International Journal of Healthcare Sciences ISSN 2348-5728 (Online)
Vol. 9, Issue 2, pp: (59-63), Month: October 2021 - March 2022, Available at: www.researchpublish.com

A Literature Review on Big Data and Data Science Application in Healthcare

Nattapong Chitsongsawat

Roi-Et Wittayalai School

***Abstract: Background*: Healthcare systems have been digitally transformed by technological advancement in medical information systems, electronic medical records, wearable and smart devices, and handheld devices. In addition to technological advancement in the healthcare field, this increased number of medical and health data enabled researchers and practitioners to extract and visualise medical large data in a new spectrum.**

***Objective*: This study aimed to review big data and data science applications in healthcare and present in a summary for junior learners to easily understand.**

***Methods*: A literature review on big data and data science application in healthcare and technology-related articles and researches.**

***Result*: Big Data and Data Science application in healthcare mainly to increase effectiveness and efficiency of diagnosing, treatment, taking care of patients, healthcare service & operation, prediction and prevention of illness. The application of data science in healthcare required comprehensive and structured data in a digital format with sophisticated data analytical systems. Applying big data and data science in healthcare requires the availability of many matters in different dimensions, besides policy and budget, collecting health data, data integrity, storing it in digital form, availability of hardware and software, trained staff, responsible persons, data workers, data sharing, connectivity, utilisation of data analytics and availability of health data scientist.**

***Keywords*: data science application, healthcare, big data.**

1. INTRODUCTION

There's an increasing need for big data in healthcare due to rising healthcare costs in nations like the United States. A McKinsey's report stated that healthcare spending has been steadily increasing for more than 20 years and represented 17.6 percent of GDP which was close to $600 billion (1). This level of spending on Healthcare costs is considered quite high. Clearly, the need to cut spending and increase efficiency have been worked on. There are many technology adoption in healthcare to increase efficiency and reduce the overuse of unnecessary tests and treatments. Optimizing patients' health data has been found to be one smart way due to the sophistication of patients' health data could support medical diagnosis to be more precise, therefore the patients' illness could be treated right to its' causes and the accessories could be avoided. Current incentive systems for health care providers in the U.S. are not encouraging to minimize healthcare costs, many insurance companies are switching from fee-for-service plans, which reward using expensive and sometimes unnecessary treatments and treating large amounts of patients quickly, to plans that prioritise patient outcomes. In the U.S, healthcare providers had no direct incentive to share patient information with one another, which had made it harder to utilise the power of analytics. However, nowadays more patients are getting paid based on patient outcomes, making the financial incentive to share data which can be used to improve the lives of patients while cutting costs for insurance companies. Finally, physician decisions are becoming more evidence-based, meaning that they rely on large swathes of research and clinical data as opposed to solely their schooling and professional opinion.

Today, there are many types of health information, including information about health conditions in the hospital such as medical history, medication intake, results from blood tests, urine tests, data from CT-Scan, MRI or X-ray images, data from wearable devices such as smartwatches such as apple watches that can measure EKG, preliminary blood oxygen level, etc. (2). Handling information from many sources can be done in many forms. Currently, in a healthcare facility,

there is a Health Information System (HIS) such as Electronic Medical Record (EMR); however, it does not support the exchange of information between hospitals on the necessary information such as treatment history. The development of public health systems, for instance,the National Health System (NHS) in the United Kingdom has revolutionised the way data is collected. Originally data was stored in the database of each hospital. However, today, data is collected in the central cloud-based storage, using cloud computing technology. (3).

Cloud computing technology together with the increasing internet speed makes it possible to upload data to the database anytime, anywhere, generating huge amounts of data at a very fast speed. Information overload has turned into Big Data. This information has been put to good use by applying the basics of statistics, information, computers, communication and management to synthesize new information. The information is born into a science called. Data Science or Data Science (4).

Data analytic in Healthcare gives insightful information supporting efficient diagnoses, treatment and patient care. This research examines a case study of how data science can be used in healthcare to help diagnose, monitor and treat patients better. In addition to helping to increase the efficiency of management

What is Big Data for Health Care?

As the volume of data is significantly increasing due to technology advancements such as internet speed is much faster than ever before make a way for gathering informations from smartphones, tablets, smartwatches, wearable devices, CCTV or other IoT appliances which can generate and upload huge amounts of data at the same time whether it is in the form of still images, video, text, sound and other forms of information. The development of storage systems such as Cloud allows data connectivity to be done anywhere, resulting in a rapid increase in data volumes. The data that is created, stored, copied and used around the world is increasing rapidly. By 2020, 64.2 zettabytes of data are being added to a new high. This data will continue to increase to more than 180 zettabytes by 2025 (5).

With some reports estimating that healthcare data increases by 48 per cent annually which can be problematic to do the manual analysis. Moreover, it is possible for the data to exceed the capacity of conventional databases to store, categorise, and analyse these growing data sets. According to the increases in volume, healthcare data are highly complex due to the presence of many data standards and an estimated 80% of information being unstructured (6). Hence, computers are involved in the storage, analysis, and prediction section. Data science is a fundamental principle that supports and guides the principled extraction of information and knowledge from data. Besides, the statistical field is one of the fundamentals of data science because it has to use vast amounts of data (7). In addition, Big Data technologies and techniques can help researchers to understand a plethora of rich and diverse health data and benefit the research and development sector.Big Data can be described as a bridge that brings precision medicine and public health together, allowing researchers to study the relationship between categorised data (8).

Where is the data coming from?

Health data can be collected by either wearable devices or traditional methods which are collected face to face. In this era, people are living in the age of mobility. Many technologies involve collecting data such as phones, tablets, notebook computers, smartwatches, etc. Most wearable devices use sensors and signal processing to collect health data (9). Wearable sensors are used to monitor patients' health and involve precision medicine (10). Cloud storage, an off-site location, can help researchers, doctors, hospitals, insurance companies, the ministry of health, and other health organisations access the patients' data (11). Cloud storage - such as Google Drive, DropBox, and iCloud - enables people to upload data through the Internet to cloud-based servers. Thus, anyone can access the data from multiple devices by using the Internet as the bridge (12).

Big data Application for Healthcare

Big Data can be applied in medical and public health. Everything starts from collecting basic patient data from the patient's history, patient symptoms, results from laboratory tests, diagnostic results, to methods of care in the convalescent and long-term periods. Sandra Durcevic (2020) claimed that big data can be applied in Healthcare as follows:

1) **To alert about the epidemic situation**: to remind patients to come for treatment on time, to arrange appropriate personnel to close-reduce weaknesses-deficiencies of the public health system for hospital budget management services

2) **To recommend the most suitable hospital for patients:** by processing Big Data from linked electronic medical record systems across the country.

3) **Real-Time Alerting:** data from wearable devices are analyzed and processed to predict future trends, to formulate appropriate treatment guidelines. Big Data also allows physicians to monitor patients' symptoms anywhere, anytime, in a timely manner in order to increase medical efficiency and reduce errors in the diagnosis process.

4) **Enhancing Patient Engagement**: people nowadays prefer to keep track of their health information all the time which may help them find hidden health risks. For chronic patients when they are using wearable devices, doctors can monitor results at all times and could reduce hospital visits.

5) **Prevent Opioid Abuse In The US:** by using information about drug reception and insurance information to predict if the drug is at risk of addiction so that it can be prevented according to the information.

6) **Using Health Data For Informed Strategic Planning:** relevant personnel can analyse demographic data in each location to gain insights into the motivations of each population such as the reasons why people choose not to treat.

7) **Data Analytics for curing Cancer**: researchers use the data on the treatment to analyse the success rate of that treatment and find a way that has the highest chance of success.

8) **Predictive Analytics In Healthcare:** with decision support data, physicians can make informed decisions in a short time. This can be very helpful in cases where patients have a medical history and complex symptoms.

9) **Reduce Fraud And Enhance Security:** 93% of public health organisations have experienced data breaches, data analytics can address this problem by alerting them when network traffic changes abnormally or when there are other potential cyber-attack indicators. Anti-corruption devils such as Misappropriation or false insurance claims can also speed up the insurance claim process.

10) **Telemedicine:** with the advancement of technology such as video conferences, smartphones, wireless, wearable devices, telemedicine is growing rapidly. Telemedicine services can be provided, especially with basic services such as initial consultation or initial diagnosis, or even telesurgery, which is a remote surgery with robots where the surgeon and the patient are in different places. In addition, telemedicine gives doctors more personalised treatment. With analytics, doctors can predict potential future crises and find ways to prevent them.

11) **Integrating Big-Style Data With Medical Imaging:** processing a medical image takes a lot of time and money because radiologists have to manipulate the results and hospitals still store those medical images for a long time. Hundreds of thousands of images can be identified in pixel format and numerically converted to help physicians make a diagnosis. It was forecasted that in the future, radiologists may not need to read the images instead of analysing the results from the algorithm.

12) **A Way To Prevent Unnecessary ER Visits:** the problem of unnecessarily repurposing the ER is caused by the fact that each hospital's emergency room doesn't share information. Hospitals in Alameda county have come up with a solution to this problem. By programming PreManage ED, patient history is shared between emergency rooms.

13) **Smart Staffing & Personnel Management:** effective personnel management can be done by collecting statistical data of each department in each period of time which is later used for data visualisation. This may be done in a form of HR analytics in order to be able to allocate the appropriate number of personnel. It also allows executives to reach out to help appropriately and also allows them to plan various strategies as well.

14) **Learning & Development data analytics:** a performance of each staff member can be analysed in order to gain insights about what needs to be improved or is a bug

15) **Advanced Risk & Disease Management:** hospital management can be done more efficiently utilizing correlation analysis between the nature of transmission and the hospital environment

16) **Suicide & Self-Harm Prevention** data analytics can predict a person's tendency to commit suicide with such an information available it can be preventable.

17) **Improved Supply Chain Management:** big data analytic with leverage analytics tools can help supply chain more fluidly and efficiently. Leverage analytics tools can be used to track supply chain performance to make informed decisions. Descriptive and predictive analytics models can also help in negotiating prices, reducing variables in supplies and improving the overall procurement process which leads to the prevention of shortages and delays.

18) **Developing New Therapies & Innovations:** data analytics assists research and development process which may lead to the discovery of new treatments and new drugs much faster than ever before.

Implementing Big data for healthcare

Applying Big Data to medical, public health, everything starts with collecting basic patient data from the patient's history, patient symptoms, results from laboratory tests, diagnostic results, and methods of convalescence and long-term care may be collected in the form of an Electronic Medical Record or EHR, for example in the United States, or from wearable devices including information from insurance companies. This has been analysed and processed for many benefits in both the public health system and at the level of the hospital itself, in both macro and micro aspects.

Health data system in Thailand

In addition, Thailand healthcare providers adopted a wide range of hospital software (13) creating difficulty for sharing of information. Interoperability of health data is a crucial feature and only some healthcare providers had already implemented this; such as Fast Healthcare Interoperability Resources (FHIR) (14).

In Thailand, several health data systems had been deployed such as 1) Hospital Information System (HIS), the HIS holds and maintains all of the information of institution's health records, medical records, personal health records, including personnel, costs, strategies, and management information (15). Different healthcare providers use different software. HOSXP is one of the most adopted systems in government healthcare facilities; other software is EPHIS, EMR Soft, SSB and etc. (13). 2) 43 Files (HDC), this is a standard formation of a medical database system that collects some parts of the medical records in every medical institution in Thailand for studying the population's overall health, as well as planning the national healthcare strategy and policy. (16) 3) ICD-10, The International Classification of Diseases, 10th Revision (ICD-10) has been used together with the HDC for classifying and grouping diseases. (17)

In Thailand, medical records were first assembled in the level of health facilities using a Hospital Information System (HIS) then some medical records from every medical institution would later be collected into a central database called a Health Data Center (HDC), which works as a database for research and planning national strategies (16).

Challenges of data science Implementation in Thailand

This paper summarises key challenges in applying data science in Thailand 1) Health data has been kept in both digital and non-digital formats. Thongthai et al. (2017) found that only 28.5% of all dental and medical clinics adopted EHR, and 25.4% of which only used EHR for collecting patients' personal information such as, telephone number, name, age, and gender (18). Without fully transforming health data into electronic forms, utilisation of health data hardly occurs. 2) Data Integrity, quality of data in both correctness and being recorded in a timely manner. It was found that data entering consumed time due to the complexity of the platform used, and the instability of the Internet connection (19). Furthermore, many physicians used abbreviations, which could be interpreted into many definitions. Hence, the data were returned for clarification and consumed more time (20). 3) Human resource issue, there was insufficient staff who are specialised in working with data at different levels such as data entering (20). Despite a fast-growing demand for data scientists globally, specifically in health data scientists, there is a serious shortage of data scientists worldwide (21).

2. CONCLUSION

Big Data and Data Science application in healthcare mainly to increase effectiveness and efficiency of diagnosing, treatment, taking care of patients, healthcare service & operation, prediction and prevention of illness. The application of data science in healthcare required comprehensive and structured data in a digital format with sophisticated data analytical systems. Applying big data and data science in healthcare requires the availability of many matters in different dimensions, besides policy and budget, collecting health data, data integrity, storing it in digital form, availability of hardware and software, trained staff, responsible persons, data workers, data sharing, connectivity, utilisation of data analytics and availability of health data scientist.

REFERENCES

[1] Durcevic S. 18 Examples Of Big Data Analytics In Healthcare That Can Save People [Internet]. 2020 [cited 2021 Sep 27]. Available from: https://www.datapine.com/blog/big-data-examples-in-healthcare/

[2] Apple. Apple Watch Series 6 [Internet]. [cited 2021 Sep 24]. Available from: https://www.apple.com/th/apple-watch-series-6/

[3] Phanhong M. Digital Transformation in Health Services [Internet]. 2021 [cited 2021 Oct 25]. Available from: https://bigdata.go.th/movements/digital-healthcare/

[4] Kullimratchai P. Entering into the Era of Data Science. EAU Heritage Journal Science and Technology. 2018;12(2):120–9.

[5] Holst A. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025 [Internet]. 2021 [cited 2021 Sep 26]. Available from: https://www.statista.com/statistics/871513/worldwide-data-created/

[6] McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, et al. Health care and precision medicine research: analysis of a scalable data science platform. J Med Internet Res. 2019 Apr 9;21(4):e13043.

[7] Provost F, Fawcett T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data. 2013 Mar;1(1):51–9.

[8] Velmovitsky PE, Bevilacqua T, Alencar P, Cowan D, Morita PP. Convergence of precision medicine and public health into precision public health: toward a big data perspective. Front Public Health. 2021 Apr 6;9:561873.

[9] Edwards J. Medicine on the Move: Wearable devices supply health-care providers with the data and insights necessary to diagnose medical issues and create optimal treatment plans [Special Reports]. IEEE Signal Process Mag. 2019 Nov;36(6):8–11.

[10] Vijayan V, Connolly JP, Condell J, McKelvey N, Gardiner P. Review of wearable devices and data collection considerations for connected health. Sensors. 2021 Aug 19;21(16).

[11] Rui L, Liang Z, Xiaming Z. Health data processing method and system based on terminal equipment and cloud computation storage. Europäisches Patentamt. 2015;

[12] Cloud Carib. Cloud Storage vs. Cloud Computing: What's the Difference? [Internet]. [cited 2021 Sep 25]. Available from: https://info.cloudcarib.com/blog/cloud-storage-vs.-cloud-computing-whats-the-difference

[13] Panyowat R. editor Hospital information system (HIS) in Thailand [Internet]. 2019 [cited 2021 Sep 28]. Available from: https://rath.asia/2019/11/his-vendors-thailand/

[14] Adler-Milstein J, Pfeifer E. Information blocking: Is it occurring and what policy strategies can address it. Milbank Quarterly 2017;95(1):117-135. https://doi.org/10.1111/1468-0009.12247

[15] Mehdipour Y, Zerehkafi H. Hospital Information System (HIS):At a Glance. Asian Journal of Computer and Information Systems. 2013 Aug;1(2):54–61.

[16] Yoheung, D, editors. Practice guide. Storage and delivery of data in accordance with the health information Standards structure version 2.4. 2021. http://spd.moph.go.th/healthdata/wpcontent/uploads/2021/03/2.4- 43File_ Structure-020364.pdf

[17] Intelligence and Emergency Response Group, Bureau of Occupational and Environmental Diseases, Department of Disease Control Ministry of Public Health. editors. ICD-10 coding for occupational and environmental diseases. 2016.

[18] Thongthai K, Kunawaradisai N, Watcharathanakij S. Adoption of electric health records in medical clinics and dental clinics. Thai Journal of Pharmacy Practice. 2017;9(2):490–8.

[19] Wipak R, Turnbull N, Siwina S. Factors related to data quality management for 43 public health data. The Southern College Network Journal of Nursing and Public Health 2017;4[Special Issue]:162-179.

[20] Habusaya S,Ditcharoen N. Analysis of ICD-10 coding errors in 43 files database systems for medical record department using data mining techniques. KKU Science Journal, 2020;8(1):141- 155. http://203.157.196.7/web_ssj/ webmanag er/uploads/2018-03- 2614320043_Data_Analytic.pdf

[21] Zita C. Is Data Science Still a Rising Career in 2021 [Internet]. 2021 [cited 2021 Sep 25]. Available from: https://towardsdatascience.com/is-data-science-still-a-rising-career-in-2021-722281f7074c#:~:text=So%20is%20 data%20science%20still,option%20for%20a%20career%20path.