated, analysed, reviewed, digested and reprocessed with high speed. In health care organizations, data collected from medical devices which monitor patients in the hospitals reach the destination quickly

III. SOURCES FORBIG DATA [2]

a) EHR – height, weight, blood pressure readings, glucose measurements, pulse oximeter readings
b) Clinical decision support systems - Doctor's prescription, Diagnostic reports, Diagnostic images, medical reports, Records, Patient medical history, Demographics, Genomic data, medical insurance data.

c) Social media – data from social media posts, articles in medical journals, blog, health plans, Facebook
d) Finger prints, retinal scans, X-rays, medical images
e) Regulatory requirements
f) Biomarkers data
g) In patients and out patients' sales data
h) Sales and marketing data
i) Customer reviews

IV. SKILLS PERFORMED BY BIG DATA TECHNOLOGY

Data quality management; Understanding information operations; Decision support models Data warehousing and data mining; Prescriptive analytics; Data visualization; Data processing and governance; Data science understanding; Cloud computing; Business modelling enhancement.

V. DATAMANAGEMENT TECHNOLOGIES / TOOLS USED TO LEVERAGE BIG DATA

Pharmaceutical and health care industry need data operation tools, techniques and software to handle massive and complex data generated with an increased rate. Analytical softwareis able to analyse and perform various activities like predictive modelling, understanding complex data patterns, improving operational activities in health care centre, assisting health care professionals, analysing data to be in compliance with regulatory requirements and in optimizing pharmaceutical production processes.

1. Apache Hadoop[3]

It is most extensively used open-source technology developed by Apache software foundation in the year 2006.It is the only technology which allows to store all the health care data in its native form. Hadoopindeed stores huge amount of data from the social media. Hadoop can process and assay data in large volumes of terabyte and pet bytes. Hadoop solves issues related to storage and access and manages overheads associated with large data sets, and the operation of very fast parallel processing by adding additional clusters, Hadoop capacity can be increased.

Hadoop technology is very important and essential for processing big data as it has high processing capacity at a faster rate; offers protection from any hardware failure; has good data backup facility which stores copies of data; cost effective;reliability;pre-processing of data is not needed

But operation of Hadoop software is challenging as it requires experts and professed people. Finding the right talents to analyse big data is one of the biggest challenges for health organizations. Finding skilled persons such as competent data scientists (professionals with data mining, visualization, analysis, manipulation, and discovery skills) is challenging and expensive for most organizations. Hadoop technology is widely used in cancer treatments and genomics; monitoring patient vitals; identification and prevention of fraudulent claims in health care insurance payments.

2. Apache Cassandra

It is a Dynamo-based commercially available technology developed by Apache software foundation in the year 2008 that can store two million columns in a single row therefore allowing the utilization of a large amount of data without requiring prior knowledge of data formatting. It specially has columnar database storage with high scalability. Apache Cassandra manages health care data analytics such as electronic health records, Clinical trial reports without any failures.

3. Apache spark [4]

Spark is the most widely used technology in health care system developed by Apache Spark in the year 2014. Spark processes and integrates data from multiple sources at an extremely faster rate. Apache spark is mostly used to perform predictive analysis on patient data where the health risks can be identified and notified to the patients at the earliest. It helps in effective monitoring of patient's health; in improving treatment standards; in taking necessary preventative measures and in delivering better treatments. Apache spark in Genomic analysis – spark is considerably used in genomic sequencing and identifying the variants at a faster rate. It can also identify the genomic sequencing of genetic traits related to specific disease conditions

VI. HOW IS BIG DATA ANALYTICS TRANSFORMINGTHE PHARMACEUTICAL INDUSTRY?[5]

Big data with its greater volume, velocity and veracity requires data analytics for processing and integration of the data thereby adding its demand especially in pharmaceutical industries. As Pharmaceutical companies substantially rely on research and clinical data to identify disease patterns, test propositions and understand the efficiency of treatments. Pharmaceutical companies rely on big data analytics and algorithms in managing Scientific research labs; Hospitals; Medical institutions; Medical research; Clinical operations; R&D investigation; Designing and developing business models; Data base design; Inpatient and outpatients data management; Enhancing sales; Development of marketing strategies.

1. Big data analytics in drug discovery [6]:

Big data analytics plays a pivotal role in medicine discovery and vaccine development. predictive modeling and designing of drugs using data analytics has replaced the old method where researchers used to manually test the samples of different compounds for new drug discovery which would consume large amount of money, time and resources. Predictive modeling helps in learning drug-drug interactions, toxic effects thereby fastening the process of drug discovery.

2. Big data analytics in developing precision medicine:
Big data generated from various sources such as patient's electronic health records EHR, demographic information, medical records, patient medical history, and genetic constitution of patient would help in the designing and development of personalized medicine. Efficient processing of all the treatment patterns, results, reports, and records can be done using artificial intelligence, software tools and computer models. This further helps in clinical decision support systems (CDSS) and post market surveillance

3. Big data analytics in enhancing the efficacy of clinical trials:
Clinical trials are utmost important for any pharmaceutical products and vaccines to assure that the specific treatment is safe and effective for human subjects. Clinical trials involve selection of appropriate trial subjects which is very difficult, expensive and time consuming. Therefore, pharmaceutical companies use big data for identifying right human subjects for inclusion in clinical trials by completely assessing the data related to patient's genetic information, demography, disease state, personality traits, behavioural patterns, medical history thereby increasing the trial's success rate. This further helps pharmaceutical companies to develop personalized medicine known as precision medicine which is relevant and specific for individual patient.

4. Big data analytics in accelerating Research and Development: [7]
Large volumes of data are generated at different stages from drug discovery and drug designing to its usage by patients who include patient data, medical records. Pharmaceutical companies extract useful information and use them efficiently by using data analytics thereby aiding research and development.

5. Big data analytics in driving sales and marketing:
Interpreting the information from Real word data obtained from various sources such as electronic health records, social media, and medical records would help pharmaceutical companies to explore new markets. This further helps in developing and implementation of new strategies in sales and marketing. Big data analytics in pharmaceutical industry results in more effective treatments.

6. Big data analytics in training and skill development:
There is an immense demand for highly skilled professionals and competent data scientists in healthcare and pharmaceutical sector. They should have good understanding of big data analytical tools and techniques which include cloud computing, data warehousing and should be able to tackle complex data. Professionals need to have knowledge on a range of data analytics and different software skills to efficiently process big data.

VII. BIG DATA IN HEALTHCARE INDUSTRY [9]
Health care industry generated large volumes of data in the form of patient medical history, clinical reports, patient monitoring data, record keeping, and regulatory requirements. Most of the data are manually stored as records and reports. Currently the entire process is digitized where now the diverse data is stored and processed electronically using various data analytics and software technologies.

VIII. APPLICATIONS OF BIG DATA IN HEALTH CARE

1. Big data analytics in controlling Covid -19 pandemic [8]
Covid-19 pandemic showed drastic effects on health and economic sector. It is the big data analytics that played a vital part in controlling the pandemic and taking necessary precautions. By collecting and analyzing the real-world patient data from all sources, it helped us to gain relevant knowledge and helped in better understanding of the virus characteristics and behavior which resulted in tracking the covid-19 cases with increased rate.

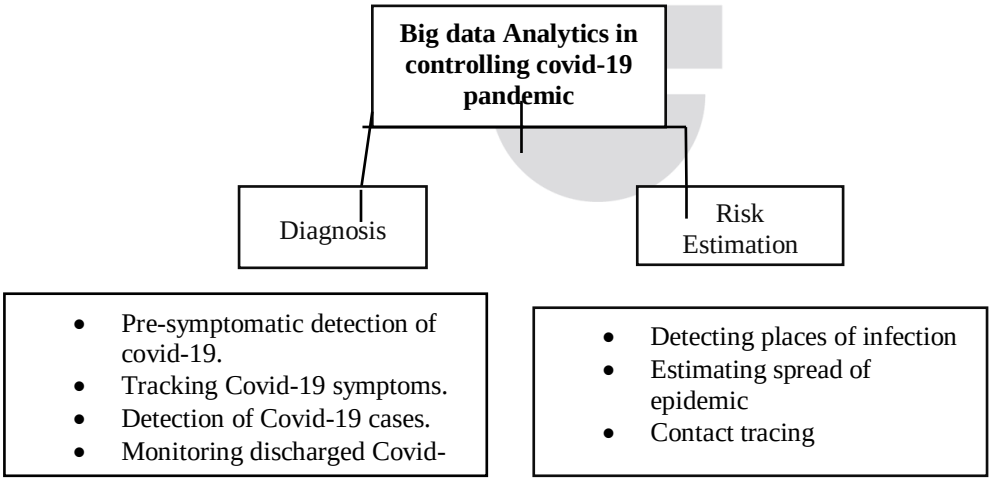

Figure: 1- *Big data analytics in controlling covid-19 pandemic.*

2. Big data analytics in supporting preventive health [9]

Big data analytics can be efficiently used in preventing the occurrence and spreading of diseases in the future.For 0people worldwide are suffering from neurological disorder which are difficult to identify at early stages By using big data analytics, they gathered data from 2,00,000 neurologically healthy patients over 20 years to identify biomarkers for brain disease for early detection and prevention of neurological illness. The data collected include cognitive function reports, eye retina scanning reports, EEG (Electroencephalogram) and MRI (Magnetic Resonance Imaging) reports and neurological information. Big data tools and technologies can be used in integrating the data thereby useful in early prevention of diseases and disorders.

3. Big data in controlling adverse drug events [10]

Big data in the ADE involves the use of the information in Spontaneous Reporting System Database and other novel electronic based systems. It is useful in estimating the risk score which helps in evaluating patient's condition and determining the level of care and priority given for each patient which in turn helps in early detection of adverse events and taking obligatory precautions. This can be made more possible through analytical technique such as data mining. Data mining in healthcare is used to associate large sets of data which includes comparison of patient's medical condition & treatment patterns through statistical analysis. This involves the identification of the humane treatment plan after the detection of ADRs.

4. Big data analytics in cancer treatments

Researchers can draw meaningful insights from patient databases available from different hospitals, universities, biopsy reports of cancer patients, their genomic data, patient treatment records which could help in identifying treatment methods that give accurate results. For Example, a team at Massachusetts General Hospital (MGH), developed a technology which could predict patient's risk of developing breast cancer. Many researchers from different institutes now started using big data analytics and tools along with Artificial Intelligence (AI) to improve cancer treatments.

5. Telemedicine

Big data is used to discover the right dose of medicine for the right course to patients in remote areas through telemedicine. Patient monitoring and medication in the rural areas is possible through telemedicine. It is useful in disease surveillance, medical education and public awareness. Medical practitioners depend on big data analytics to extract useful information from large amounts of medical data which aids in improving diagnosis and test results.

IX. ADVANTAGES IN HEALTH CARE[11,12,13,20]

1. Assures public healthy conducting public health surveillance and predicting occurrence of diseases and taking instant prevention and control measures.
2. Useful in improving the efficiency of research and clinical diagnosis by using various statistical tools and algorithms for designing patient specific treatments
3. Aids regulators to meet the requirements in the development of personalized and patient specific treatments
4. Patient follow-ups: big data analytics helps in keeping a track on patient health condition through various Biosensors, Smartphone apps, smart pills, smart bottles, etc where real-time monitoring of patient health helps in analysing the medication and treatment efficacy.
5. Predicting epidemic by keeping a track on disease outbreaks and preventing them rapid transmission.
6. Used to predict and design models for diabetes, oncology and rare diseases.
7. Used to predict cancer re- occurrence, progression and response to a therapy.
8. Determining and executing appropriate treatments procedures for patients.
9. Efficiently monitors and improves health care system.
10. Promotes health system accountability.
11. Ensures patient safety and improves quality of human life.
12. Building new tools for physicians and consumers.
13. Reduces cost and duration of treatment.
14. Useful in optimizing innovation.
15. Supports clinical decisions.

X. REGULATIONS GOVERNING BIG DATA ANALYTICS

With increase in demand for the usage of Big Data in pharma and health care industry, there are immense challenges faced by the organization in terms of collecting, integrating and storing of data. If these challenges are not appropriately addressed, they may affect quality, accuracy and consistency of data which in turn affects the public. In Europe, the European Commission has established Big Data Task Force to assess the role of big data in health care. However, the process of utilization, development and regulation of Big Data Analytics has been slow in different countries like USA and India.

1. EUROPE

Big Data Task Force (BDTF)

In order to evaluate the role of Big Data in health care systems, the Heads of Medicines Agency (HMA) and European Medical Agency (EMA) together established a Big Data Task Force. This group mainly involves in regulation and evaluation of Big Data to support drug product development, marketing authorization and post- market surveillance. Big Data Task Force generates the roadmaps and list of recommendations for use of Big Data in the evaluation of medicines.

HMA/EMA Big data Steering group [14]

The joint HMA/EMA Big data steering group advices the EMA Management board and HMA on prioritization and planning of actions to implement the 10 priority recommendations under Big Data Task Force

1. Deliver a sustainable platform to access and analyse healthcare data from across the EU (Data Analysis and Real-World Interrogation Network -DARWIN).
2. Establish an EU framework for data quality and representativeness.
3. Enable data discoverability.
4. Develop EU network skills in Big Data.
5. Strengthen EU network processes for Big Data submissions.
6. Build EU network capability to analyze Big Data.
7. Modernize the delivery of expert advice.
8. Ensure data are managed and analyzed within a secure and ethical governance framework.
9. Collaborate with international initiatives on Big Data.
10. Create an EU Big Data 'stakeholder implementation forum'.

Role of BDTF:

a. DATA STANDARDISATION- To establish the regulatory requirements which is used to confirm the validity of the mapped data and also promotes the use of global open-source standards.
b. DATA QUALITY- To establish the data standards and data quality attributes.
c. DATA SHARING AND ACCESS- To develop structured and supportable frameworks for data sharing; promote sharing of qualified models and includes submissions for data management plans.
d. DATA LINKAGE AND INTEGRATION- Facilitate harmonization of similar data sets and support mechanisms to maintain mappings up to date.
e. DATA ANALYTICS- To support novel approaches in modeling of Big Data sets and apply these methodologies for regulatory decisions.

General Data Protection Regulation (GDPR)[15]– GDPR is a regulation in EU law on data protection and privacy in European Union (EU). GDPR was approved by the European parliament on April 14, 2016 and came into effect on May 25, 2018. The main goal of GDPR is to strengthen and join data protection. Especially pharma companies should manage to protect and store data related to medical consent forms, clinical trial subjects, patient data base and medical records.

2. USA

USA does not have well-established regulations governing Big Data specifically, but if any company, organization or health care sector wants to utilize the Big Data activities, they must follow different regulations such as sector-specific privacy laws and region-specific regulations. In US, the US Federal Trade Commission (FTC) regulates the data protection but these regulations are mostly at state level.

3. INDIA

At present, India lacks any specific regulations for Data Analytics. The provisions such as Information Technology Act (IT Act) 2000, Information Technology Rules (IT Rules) 2011 are implemented to lay down protection of personal and sensitive information of patients.

XI. CHALLENGES [16, 17, 18, 19]

- Data Security risks–While the big data analytics helps the organizations, companies to take better decisions but it brings security issues when the companies work with the sensitive information.
- Privacy risks – Lack of privacy in maintaining the confidential medical records is one of the major challenges. There are many laws to maintain privacy of the data such as patient data, medical records etc., but some of them are not applicable for big data sharing which has become a drawback.
- Need for new technologies – The process of incorporating new technologies is veritably slow which could affect the health care systems and research.
- Need for human skills– It is a difficult task in selecting the trained and qualified data scientists to manage and handle the data generated in the organizations or companies.
- Data ownership and Governance – Capturing data has become major hindrance in the health care organizations and companies, due to the lack of effective data governance. The data generated must be clean, precise and properly formatted so that it can be used efficiently.

XII. CONCLUSION

Electronic health records, clinical decision support systems, wearable device technology used to record patient data are considered to be major sources of health information and they are being continuously approved by regulatory agencies such as Food and Drug Administration (FDA) and European Medical Agency (EMA). Such approvals are clear examples as how pharmaceutical and health care industries are depending on digital platforms, Artificial intelligence and big data analytics to collect more reliable patient related information. Traditional data operation systems cannot handle the large volumes of data generated from various sources with an increased rate. Hence Big data technology is a boon. Big data analytics gives us

insights into innovative and cost-effective treatments. During Covid-19 pandemic, big data analytics had played a pivotal role from identifying the source of infection to diagnosis, treatment, prevention, management and control of further spreading of disease. In spite of many advantages and widespread usage of big data, the medical sector face challenges like privacy and data security risks, so there is a need for regulations governing the big data analytics.

REFERENCES

[1] Wullianallur Raghupathi and Viju Raghupathi, "Big data analyticsin healthcare: promise and potential," February 2014.doi: 10.1186/2047-2501-2-3

[2] "How Big Data in Health Care Influences Patient Outcomes"; home/blog/how big data in health care influences patient outcomes; July 7 2021.

[3] **"**Healthcare Big data projects, applications and examples," June 2022; ProjectPro

[4] J. Archenaa and E.A. Mary Anita, "Interactive Big data management in healthcare using Spark," February 2016.doi: 10.1007/978-3-319-30348-2_21.

[5] Tushar Sonal, "8 Ways pharmaceutical companies ensure success with Analytics," August 2019.

[6] "Big data Analytics in the pharmaceutical industry. How is big data analytics revolutionizing the pharma industry," May 2022; Intone

[7] **"**How Big data means big opportunities for Pharma industry," January 2022, i2e consulting

[8] Shikah J. Alsunaidi, Abdullah M. Almuhaideb, Nehad M. Ibrahim, Fatema S. Shaikh, Kawther S. Alquadaihi, Fahd A. Alhaidari, Irfan Ullah Khan, Nida Aslam, and Mohammad S. Alshaharani, "Applications of Big data Analytics to control COVID-19 pandemic," March 2021.doi:10.3390/s21072282.

[9] **"**How Big data analytics can support preventive health"; Features; Health IT Analytics; Xtelligent health care media; May 2022.

[10] C. Lee Ventola, "Big data and pharmacovigilance: Data Mining for adverse drug events and interactions," June 2018.

[11] "Big Data in Healthcare; Examples, Advantages and Disadvantages"; May 5, 2022

[12] William Goddard "Pros and Cons of Big Data"; IT Chronicles

[13] Ashwin Belle, Raghuram Thiagarajan, S.M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard and Kayvan Najarian, "Big data analytics in Healthcare," July 2015.doi:10.1155/2015/370194.

[14] European Medicines Agency; "Big data steering group (BDSG):2021 report," February 2022, Available at: www.ema.europa.eu/en

[15] "What is GDPR, the EU's new data protection law?" GDPR.EU

[16] Raghunath Nambiar, Ruchie Bhardwaj, Adhiraaj Sethi and Rajesh Vargheese, "A look at challenges and opportunities of big data in health care," 2013.doi:10.1109/BigData.2013.6691753.

[17] Nishita Mehta, Anil Pandit, "Concurrence of Big data analytics in healthcare: A systematic review," March 2018.doi: 10.1016/j.ijmedinf.2018.03.013.

[18]"Big data and data analytics"; The potential for Innovation and growth; Briefing, September 2016; European Parliament.

[19]Adriana Gabriela ALEXANDRU, Irina Miruna RADU, Madalina"Big Data in Healthcare – Opportunities and Challenges; June 2018.doi: 10.12948/issn14531305/22.2.2018.05

[20]Antonio Pesquera, Maria Jose Sousa and Alvaro Rocha, "Big data skills sustainable development in healthcare and pharmaceuticals," October 2020.doi: 10.1007/s10916-020-01665-9.