Manoj Kashyap

kashyap.ma@northeastern.edu

big data analytics in medical and healthcare

Industrial Analysis(May 2020)

Abstract

In the 21st generation, both private and public sectors industries generate, store, and analyze data to improvise their services. Healthcare industry and Biomedical research areas are generating a significant amount of Big Data related to hospital records, patients’ medical records, results of medical examinations, and research in related to vaccines. Analyzing Big Data has given solutions to various challenges in the healthcare sector. In today’s generation healthcare sector is supported by many advanced data-driven devices and infrastructures. With the ample data generated in the biomedical sector, modern healthcare organizations are about to make a revolution in medicine and public health.

Introduction

Information is the fuel and insights are the keys to every organization in the modern world. That is the reason why the collection of data and analysing them is the most important aspect of every organization. Almost every organization is generating tones of data related to work, health, banking, retail, insurance, energy, transportation, science, activities, etc. This data deluge is creating a large amount of unmanageable data which is referred to as ‘Big Data’. Now organizations are focused on managing and organizing this data to develop new strategies and derive meaningful information out of it. Businesses are boosting their profits and efficiency using advanced analytical tools to make data-driven decisions. One such mandatory organization in meeting social needs in healthcare. “Beyond improving profits and cutting down on wasted overhead, Big Data in healthcare is being used to predict epidemics, cure disease, improve quality of life, and avoid preventable deaths”[2]. In this review, we will be discussing Big Data, organizing and analysing big data, and the impact of big data analytics in the medical and health care sector.

Big Data

Internet giants like Google and Facebook are generating a massive amount of data every day. In one or the other way people working in various organizations are contributing to generate this digital form of data in this “digital universe”. International Data Corporation (IDC) estimated the total size of data in this digital universe in 2005 to be 130 exabytes (EB), by 2017 it was expanded to 16,000 EB or 16 zettabytes and in 2020 it is predicted to reach 40,000 EB or 40 zettabytes [[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0). Upon distributing it, it would count to 5,200 GB of data to each individual. This humongous amount of data shows the speed in which the digital universe is expanding and generating significant revenue in all the industries. Since the data is too abundant to manage by conventional internet-platforms and software, the term ‘Big Data’. Both the strategy development process and the actual strategies developed are impacted by Big Data in majorly 6 ways, often called as 6v’s, the below Figure 1 shows it [[9]](https://www.researchgate.net/publication/318338677_Components_of_Big_Data_Analytics_for_Strategic_Management_of_Enterprise_Architecture).

A screenshot of a cell phone

Description automatically generated

Figure

Almost every sector of research is generating and analysing big data for various purposes. So to handle this unmanageable big data, and to use it for the decision making processes, various organizations are trying to automate the process by advanced machine learning algorithms and AI techniques.

Big Data in Healthcare

Healthcare has turned out to be a big data repository with the aim of prevention, diagnosis, and treatment for the welfare of human beings and other animals. Implementation of big data analytics in healthcare can potentially reduce the cost of treatment and minimize waste and error, with improved quality of service. The majority of the healthcare industries are using natural language processing (NLP) as a big data analytical technique with the help of Hadoop for the optimization of clinical operations and reduction of cost of care [[8]](https://www.sciencedirect.com/science/article/abs/pii/S1386505618302466). Healthcare professionals serve patients on several levels such as the first point of consultation (primary care), acute care requiring skilled professionals (secondary care), advanced medical investigation and treatment (tertiary care), and highly uncommon diagnostic or surgical procedures (quaternary care) [[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0). In all the different levels of treatment healthcare professionals are responsible for the gathering of patients data such as medical history, clinical data, and other personnel information of patients. Murphy, Hanken, and Waters termed computerized medical data of patients relating to the past, present or future physical/mental health as “Electronic Health Records”(EHR). EHR uses patients' data stored in electronic systems to capture, transmit, receive, store, retrieve, link, and manipulate multimedia data to provide improvised health-related services [[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0).

Similar to EHR, electronic medical record (EMR), personal health record (PHR), medical practice management software (MPM), and many other records related to healthcare data components collectively work together to improve the quality of medical service.

A close up of a device

Description automatically generated

Figure 2

*Figure 2* shows the data warehouse, analytical pipelines and workflow of Big Data Analytics in healthcare to provide smarter and cost effective healthcare services.

Big Data in Biomedical research and omics studies

Understanding the biological system in cellular and genetic level requires multiple experiments to generate a wide spectrum of data. More the data is available, it’s better to understand complex biological processes. Many technologies such as next-generation sequencing (NGS) and Genome wide association studies (GWAS) are gathering data to decode human genetics[[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0). The ‘omics’ discipline of study has helped scientists to broaden their area of research. Instead of studying single ‘gene’, now scientists are studying entire ‘genomics’ of an organism, instead of studying ‘transcription’ of single gene now, we are studying entire ‘transcriptomics’ of an organism. This shows the depth of information that is been generated in the omics era.

NGS has reduced the cost of genomics and transcriptomics sequencing from millions to a couple of thousand dollars with simplified sequencing [[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0). The genomics and transcriptomics data are now combined with proteomic and metabolomic data to enhance the individual profile of the patient, which is ascribed as “individual, personalized or precision health care”.

A screenshot of a cell phone

Description automatically generated

Figure 3

*Figure 3* shows A framework for integrating omics data and health care analytics to promote personalized treatment. This helps healthcare professionals in making personalized medical decisions. Upon analysing this big data, this technology is about to change the future of medicine and healthcare.

Image Processing Analytics

The medical domain also uses some of the Image processing and analysing techniques such as computed tomography (CT), magnetic resonance imaging (MRI), X-ray, molecular imaging, ultrasound, photo-acoustic imaging, functional MRI (fMRI), positron emission tomography (PET), electroencephalography (EEG), and mammograms which captures various anatomical high definition images. Picture Archiving and Communication System (PACS) helps to store and access medical images and reports in local workstations with the help of protocol such as digital image communication in medicine (DICOM). Various tools have been built for the visualization and simulation of medical images in order to analyse and dig hidden information out of them. Even there are toolkits available that produces 3D images of medical data for the detailed analysis.

Commercial platforms

There are many commercial platforms available for analysing healthcare data such as:

* IBM Watson which collaborated with Pfizer, discovering novel immune-oncology combinations to target novel drugs specifically for cancer disease model.
* AYASDI analyses and manages clinical variations and healthcare costs, it also helps in managing hospital organizations and provides a conversation platform for the doctors in treatment.
* Linguamatics implements the NLP algorithm to provide information on genetic relationships, with the help of an interactive text mining algorithm (I2E).

Analytical Challenges in Healthcare

People have accustomed to various personalized services in many sectors such as banking, insurance, retail, and others. But even in 2020 healthcare domain seems to be lagging in the usage of advanced analytical tools for the decision-making process. Many hospitals, especially in rural levels, are staggeringly inefficient in providing accurate medicine for lower costs [[6]](https://www.forbes.com/sites/forbesbusinesscouncil/2020/03/11/the-future-of-healthcare-data-driven-personalized-medicine-at-scale/#1edce2013629).

Handling Big Data in the medical domain will face challenges in multiple levels:

* The storage of data should be in a hybrid way to be flexible.
* Accurate ML algorithms need to be implemented to maintain clean and scrubbed data to ensure accuracy, consistency, and relevance.
* Since patients produce a massive amount of data, it is crucial to maintain it in a Unified format using medical coding systems such as Current Procedural Terminology (CPT) and International Classification of Diseases (ICD).
* Accuracy is one critical aspect when it comes to healthcare. Many studies show that because of complex workflows, EMRs or EHRs are not entirely accurate yet [[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0).
* Misinterpretation in medical Image processing because of noise and contrasts leads to poor results.
* There are many security breaches reported by hackings and phishing attacks. HIPAA Security Rules have been implemented by protected health information (PHI) to safeguard medical data. Using anti-virus software, firewalls, encrypting sensitive data also helps in maintaining Security.
* Absence of proper interoperability between datasets the query tools may lead to difficulty in querying data to get answers. Medical coding systems like ICD-10, SNOMED-CT, or LOINC can be implemented to improvise querying.
* Patients in remote places may not be updated with their results in time. Fast Healthcare Interoperability Resource (FHIR) and public APIs with mobile healthcare(mHealth), Data sharing could be made easier.

Big Data Analytics in understanding COVID-19 pandemic

Upon declaring COVID-19 as a pandemic by WHO, there has been rigorous usage of big data analytical tools in research, to monitor and reduce the impact of the virus. “There are several big data components to this pandemic where artificial intelligence can play a big role,” said James Hendler, an artificial intelligence researcher at Rensselaer Polytechnic Institute, United States [[7]](https://healthitanalytics.com/news/understanding-the-covid-19-pandemic-as-a-big-data-analytics-issue). Supercomputers at IBM are working rigorously in collaboration with medical research teams in finding a cure. Industry giants like Google and Amazon have launched projects in cloud and AWS offering free access to open datasets and analytics tools for researchers, helping them develop COVID-19 solutions faster [[7]](https://healthitanalytics.com/news/understanding-the-covid-19-pandemic-as-a-big-data-analytics-issue).

Career as a Healthcare Data Analyst

In the healthcare industry, I personally aspire to play the role of healthcare data analyst--sometimes called healthcare business analysts or health information management (HIM) analysts [[10]](https://himt.wisconsin.edu/experience-uw-himt/healthcare-data-analyst/). With the improved Big-data based advance medical practices in the past 20 years, 18% of the GDP accounted for healthcare in the United States, totalling about US$3.4 trillion [[11]](https://www.evariant.com/faq/what-is-healthcare-big-data). By 2018 global healthcare analytics was worth US$19.6 Billion, and the market is projected to reach about US$ 47.7 Billion by 2024 with a CAGR of around 16% in the years 2019-2024 [[12]](https://www.businesswire.com/news/home/20191121005779/en/Global-Healthcare-Big-Data-Analytics-Market-Report). This shows the high demand for data analytics in healthcare domain.

An HIIM analyst required to gather and interpret various medical data such as electronic health record, billing claims, cost reports, and patient satisfaction surveys, helping an organization to improve the quality of service. Being a person with a lot of interest in the medical sector and social service, I believe this position perfectly suits me.

Ethical issues with Big Data Analytics in Healthcare

“Ethics is about goods that we have a reason – and sometimes even an obligation – to pursue, such as the good of knowledge that can be used to bring about significant improvements in health” [[4]](https://www.sciencedirect.com/science/article/pii/S2452310017300264). In the 21st century, ethical guideline are about to collide with Artificial Intelligence(AI). The healthcare industry is implementing big data analytical tools to transcribe medical records, assist surgery, investigate insurance claims for fraud, and for many other services which are predicted to reach a market value of about $20 billion by 2024 [[5]](https://www.forbes.com/sites/insights-intelai/2019/02/11/rethinking-medical-ethics/#6b42ef466f03).

Big data analytics is implemented through Artificial Intelligence(AI) models. The question is, who is responsible if the AI system makes a wrong decision? Who is responsible to protect patient privacy?

In 2018 the American Medical Association (AMA) first designed some guidelines on usage AI systems, referring AI as “augmented intelligence”, says AI should enhance, not replace, the work of physicians [[5]](https://www.forbes.com/sites/insights-intelai/2019/02/11/rethinking-medical-ethics/#6b42ef466f03). It also adds that it should identify and address bias, and should be transparent and protect patient privacy. But these policies do not answer all the ethical challenges.

In 2017, the data analytics team at the University of Chicago Medicine (UCM) designed a machine learning algorithm to predict how long a patient might stay in the hospital. They found the ZIP code attribute to be in strong correlation, which was strongly connected with the patient’s race and socioeconomic status. This had scrutiny on African-Americans from Chicago’s poorest neighborhoods, who stayed for a long time in hospitals. The team observed that this algorithm seems to be biased and unethical, which may contribute to perpetuating bias and lock in existing inequalities in healthcare. “How do we, as machine-learning practitioners, handle societal problems while trying to make accurate, actionable models?” asks John Fahrenbach, a UCM data scientist who worked on the algorithm. He suggests seeking expertise outside the UCM diversity group [[5]](https://www.forbes.com/sites/insights-intelai/2019/02/11/rethinking-medical-ethics/#6b42ef466f03). This shows there is always a Black Box problem in understanding the thought process of algorithm, which absorbs large quantities of data and identify statistical patterns often without explicit instruction.

Many biomedical organizations are conducting research on human genomes for the last two decades. Considering privacy of such data which may cause genetic discriminations, in 2008 United States adopted the Genetic Information Non-discrimination Act (GINA), in 2016 the Council of Europe recommended that for insurance purposes genetic tests neither be required nor used and in 2017 Canada has adopted legislation prohibiting the use of genetic data by life insurers [[4]](https://www.sciencedirect.com/science/article/pii/S2452310017300264). Council for International Organizations of Medical Sciences (CIOMS) has issued common rules and ethical guidelines on health related research.

Conclusion

In today’s world, various biomedical and healthcare tools such as genomics, mobile biometric sensors, scanning devices, and smartphone apps generate a huge amount of data. Advanced personalized healthcare frameworks have resulted in building better healthcare and clinical systems. Analysing big data has helped biomedical and pharmaceutical researches in finding accurate vaccines and medicines by studying the pattern and behaviours of the disease and related DNA or RNA strains.

Big data analysis can provide tremendous solutions to the challenges in healthcare, that haven’t seen an answer in centuries. Big Data can be gathered for improved health research and clinical care since quality data can improve the quality of clinical care. We must ensure that there is no misuse of the data gathered and the organizations' privacy policies fit the 21st century’s standard ethical and privacy policies.

“It takes time to recruit people, find sites for clinical trials, and establish protocols. Once all is done, these large vaccine trials can go anywhere between one to five years” says Supriya Munshaw Senior Lecture at Johns Hopkins Carey Business School on today’s pandemic COVID-19. This shows that the utility of big data analytics in healthcare and medicine is not yet in its peak, and is yet waiting to leverage at its best to improve the quality of life.

References and Citations:

[[1]](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0)

[[1] Big data in healthcare: management, analysis and future prospects](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0)

[[2]](https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#244c1e172873)

[[2] Forbes : How Big Data Is Changing Healthcare](https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#36b480e22873)

[[3]](https://www.datapine.com/blog/big-data-examples-in-healthcare/)

[[3] 12 Examples of Big Data Analytics In Healthcare That Can Save People](https://www.datapine.com/blog/big-data-examples-in-healthcare/)

[[4]](https://www.sciencedirect.com/science/article/pii/S2452310017300264)

[[4] Ethics and Big Data in health](https://www.sciencedirect.com/science/article/pii/S2452310017300264)

[[5]](https://www.forbes.com/sites/insights-intelai/2019/02/11/rethinking-medical-ethics/#6b42ef466f03)

[[5] Forbes : Rethinking Medical Ethics](https://www.forbes.com/sites/insights-intelai/2019/02/11/rethinking-medical-ethics/#42a224196f03)

[[6]](https://www.forbes.com/sites/forbesbusinesscouncil/2020/03/11/the-future-of-healthcare-data-driven-personalized-medicine-at-scale/#1edce2013629)

[[6] Forbes: The Future Of Healthcare: Data-Driven Personalized Medicine At Scale](https://www.forbes.com/sites/forbesbusinesscouncil/2020/03/11/the-future-of-healthcare-data-driven-personalized-medicine-at-scale/#470221a33629)

[[7]](https://healthitanalytics.com/news/understanding-the-covid-19-pandemic-as-a-big-data-analytics-issue)

[[7] Understanding the COVID-19 Pandemic as a Big Data Analytics Issue](https://healthitanalytics.com/news/understanding-the-covid-19-pandemic-as-a-big-data-analytics-issue)

[[8]](https://www.sciencedirect.com/science/article/abs/pii/S1386505618302466)

[[8] Concurrence of big data analytics and healthcare: A systematic review](https://www.sciencedirect.com/science/article/abs/pii/S1386505618302466)

[[9]](https://www.researchgate.net/publication/318338677_Components_of_Big_Data_Analytics_for_Strategic_Management_of_Enterprise_Architecture)

[[9] Components of Big Data Analytics for Strategic Management of Enterprise Architecture](https://www.researchgate.net/publication/318338677_Components_of_Big_Data_Analytics_for_Strategic_Management_of_Enterprise_Architecture)

[[10]](https://himt.wisconsin.edu/experience-uw-himt/healthcare-data-analyst/)

[[10] WHAT IS A HEALTHCARE DATA ANALYST? HOW DO YOU BECOME ONE?](https://himt.wisconsin.edu/experience-uw-himt/healthcare-data-analyst/)

[[11]](https://www.evariant.com/faq/what-is-healthcare-big-data)

[[11] What is Healthcare Big Data?](https://www.evariant.com/faq/what-is-healthcare-big-data)

[[12]](https://www.businesswire.com/news/home/20191121005779/en/Global-Healthcare-Big-Data-Analytics-Market-Report)

[[12] Global Healthcare Big Data Analytics Market Report 2019-2024 - ResearchAndMarkets.com](https://www.businesswire.com/news/home/20191121005779/en/Global-Healthcare-Big-Data-Analytics-Market-Report)