

# Big Data Analytics in Healthcare

Sakshi Raj, Swati Sharma



**Abstract:** *Today, data is constantly increasing and it becomes very hard to handled data skilfully in the usual way. It is also quite expensive, unproductive and very hard to manage data. That's the reason why the super-important technologies are entering in the work place. In the medical field data is increasing day by day and it becomes very hard to manage, analysis and store data in paper pen work. Recent, big data technology, plays major role in the management, organization and analysis of data. Big data technology has been applied towards improving patient profile of care delivery. Biomedical image, generating each day in huge number can be analyzed better with big data technology along with machine learning and artificial intelligence. Research rate in this field has the potential to provide meaningful result in identification of diseases. Change in DNA can predict many future chronic disease. Research also demanded useful, updated, and accurate data. . It becomes possible for public to identify their health risk by their own.*

**Keywords:** *Big data, Big data analytics, Healthcare, NGS, Personalized medicine.*

## I. INTRODUCTION

Now data plays an important role for the development of any part/area. More data we have more organize we are, to produce better results. Healthcare industry has already large data driven by patient care and record keeping. According to a survey carried out in 2011, US health care alone has 150 Exabyte data and didn't take time to reach the scale of zettabytes. Also California based health care network part/area has 44 (quadrillion bytes) of possibly rich data from HER, pictures and notes. These large data helps health care part/area in many ways like improves the quality of health care, reduce price, sicknesses and takes lesser time. These large and complex data can be easily handled by the introduction of electronic health data in health care part/area. It also becomes fast to analyse and distribution of data from point to point. Healthcare information data classified as structured, semi-structured and unstructured (without rules, schedules, etc.) for example, sensor data, (putting pictures into your mind), organizes laboratory testing, free text. Most data in the health system (e.g., doctor's notes, lab test results, and medicine-based data) is used electronically. Significant body-structure-related and pathophysiological events appear at the same time as changes in medicine-based streams.

**Revised Manuscript Received on September 30, 2020.**

\* Correspondence Author

**Sakshi Raj\***, School of Computing Science & Engineering Galgotias University Yamuna Expressway, Greater Noida, Gautam Buddh Nagar, Uttar Pradesh, India Sakshiraj344@gmail.com

**Miss Swati Sharma**, Assistant Professor, School of Computing Science & Engineering Galgotias University Yamuna Expressway, Greater Noida, Gautam Buddh Nagar, Uttar Pradesh, India swatisharma@galgotiasuniversity.edu.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

This results in strong coupling between respectful systems within the body (e.g., interactions between heart rate, breathing, blood pressure, etc.). So, an approach is needed to understand the dependencies of these data. Healthcare is rich in useful data.

Electronic healthcare makes better understanding, analyzing, processing and collection of data. Machine learning and artificial intelligence algorithm solves this complex issue faced by healthcare department in pen-paper work (also many important information remain untouched). In this way one can improve patient profile (by providing efficient personalized medicine). Big data changes challenges like increasing costs and management of data into great opportunities.

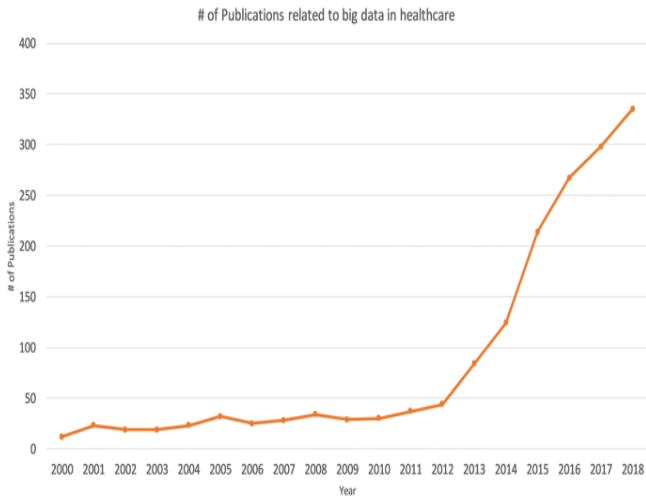
## II. MEDICAL IMAGE PROCESSING FROM BIG DATA POINT OF VIEW

In last twenty years, medical field seem rapid development and continuous revolution of technology. With advanced technology and computer scientist health risk is can identify earlier. MRI, X-ray Produced visible images of inner structure of the body. It helps in medicinal study and treatment. Image processing is needed because even a little bit difference between disease signs of sickness may fall in the different category of sicknesses. So analyzing data correctly and continuous updating is important. Modern technology coined the term computer integration, which makes it easy to disease identification. Medical image data used for reduce cost, developed new therapies, medicines and planning. Here, some sources of useful image way are "MRI, x-ray, molecular imaging, ultrasound, photoacoustic imaging, positron emission tomography-computed tomography(PET-CT)". Large volume of medical image brings transformation in healthcare sector by analyzing images online with different records. Medical image data range, collected form different platforms reached at (million bytes). According to a study conducted worldwide (in 2010) 5 Exabyte medical images processed per year. Alone America covers 50 percent of ionizing radiation exposure. Such data needs/demands large storage ability, if stored for a long period. It also calls for fast sets of computer instructions if any decision using data was to be helped by automation. Modern medical image technologies can produce bright and sharp images such as breathing and lung related problem by four-dimensional figured out tomography. The bright and sharp and dimensions of these images create high large amounts of data needing/ordering high performance figuring out (HPC) and advanced methods for its use. Medical image find out the state of image by analyzing cells of human, animal anatomy. MRI records sound (wave) of internal structure of body. Tumors in lungs and cancer can be identified.



### III. BIG DATA ANALYTICS IN BIOMEDICAL RESEARCH

Any research needed/demanded large and different sets of data from many years. Bio medical research required data from different inputs. Input covers areas like origin of disease, topography of particular disease, rate of spread and discovery rate, medicines, patient personal record and many other. For example, in (1831) a fatal disease called cholera hits America, kills many of its citizens. After few days cholera exposure happens. Many scientist comes and give different theories (one common theory is that cholera is air borne disease). Until a scientist ‘John Snow’ in 1854 stated that cholera is a water borne disease. He makes it possible by tracking all the patient and family and find that all deaths are involved around a public hand pump in ‘Soho District’. Later measure step taken by local authorities to remove that pump. As research required large sets of data and genomic record, new generation sequencing (NGS) started to facilitate in healthcare sector. It mainly used for clinical practices and becomes the part of routine checkup. It allows data pooling in research data base. The main goal is the development of effective therapies and so the improvement in health care. It allows to understand the interdependencies of complex diseases. Researchers can analyze data from different dimension from point to point. “Publications shows a positive, continue growth in health care in large figures”. Latest technology being introduced in this sector raised revolution of near future.



### IV. HEALTH CARE SECTOR AND BIG DATA ANALYTICS:-

- **Patient role in health care analytic**  
Healthcare field can be improved by anyone through small basis and personal data. In 2012, Swan created the term “personscience” where non-professional are educated and skilled enough to operate and support healthcare system.
- **Connectivity between medical staffs and patients**  
Connectivity is important so that people get connect and share their thoughts. In (2009) McHorn stated that healthcare industry is more than technology and knowledge. It should also connects people which make them able to share their personal skills such as ability to learn and adopt in their personal lives.

- **Healthcare and predictive decision support**  
Advanced healthcare not only reduces costs but also can (describe a possible future event) the sicknesses. Which can be reduced by taking proper steps towards it.

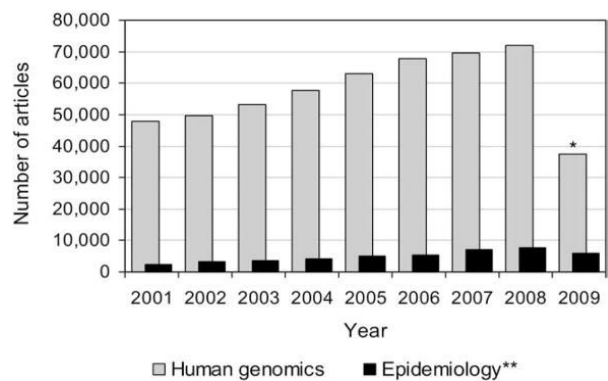
For example, a hospital named Parkland in Dallas introduced forecasting system. Which scans the patient details and identifies possible risks. Hospital saved half a million dollars in heart failure, which is a heart success

- **Healthcare (information-giving numbers) and Real Time**

Murphy (2013) stated that real time (information-giving numbers) produces more results and information, as it has history of patients, therefore providing the best treatment by properly examining the of patients record.

### V. BIG DATA APPLICATION IN GENOMICS

Genomics stand for the study of all individual genes, its character, and qualification. This covers the overall structure of human/animal DNA its function, evolution, mapping and editing. Till date total 30,000 to 35,000 genes were discovered. Disorder in DNA cause a number of chronic disease which can be controlled by taking major steps, by analyzing change in DNA. That’s the reason why modern technology (NGS) new generation sequencing being introduced. Experts can find the interdependencies between diseases just by observing change in DNA. In 1866, Gregor Mendal recognize a genetic disorder ‘Pleiotropy’ which carry multiple disease manifestation even a small change in genes happens. Also diseases like cancer can cured on time by predictive analyzing algorithm. In 2013, ASCO launched ‘Cancer LINQ’. ‘Data lake infrastructure’ used to store large volume data by using SAP’s HANA system. These data shared by many organization. Reconstruction of network got advanced over last twenty years.”figure showing result of publications discovery of human/animal genes”. Electronic health care is like biggest boost in health care sector. Psychological disorder, that positive and negative thinking can be identified at first state by observing change in DNA.



\*PubMed and HuGE Navigator queries performed October 1, 2009  
\*\* population-based association studies, meta-analyses, gene-environment interactions, evaluation of genetic tests

## VI. CHALLENGES ASSOCIATE WITH HEALTHCARE

Despite of many benefits of big data in healthcare. It carries many challenges like cleaning of data, updating, storage and extraction of data, security and privacy. On which steel need to work for better future. So that organization can take more advantages of it.

### ● Data cleaning

Cleaning of data improves data quality. Hence, improves overall productivity. Dirty, incorrect and outdated data doesn't make any sense. This is done by manually and automated. This method is termed as data scrubbing. Useful data makes sense for accuracy and decision support.

### ● Security and Privacy

Big data healthcare has large volume and variety of data. Continuous streaming of data from dataset, makes difficulty in data security. MapReduce framework is being used for data processing. It splits data, then processes data by mapper algorithm, which is not good in security.

### ● Data capturing

It became difficult to gather data from different platform and organize them in database. According to a study, perform by eye specialist clinic 'ophthalmology clinic' patient with only 2-3 symptoms eye matched with 23.5 percent of HER records.

### ● Querying

Data querying allows great use of data form database. It basically uses hive for table creation and underlying concept

### ● Updating

To avoid duplicate records, continuous updating required. Organization should be sure that the information up to date. So that patient care done more accurately and affectively.

### ● Sharing

Every individual organization should their own data with larger database of healthcare. So that overall health organization utilize these data.

## VII. CONCLUSION

There is a great need to launch big data analytics in health care for each and every department so that it becomes possible to gather data from large to small scale. As those data is essentially important. Today's world population is around 8-billion (according to world meter 2020) and health care has only 153 Exabyte's useful data. More data gives more accurate result. Also day to day new genetic disorder recognize. If new technology introduced, it reduces many disease failure like from lungs tumor to spinal cord disorder, from cancer to many chronic diseases. Heart failure can also be under controlled. All hospitals, organizations should corporate with electronic health care. Not only hospitals but non-professional educated, skilled person should also share their experience. So that healthcare reached at its highest peak to help every individual by analyzing data and predicting future diseases accurately. Health data found in many forms that is structured and unstructured data. All consistent and organized data comes under structured data account. Like rate of flow of blood, height, weight and blood type. What absent in structured comes under unstructured. It mainly consists of paper-pen work that is images, faxed copies and physician copies. Research

required both structured and unstructured data. For quality control of data with its accuracy machine learning technology should be developed.

## REFERENCES

1. Andrew McAfee, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. Big data: the management revolution. Harvard Bus Rev, 90(10):6068, 2012.
2. Wen-Syan Li, Jianfeng Yan, Ying Yan, and Jin Zhang. Xbase: cloud-enabled information appliance for healthcare. In EDBT, pages 675680, 2010.
3. Mendelson, D. (2017). Legal protections for personal health information in the age of Big Data—a proposal for regulatory framework. Ethics, Medicine and Public Health, 3(1), 37-55.
4. Sacristán, J. A., & Dilla, T. (2015). No big data without small data: learning health care systems begin and end with the individual patient. Journal of Evaluation in Clinical Practice, 21(6), 1014-1017.
5. Ulfarsson, M. O., Palsson, F., Sigurdsson, J., & Sveinsson, J. R. (2016). Classification of big data with application to imaging genetics. Proceedings of the IEEE, 104(11), 2137-2154.