BIG DATA IN HEALTH CARE

Bhumika Agarwal
Ritika Agrawal
Namit Gupta
BSc (H) Computer Science, TMU, Moradabad
BSc (H) Computer Science, TMU, Moradabad
Assistant Professor, TMU, MBD
bhumikaagarwal.dpr@gmail.com
Ritikaagrawal008@gmail.com
namit.k.gupta@gmail.com

Abstract— **Healthcare is one of the business fields with the highest Big Data potential. According to the prevailing definition, Big Data refers to the fact that data today is often too large and heterogeneous and changes too quickly to be stored, processed, and transformed into value by previous technologies.**

What is Big Data in Healthcare

"Big data in healthcare" refers to the abundant health data amassed from numerous sources including electronic health records (EHRs), medical imaging, genomic sequencing, payor records, pharmaceutical research, wearables, and medical devices, to name a few. Three characteristics distinguish it from traditional electronic medical and human health data used for decision-making: It is available in extraordinarily high volume; it moves at high velocity and spans the health industry's massive digital universe; and, because it derives from many sources, it is highly variable in structure and nature. This is known as the Data. With its diversity in format, type, and context, it is difficult to merge big healthcare data into conventional databases, making it enormously challenging to process, and hard for industry leaders to harness its significant promise to transform the industry.
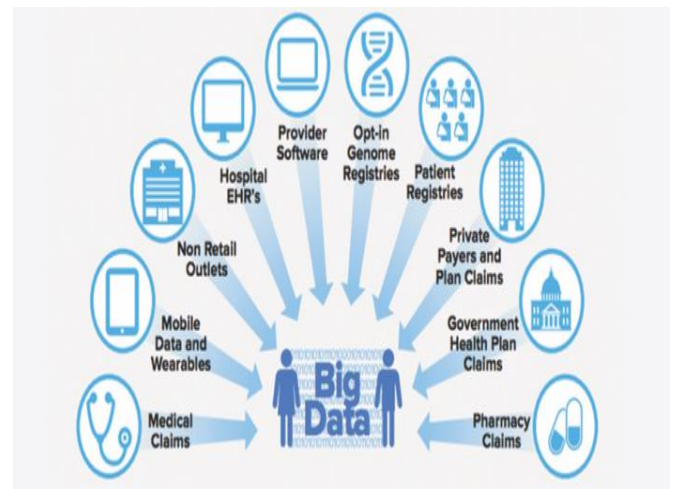
Now that we live for a long time, treatment models have changed and many of these changes are operated by data. The doctor wants to understand as much about the patient as he can understand about a patient and the disease of his life - treatment at the initial level of any disease is far simpler and less expensive. With healthcare data analytics, prevention is better than cure and provides a comprehensive package of management insurance to take a comprehensive picture of a patient. The industry is trying to solve the problems of patient data: bits and bytes are collected everywhere and are stored in hospitals, clinics, surgeries, etc., so that communication can be done properly. Indeed, over the years, consuming a large amount of data for therapeutic use is expensive and time-consuming. With today's always better techniques, not only is this easy to assemble such data, but it is converted into relevant critical insights that can be used to

provide better care. This is the purpose of healthcare data analytics: it is too late to use data-driven search before estimating and solving a problem, but simultaneously assess method and treatment.

The Usefulness and Challenges of Big Data in Healthcare

Big data in health informatics can be used to predict outcome of diseases and epidemics, improve treatment and quality of life, and prevent premature deaths and disease development. Big data also provide information about diseases and warning signs for treatment to be administered. This will help not only to prevent co-morbidities and mortality but also assists government to save the cost of medical treatment. It is very useful not only in clinical medicine for diagnosis/detection but also in epidemiological research as the big data will provide huge amount of data. The government, non-governmental organization and/ or pharmaceutical companies can use the data to formulate policies, strategies, intervention or medical treatment such as drugs development. Big data has implications on healthcare on patients, providers, researchers, health professionals . Nowadays, there is an increasing demand for more information by the patients about their healthcare options or choices, and want participation in their health decision-making. The big data will help to provide patients with up-to-date information to assist them to make the best decision and to comply with the medical treatment. The Malaysia National Health and Morbidity Survey in 2015 has revealed that the number of obese Malaysians have risen to 17.7% compared to 4.4 % in 1996 and 17.5% of those aged 18 and above have diabetes compared to 11.6% in 2006. There is a need to capture and analyse this raw data information to provide better healthcare, accessibility, affordability and quality of healthcare from diagnosis, treatment and follow-up. In 2017, the Ministry of Health Malaysia (MoH) has launched the Malaysian Health Data Warehouse (MyHDW) to share information patient's medical records and knowledge among public, private hospitals and clinics. MyHDW aims to synchronise patients' data from public hospitals (including university hospitals, armed forces hospitals), private hospital and clinics along with

National Registration Department (NRD), National Department of Statistics, and other health related agencies where it will serve as one stop centre to provide healthcare providers to make decisive decision on treatments. Most of the time, medical data are collected in silos in their respective healthcare centres and is governed and controlled by hospitals or clinics administrative departments. Should big data is successful implemented in Malaysia, it will reduce wasteful overheads and effective managed.The aim of this manuscript to highlight the usefulness and challenges of big data in healthcare worldwide generally as well as country like Malaysia.

Advantages of Big Data Big data
could reduce the recency bias or recency effect bias. Recency bias occurs when the recent events are weigh more heavily than earlier events in order to improve the situation, but it may lead to incorrect decisions. The real time information can also be incorporated into big data. Real time big data has many advantages. For example, any errors or trouble shoot in an organization can be identified immediately and the operational problem can be overcome. This will save time, cost and increase the productivity. The services also can be further improved as the real time provides the latest information on the subject matter. For instance, it will provide the complete information on the patients and at the same time able to administer medical intervention without any delay.

In healthcare, big data is also used in predictive analysis which is to identify and address the medical issues before it becoming an unmanageable problem. Healthcare professionals are able to reduce the risk and overcome the issue with the information derived from the big data.

Apart from that, big data is also able to help identify frauds in healthcare especially on insurance claims. Fraudulent, inconsistency and false claims can be flagged. This will facilitate insurance companies to prevent losses.

Big data can also benefit healthcare through data management, electronic medical records and data

analysis. The big data will help to find and identify the right population or target group. Big data consists of diverse group of population and certain group can be identified for risk assessment and screenings. The existence of big data will also allow development or modification of a program or intervention to target the health problem. It will enables clinical trials to be initiated immediately. Big data will give a clearer picture on the type of population as well as their medical problem. The pattern of the distribution or disease information will allow quick development of intervention program as well as targeting the affected group as early as possible. Data growth of pharmaceutical industries were derived from patients, caregivers, retailers and Research and Development (R&D). Big data could facilitate the pharmaceutical companies to identify new potential and effective drugs and deliver it to the users more quickly .

Issues with Big Data

There is a huge challenge in big data in terms of data protection, collection and sharing of health data and data usage. Big data analytics with the use of sophisticated technologies has the potential to transform the data repositories and make informed decisions. Issues such as privacy, security, standards and governance to be addressed. Information such as nano particulate therapy on cancer treatment could be also be incorporated in big data to provide an overview and best treatment for cancer especially when nanotechnology is important in drug delivery in cancer treatment. Apart from that adverse effects of drugs use could also be determined.

Security

Since the big data contained subject's personal information and their health history, it is important for the database to be protected from hacking, cyber theft and phishing, where the stolen data can be sold for a huge sum.Apart from the health information and personal information from the health system which can be hacked or stolen, other big data in other commercial organizations such as telecommunications companies (telcos), banks or financial institution are also vulnerable without the

knowledge of the clients. Before big data can be implemented, it is necessary to ensure that the administration, privacy, security of the big data are well protected. Protection health information via transmission security, multilayer authentication, using anti-virus software, firewalls, data encryption are indeed vital. As the data becomes more regional and global, it become more complicated and have more serious impact on security, standards, language and terminology. The accessibility of the healthcare data need to be consistently reviewed and monitored.

Data Classification

Big data is a massive, less structured and heterogeneous. As such there is a need to identify and classify the data so that it can be used effectively. However, it is laborious to search for a specific data in the big data. The big data also required to be contextualized or pooled together so that it will become more relevant to specific individuals or groups.

Data Modeling

Although big data is excellent for modelling and simulation, there is a need to identify, structure and pool the proper relevant data so that it can be used to model the problems, which later can be used for intervention. Without the proper structured data, it is challenging to analyse and visualize the output and to extract specific information or data.

Cloud Storage

The cloud storage can be used to upload data or having the whole system designed in the cloud. Thus, the cloud will need to have sufficient space for the storage and sufficient speed for data upload at the same time. The storage apart involving words documentations, it should also able to store graphic type such as X ray, CT or MRI. The system should also be able to generate graphic presentations from the available data so that clinicians are able to visualize and understand quickly and take prompt decision .

Data Accommodation

One simplified big data system is require to accommodate all the data and it has to be compatible and simplified. This is to ensure that the users are able to retrieve the information without any hassle. It is a difficult task to get all the relevant systems to link to each other.

There is a culture of dissonance within individual organizations, where some parties may control the data for their own needs rather than for the organization as a whole.

Data Personnel

At this time, it is still an arduous task to find data scientists with expertise in statistics, computer science or information technology (IT). A standard protocol need to be in place for data entry so that all information entered are standardized by data entry person even though there will be changes in the data entry personnel. This is to ensure the continuity and standardized format of data entry.

Miscommunications Gap

The miscommunications or the gap between the users and data scientists is one of the biggest problems in relations to big data. The understanding of the users on data generated by data scientists' maybe low and this may affect the effective usage of big data. The health data from all clinics and hospitals need to be pooled together as stored at one-stop centre (big data). At the moment, all the information are kept separately. As such, it is difficult to get a clearer picture of the patients due to the incomplete information gathered. Thus, this waste a lot of time as the doctor will need to start all over from the beginning taking the patients history. Since big data has the ability to predict future medical issues which is a positive thing, big data can also pose risk and undermine doctors. The patients too will rely on the technology rather consulting the healthcare practitioners.

Data Nature

The integration of data will not only involve data within the healthcare system but also external data. Although it gives potential benefits, it is also challenging in terms of privacy, security and legal matters. The healthcare data usually consists of patients who are seeking treatment in the hospitals or clinics but none on healthy individuals. With the inclusion of healthy individuals in the database, it will help to provide better understanding on the nature of the disease and intervention. As the data becomes more current, it is necessary that the information are passed to the users immediately for clinical decision making and to improve the health outcomes.

Technology Incorporation

Lack of information to support the decision making, policy planning or strategy is one of the problems in big data. The processes of redefining and in adopting of technology is slow and this can impact the healthcare, care delivery and research. Without the technology, big data is unable to generate and disseminate information. Most the time, data are fragmented and dispersed among various stakeholders such as providers, vendors, organizations and payers. The solution to this is to have all the data uploaded in one 'warehouse'.

Conclusion

Big Data has a great potential changing the healthcare outlook such as in drug discovery, patients personalization care, treatment efficiency, improvement in clinical outcomes, and patients safety management.