

## Big Data Analytics for Health Care: A Review

Mohammed Iqbal Fathima Nihla<sup>1</sup> and Abdul Raheem Fathima Shafana<sup>2</sup>

<sup>1</sup>Faculty of Information Technology  
University of Moratuwa  
Sri Lanka

<sup>2</sup>Department of Information and Communication Technology  
Faculty of Technology  
South Eastern University of Sri Lanka  
mifnihla@gmail.com, arfshafana@seu.ac.lk

**Abstract.** Sequences of revolutions and disruptions in medical learning and health care coupled with Information Technology have motivated and triggered the usage of big data which in turn has impacted health care too as many other fields available today. The word big data has become a slogan in the recent past, through its growing usage every year. Big data applications play a major role in health. Big data in health sector are alarmed with a huge amount of data sets that are too large, too complex, and too fast to process, manage and analyze. The proper management and critical analysis of these data sets can be applied into health care. This would give a very optimal result in wellness, personalized health, cost efficiency and predictive analysis. But the challenges in handling big data and the poor knowledge in using that would affect the applications of it into health care sector. The purpose of this study is to gain insight on big data and its applications in health care and to analyze the roles and applications of big data for the improvements in health care.

**Keywords:** Bioinformatics; Medical informatics; Bio curation; Personalized health; Omics data

### 1 Introduction

Big data is a buzzword today in many fields. Big data can be simply referred to as a huge amount of data which is totally beyond the processing capability of traditional database management system. Unlike the traditional system, it is hard to process, manage and analyze data in a specified time. The last decade has seen huge advances in the amount of data we routinely generate and collect in pretty much everything we do, as well as our ability to use technology to analyze and understand it" [1].

Big data is guiding towards the betterment of this world and it is widely used in many fields and sectors like health care, financial services, marketing, and telecommunication etc. Among all the fields stated above, application of big data in health care sector is highly beneficial.

There is a huge number of scenarios in health care that are well suited for solutions through big data. Big data in biomedicine is driven by the single premise of one day having personalized medicine programs that will significantly improve patient care. Constant advances in understanding of different omics information are providing the footholds into

establishing, for the first time the casual genetic factors, that could help manage the golden triangle of treatment; the right target, the right chemistry and the right patient.

However, challenges ahead funneling clinical data, omics data, administrative and also financial information securely into a unified system to achieve better patient outcomes, advance research and continually improve the quality of patient care while reducing costs"[2]. Characteristics of big data is defined the three V's such as Volume, Velocity and Variety. Along with these three there are other characteristics like variability, veracity, and value (worth). Veracity- trustworthiness of the data is considered is one of the most significant things for big data in health sector. One example of maintenance of personal health records is the maintenance of records free of errors and cryptic notes.

The worthiness of big data in healthcare can be realized only if the challenges in big data applications are considered properly. Therefore, if used in a coherent fashion, big data can be a valued source that can be a significant one towards enlightening current health services and dropping health care charges.

Organization of the paper is as follows: Section 2 provides the overview of the applications of big data in health care which includes a small description on bio curation. Section 3 is on the prevailing activities and future opportunities related to big data for health care sector and it includes various informatics related to the study. Section 4 presents how these data can be analyzed and applied, and Section 5 analyzes as challenges faced. Section 6 is about the Sri Lanka's situation about the study being discussed and finally the Discussion and conclusion of the research is presented in section 7.

## **2 Overview of applications of big data in health care**

In the research report published by Matthew Herland et. al., it is said that "The amount of data produced within Health Informatics has grown to be quite vast, and analysis of this Big Data grants potentially limitless possibilities for knowledge to be gained. In addition, this information can improve the quality of health care offered to patients."[3].

Health Informatics can be defined as the integration of information science and computer science within the realm of healthcare. There are numerous current areas of research within the field of Health Informatics that includes Bioinformatics, Image Informatics (e.g. Neuro informatics), Clinical Informatics, Public Health Informatics, and Translational Bio Informatics (TBI). [3].

As stated before, any company in an industry can get advantage from big data applications regardless of its size and location, but when analyzing thoroughly, it is understood that it is the health care that is benefited the most. Each time hospitals produce a lot of data for patient care and things related to it and on things like drug administration and so on, and all that data can be used to mend the worth of care and operations. The vast amount of data produced is one of the best pools of knowledge which supports the health care professional greatly.

Therefore, by analyzing the amount of data that is produced in health care and studying about various types of health informatics, the need of big data is felt and its applications in health care have turned out to be much essential.

Bio curation is the field that connects biologists with their need for data, identification, and support. The exponential growth in the amount of biological data has demanded the necessity of revolutionary measures for successful data management, analysis, and accessibility. Further, online databases have gained much importance since a strong need was felt to identify the suitable avenues for publishing biological data. In simple terms, the activity of organizing, representing and making biological information accessible to both humans and

computers can be referred as Bio Curation and this field has become much promising and an essential part of biological discovery and biomedical research [4].

### **3 Existing opportunities and future activities related to big data in health care**

#### **3.1 Medical and health informatics**

There is a great opportunity to integrate old-fashioned medical informatics with mobile health and social health data when dealing with the huge capacity of data.

##### **Electronic Health records (EHRs)**

Traditionally health data cores capture and store a massive volume of data regarding diagnostics of diseases, laboratory test results, medication, and clinical data records. Researching on EHRs plays an important role in maintaining clinical knowledge of patients and supporting clinical research in academic institutions. Although when analyzing through these, most of the clinical databases provide chronological resolution information due to the difficulty in gathering long term data. Therefore, to overcome this cavity, clinical databases are linked by mobile health platforms so that information can be fed into the system to enhance and improve clinical decision making.

“Large-scale population screening for this rare disease will, therefore, be useful in identifying people who are at higher risk of developing aortic dissection.” [5]. Further, using clinical databases, acute diseases can be predicted at an early age.

##### **Social Health**

Clinical patients and doctors are connected through telemedicine. This new feature gives possibilities to patient to patient-based communication beyond traditional doctor to patient-based paradigm. In this scenario, the patients tend to share their experiences with others over social media which paves way to potential use of big data applications.

##### **Personalized Health**

Traditionally epidemiology or biology alone is not sufficient to address the disorders and prevention mechanism. Therefore, a well-recognized approach must be considered to get understandings of the biological systems. Also, the access to large capacity of omics data have paved the way to rise of personalized health. And this has led the way to development of ‘systems biology’ as an area of research where the complex diseases can be modelled. Personalized health can be achieved through systems biology or systems medicine, which is a biological mechanism.

“Systems medicine, the application of systems biology to human diseases, requires investments in infrastructures with cutting edge omics facilities and analytical tools, advanced digital technologies (high computing performance and storage resources) and highly qualified multi – disciplinary teams (clinicians, epidemiologists, computer scientists, statisticians and mathematicians) in addition to investments in security and privacy”[6]. The researcher has

identified that system biology is the tool to address the gap between socio economic and scientific evolution towards personalized medicine.

### **3.2 Translational Bioinformatics**

Translational bioinformatics is one field that developed because of human genome mapping. This field focuses on to bridge molecular biology, biostatistics, and statistical genetics with the clinical informatics. The field is evolving at a rapid pace, in turn, has proposed certain other related areas [5].

In another research, translational bioinformatics has been identified as one of the most trending areas in genomics that is concerned with individuals showing different responses to drug due to genetic differences. Human health and healthcare can be easily revolutionized by the translational informatics by means of large-scale measurements on individuals. The researcher also believed that the data-centric approaches that compute on enormous amount of data (also known as “Big Data”) in order to discover new patterns and to make clinically relevant predictions will gain an important adoption [7]. Gathering the huge amount of data, analyzing the patients’ medical and functional data, sequence variations in genomic information can be predicted in advance. This pattern mining technique is believed to be supported with an added advantage in the field of health care.

### **3.3 Sensor Informatics**

“Enabling technologies ranging from Nano and microelectronics, advanced materials, wearable/mobile computing, and telecommunication systems, as well as remote sensing and geographic information systems have made it possible for sensing health information to be collected pervasively and unobtrusively” [5].

These sensors can be found as wearables, implantable and ambient sensors. These sensors are fed with high data processing power, higher bandwidth which enables faster wireless communications, and improved designs. “Example platforms include earlier systems with limited connectivity and single sensing elements developed solely for use in research laboratories to more recent ambient sensors as well as easy-to-wear wearable/implantable devices equipped with continuous multi modal sensing capabilities and support for data fusion deployed in a wide range of clinical applications” [8]–[10].

Challenges of both acute and chronic disease monitoring can be done by implantable sensors which facilitate with means of capturing adverse and critical information of health-related data and information.

Mobile health is another concept that comes under this area. Smart phones are at a high usage now days. They have become inseparable with everyone. Also, when considering about the data generating capability of smart phones, they are at a high speed and very much descriptive. Therefore “They can potentially serve as a platform to centralize health data, from which additional new information that was previously untraceable by individual sensors can now be mined.”[5].

### **3.4 Neuro informatics**

Neuro informatics is a young field that concerns about analysis about the brain image data to understand the working patterns of the brain and how it correlates between information gained

from brain. The field of Neuro informatics is widely chosen to represent the broader domain of Medical Image Informatics due to the reason that, by limiting the scope to brain images, more in-depth researches may be undertaken while still gathering enough information to constitute Big Data [3].

#### **4 Data analysis and Application**

Using various levels of data (molecular level data, tissue level data and patient level data) the applications of big data is possible. If molecular level data is taken into consideration it uses the data collected at a molecular level. That means, “The studies in this subsection use gene expression data to answer clinical questions. Two research efforts are reviewed in this subsection, both of which focus on cancer: the first uses gene expression profiling to categorize leukemia into two different subclasses, while the second study uses gene expression data to predict relapse among patients in the early stages of colorectal cancer (CRC). Both studies (as well as similar studies) can help physicians guide, advice and treat their cancer patients.”[3]. Forming the connectivity plan of the brain is discussed in tissue level data gathering. Mapping brain images will help to understand how human brain works and there after leading to many predictions of human brain’s activity. Use of MRI data can be a better way to analyze brain images.

Clinical data is best when taking about the patient level data analyzing. Mortality rate predictions are done by use of patient level data. “This is a useful line of research in that it can potentially help physicians know what to look for in their patients, determine which patients should have their ICU stay extended, and better tell which patients should receive particular treatments.”[3], “Instead of predicting the patient’s condition in the future (i.e. ICU readmission or 5 year survival), the research here will be using data streams in order to predict patient’s conditions in real-time. Data streams are never ending torrents of data that requires continuous analysis giving the possibility for real-time results (a feature not available when using static data sets).”[3].

#### **5 Challenges Ahead**

The remarking changes in big data have created issues in storage, transformation and security in data and information. These days generating data can be less expensive than securing it and analyzing it. “In addition, biological and medical data are more heterogeneous than information from any other research field.”[2]. Therefore, this imposes a great challenge for the future of bio medical research.

Transfer of data from one place to another is the other challenge encountered. Another challenge about this is that the computer organization and software implements must be up to the mission that is being done. “They wanted to compare their data with the thousands of other published breast cancer – genomes and look for similar patterns in the scores of different cancer types. But that is a tall order: downloading the data is time consuming,” [11].

The security and privacy of data is also considered as a challenge in big data. As data is generated at increasingly high speed with various data capturing devices these days, the decision must be made to available always in real time to face the continuous evolution of technologies.

## **6 The view on Sri Lanka's context connecting to the study conducted**

This sub part of the paper is going to discuss mainly about how the study conducted is viewed at Sri Lanka's context. Sri Lanka being one of developing countries, in the World and in South Asia, to provide free health care solution to its public, has been a center for rich and well facilitated medicine since the ancient time. Time to time there had been various health care habits conducted in Sri Lanka, as an example, Ayurveda – the indigenous medical tradition of Sri Lankans can be considered. Providing this background information, it can be said that the country is now heading towards various technological transformations and studies to adopt the high-tech changes to its various sectors ranging from micro to macro level.

Under this phenomenon, the health care sector of Sri Lanka is also becoming revamped towards data engineering; making a way to approach and analyze bigdata with the huge amount of health data that is open, clear, and accessible. Data Engineering promotes an engineering approach to analyze big “imperfect” data by creating and using new smart algorithms such as Unsupervised Deep Learning, appropriate electronic hardware platforms and mechanical/mechatronic/material/ chemical-based approaches to interrogate and acquire missing data or information [12].

In the research paper submitted by GGT Dantanarayana et al, ‘Quality of Information for Quality of Life: Healthcare Big Data Analytics’, the paper emphasize on Sri Lankan hospitals of both government and private sector where it proposes a methodology to use Electronic Health Records (EHRs) as a way of providing a better and quality healthcare to patients by analyzing the quality information such records carry within them during, the bigdata analytics for healthcare. Also, the study is highly motivated on identification of data sources, point of information extractions for clinical and healthcare decision-making and analysis of Information Exchange measures using predictive, descriptive, and prescriptive analytics. The research methods aim to implement six to eight months' longitudinal studies from selected private hospitals and government hospitals in Sri Lanka when identifying healthcare information sources and its transmission [13].

The research also suggests having clinical and EHRs to assist the clinical decision making using the analytics applied. From this study which is based on Sri Lankan government and private hospitals it is well understood that, the Electronic Health Records play a major role and are highly considered as a key information resource when it comes to big data analysis.

A systematic approach containing comprehensive guidelines with the correct data must be provided to apply and evaluate big data analysis until the necessary decision-making requirements are fulfilled to improve quality of healthcare service delivery. In this research, the healthcare information sources, and their transmission will be examined while quality Healthcare data analytics is implemented [13].

On another study conducted by Chathurika Wickramage et al. ‘Information Accountability of Healthcare Big Data’, it has been proposed about the implementation of information accountability during identification of data sources and information extractions points which could be considered as a practical solution when dealing with the complexity and multi-varied dimensions of healthcare bigdata. The study discusses a methodology to be executed during information sharing in an e-Health scenario where both the public and professional view are taken into consideration when it comes to information flow. The professional view corresponds to the medical professionals who are involved in providing healthcare service while the public view corresponds to the customer (patient) who acquire the service [14].

The motive of this research is to identify the procedures and measures that are used to recognize the accountability of information and accurate information extractions using the Electronic Health Records (EHRs) for better large-scale processing analyses of data.

Accordingly, it can be applied to healthcare big data for clinical and healthcare decision making. Correspondingly, the study indicates the proper authorization of data as an important feature to the research conducted based on Sri Lankan hospitals.

In another research conducted, titled 'How Sri Lanka can leverage genome sequencing and editing technologies to become a global leader in the precision medicine initiative' by Prof. Preethi Gunaratne, the study is aimed at presenting a framework for Sri Lanka to integrate genome sequences of every citizen into the electronic health records or electronic medical records (EHRs/EMRs) by creating a database. By doing so will make the country to emerge as the first nation to address such a problem with the use of emerging technologies such as big data.

The value proposition is to create a powerful database for solving the genetic basis of diseases common to Sri Lanka like diabetes, kidney disease, and cardiovascular disease. The infrastructure established can be leveraged to create two additional value propositions 1) Become a global leader in health tourism—leveraging Sri Lanka's reputation as a top tourist destination with a strong national health care system. 2) Building on a long, strong tradition of classical ayurvedic medicine, to integrate natural product derivatives from plants indigenous to Sri Lanka into next-generation drug discovery platforms [15].

Therefore, it is clear that the studies based Sri Lanka have focused on incorporating bigdata to healthcare due to the usage of electronic health records where the health communities have felt those types of data sources are more complex and are of huge amount of data content in them. Also, in view of Sri Lanka, adopting bigdata solution to healthcare domain is considered more challenging than other sectors because of the fact that , this sector is consisted of data from numerous sources. And one important thing to be highlighted from the studies is that, compared to bigdata analysis for business sectors, the healthcare sector is still in its early stages.

## **7 Discussion and conclusion**

Throughout this paper, the big data applications in health sector has been discussed. When thinking in the view of applications of big data in health field, different informatics related to health are taken into consideration and by analyzing through various levels of data the applications can be done in health care sector.

With the rapidly changing health care environment there should be applications that can generate clinical trials data analysis, disease pattern analysis, patient care, and quality analysis and so on. So big data has become a tool technology, which is very much helpful in meeting the above challenges faced by the health care industry.

The field of health care is rapidly evolving, and it produces millions of data in every other second. For example, Electronic Medical Records (EMRs) alone could be used to collect a huge amount of data. Considering this, it can be said that there is certainly a vast amount of data that is collected every time in this sector which needs to be processed and managed within a specified time span. Therefore, big data has a potential in this promising field to closely monitor all the quintillion amount of data collected.

The current system of disease centric health care system is not much workable. The cost related with treatments of this type of health care is rising continuously in huge amounts. Therefore, integration of data from whole health care and single person are used to provide results to coach with the aim of optimizing health and reducing disease. The concept of P4 medicine, where P4s stands for Predictive, Preventive, Personalized and Participatory respectively, bridges the gap between prevailing system approaches to diseases through applications of big data.

Combination of the various types of data can provide more complete assessment and ultimately more reckoned metrics of wellness. With the available genetic data or laboratory data, the individual can get expert advises from health coaches. With these integrated data types, recommendations can be provided to individuals. Further, the early transitions to diseases can also be identified.

This paper has discussed about bio curation which has its main role as connecting information from different sources and extracting knowledge from published papers. As health sector moves towards a more effective route, bio curation is recognized an important sector of research and education in terms of technology. And it is obvious from the study that bio curation and bio curator are the cores to the future of health care field.

To put the findings in a nut shell, it can be concluded that the use and applications of big data in health care is an approach which is more effective, economical, innovative way to optimize wellness and it still can delay or prevent a disease at a relatively earlier stage.

## References

- [1] Bernard Marr, *How Big Data is Changing Healthcare*, Enterprise Tech, Apr 21, 2015. Accessed on: June 13, 2020. [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#475a02da2873>
- [2] F. F. Costa, "Big data in biomedicine," *Drug Discov. Today*, vol. 19, no. 4, pp. 433–440, 2014.
- [3] M. Herland, T. M. Khoshgoftaar, and R. Wald, "Survey Paper, Open Access. A review of data mining using big data in health informatics," *J. Big Data*, vol. 1, no. 2, pp. 1–35, 2014.
- [4] D. Howe *et al.*, "Big data: The future of biocuration," *Nature*, vol. 455, no. 7209, pp. 47–50, 2008.
- [5] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Z. Yang, "Big Data for Health," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 4, pp. 1193–1208, 2015.
- [6] A. Alyass, M. Turcotte, and D. Meyre, "From big data analysis to personalized medicine for all: Challenges and opportunities," *BMC Med. Genomics*, vol. 8, no. 1, 2015.
- [7] N. H. Shah, "Translational bioinformatics embraces big data," *Yearb. Med. Inform.*, vol. 7, pp. 130–134, 2012.
- [8] Y. L. Zheng *et al.*, "Unobtrusive sensing and wearable devices for health informatics," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1538–1554, 2014.
- [9] Y. L. Zheng, B. P. Yan, Y. T. Zhang, and C. C. Y. Poon, "An armband wearable device for overnight and cuff-less blood pressure measurement," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2179–2186, 2014.
- [10] C. C. Y. Poon, B. P. L. Lo, M. R. Yuce, A. Alomainy, and Y. Hao, "Body Sensor Networks: In the Era of Big Data and beyond," *IEEE Rev. Biomed. Eng.*, vol. 8, no. September, pp. 4–16, 2015.
- [11] V. Marx, "The big challenges of big data," *Nature*, vol. 498, no. 7453, pp. 255–260, 2013.
- [12] S. K. Halgamuge, "Keynote address: Data engineering meets health," *2016 Moratuwa Eng. Res. Conf.*, pp. xvi–xvi, 2016.
- [13] G. G. T. Dantanarayana, T. Sahama, and G. N. Wikramanayake, "Quality of information for quality of life: Healthcare big data analytics," *15th Int. Conf. Adv. ICT Emerg. Reg. ICTer 2015 - Conf. Proc.*, no. November 2016, p. 281, 2016.
- [14] C. Wickramage, T. Sahama, and G. N. Wickramanayake, "Information Accountability of Healthcare Big Data," *15th Int. Conf. Adv. ICT Emerg. Reg. ICTer 2015 - Conf. Proc.*, no. September 2016, p. 279, 2016.
- [15] N. Pokuna, M. Rajapaksa, and W. Edge, "Science and Technology for Society Forum Sri Lanka 2016."