



# Data Science for Healthcare: A Brighter Clinical Future

Elnaz Amanzadeh Jajin <sup>1,\*</sup>

<sup>1</sup>Cellular and Molecular Biology/Microbiology, Faculty of Biological Sciences and Technologies, University of Isfahan, Isfahan, Iran

\*Corresponding author: Cellular and Molecular Biology/Microbiology, Faculty of Biological Sciences and Technologies, University of Isfahan Isfahan, Iran. Email: [elnaz\\_amanzadeh@sci.ui.ac.ir](mailto:elnaz_amanzadeh@sci.ui.ac.ir)

**Received** 2021 April 07; **Revised** 2021 April 30; **Accepted** 2021 May 09.

**Keywords:** Neurological Diseases, Big Data, Precision Medicine, Data Science, Electronic Health Record

The last decade is known for the exploitation of clinical data and huge advancements in computational methods, which resulted in the introduction of large amounts of data with detailed information at the molecular level. These data can be used by health systems to increase the accuracy of diagnostic techniques as well as the efficiency of treatments. Clinical data, especially those collected by Electronic Health Records (EHR), have turned into a new approach for recording patients' health information (1). Computer science has provided a great opportunity to record detailed data of patients. EHRs include data on, for instance, demographic characteristics, medical images, medical signals, genomics data, and pharmaceutical data. EHRs allow hospitals 24/7 access to the clinical history of patients. In addition, the information provided through EHRs allows physicians to make better diagnostic and curative decisions. Moreover, access to such information paves the way for achieving a better follow-up process and to reduce healthcare costs (2). EHRs also have advantages regarding health research and hypothesis to hospitals where new treatments are required and applied every day. Apart from the significant benefits of the collection of clinical data, it seems that storing and analysis of these data is becoming a challenge. The first challenge is related to data storage. Access to patients' data and accurate data analysis require the cooperation of several departments. Data analytics, big data management, programming, classification and prediction algorithms, statistics, web development, computer science, etc., are involved in data science for healthcare. Data scientists have developed infrastructures that are required to store high throughput data, known as cloud computing. Through cloud computing, partners working on a similar project can have access to all necessary data. Authorities have different responsibilities, including operators for data entry, researchers for data analysis, and programmers to update web-based platforms.

This approach also makes it possible to review databases and update them remotely (3). Accordingly, the analytical process, as the second challenge, has been improved since clouds are usually connected to super-computers. In healthcare, big data is heterogeneous and multi-spectral, because such data are collected through various resources, including diagnosis, treatment, EHRs, etc. (4). Therefore, several methods should be used to integrate data at pre- or post-analysis levels and to analyze such data. Data science provides the possibility to gather healthcare data and analyze them to return understandable outputs for healthcare staff. The importance of data analytics is highlighted when there are data (both structured and unstructured) that should be linked or should be analyzed and interpreted to provide understandable reports for physicians. Currently, with advancements in OMICS studies, in different dimensions, including genomics, proteomics, and metabolomics, huge amounts of data have been collected (5).

To analysis OMICS data, integrated packages and pipelines have been developed, which can be applied in remote super-computers. The achieved results can be used for point mutation detections, gene enrichment analysis, protein interactions, and pathway enrichment analysis studies. In addition, analysis of genomics data with the aim of detecting new markers for complex diseases, including cancers, neurodegenerative diseases, and atherosclerosis diseases, has been widely used by biological researchers. The output of these analyses can be used to design novel and more effective medications. In the case of clinical images like magnetic resonance imaging (MRI), ultrasound, and computed topography (TC), enhancement, segmentation, and denoising should be applied. Processing clinical signal data, for example, integration of huge data collected in intensive care units (ICU) with a large amount of pathophysiological data, requires a robust in-

frastructure. Text mining studies using available documents, including academic researches, are considered as a source for improvement of treatment methods.

Another application of healthcare data science is to predict future steps in diagnosis and treatment of various diseases (6). As healthcare data is devaluated and predictions lose their validity after a certain time. Hence, data scientists should have access to updated data. On the other hand, with increase in size, dimensionality, and complexity of data, analysis becomes more complicated. Machine learning (ML) and artificial intelligence (AI) are very important approaches for big data analysis and prediction. The integration of big data analysis and healthcare system has led to the appearance of a new concept called clinical decision support system (CDSS) (7). ML and AI play key roles in the development of CDSS. CDSS uses ML approaches and patient's medical history as the entry and returns medical recommendations.

Noteworthy, considering the heterogeneity of biological and clinical big data, ML methods can be applied to develop new models such as Bayesian Naïve models, and to classify data using classification algorithms, for example, random forest, support vector machine, neighbor-joining, and etc. Meanwhile, deep learning has become very popular and is frequently used in big data projects. Deep learning provides end-to-end learning methods for healthcare big data. Deep learning provides the possibility to translate the clinical complex data into an understandable human health report. Based on this feature, deep learning has been widely used by researchers to develop prediction algorithms among which early diagnosis of neurodegenerative diseases using MRI processing, detection of variations of Alzheimer's disease, early diagnosis of breast cancer using ultrasound images processing, and prediction of future diseases, based on their clinical history, can be mentioned. Moreover, some research groups have used deep learning to develop applications for the diagnosis and treatment of diseases. For example, DeepCare is a predictive application for medical recommendations designed by EHR. DeepPatient contains using clinical history of patients to predict future diseases. However, the development of deep learning models requires data availability, powerful computers, and accurate and intensive algorithms to represent reliable predictions (8-10).

Another important point is data privacy, particularly concerning health-related big data projects that should be investigated by various stakeholders, including providers, data scientists, and the users of deep learning platforms (11). In this regard, policies should be applied for data transfer (especially when cloud computing is applied), data storage, data usage, and the usage of analytics results by hospitals and organizations that cooperate in the develop-

ment of deep learning methods. Apart from the above-mentioned advantages for the improvement of patient care, healthcare data science also has economic benefits for healthcare systems. It is estimated that only in the United States the application of healthcare data science for cancer patients worth about 300\$ billion, from which two-third is achieved through declined costs, and the remaining is the value of extra benefits (12).

There is no doubt that healthcare plays a key role in healthcare advancements. So that, in a few years, EHR of patients will be available at a global scale, which will speed up both diagnosis and treatment processes. In other words, EHRs are considered as a very first step to achieve precise medicine and telemedicine (3). Currently, 24 countries and 9 consortia are planning for precision medicine, which have developed robust infrastructures, computational platforms, and mainly use ML methods.

Altogether, fast and accurate data analysis are a necessity for the success of healthcare data science. For this purpose, infrastructures and investments should be provided by various responsible institutions. Besides, the existence of healthcare departments and the availability of data scientists familiar with biological and medical information are necessary for this purpose. In conclusion, data science is showing us a new era in healthcare that can improve all aspects of patient care, will provide new job opportunities, and brings economic benefits to the health systems. Noteworthy, every day means a new step towards a more developed healthcare data science.

## Footnotes

**Authors' Contribution:** Contributions: Study concept and design: E. A.; analysis and interpretation of data: E. A.; drafting of the manuscript: E. A.; critical revision of the manuscript for important intellectual content: E. A.

**Conflict of Interests:** The author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Agrawal R, Prabakaran S. Big data in digital healthcare: Lessons learnt and recommendations for general practice. *Heredity (Edinb)*. 2020;**124**(4):525-34. [PubMed ID: 32139886]. [PubMed Central ID: PMC7080757]. <https://doi.org/10.1038/s41437-020-0303-2>.
2. Esposito C, De Santis A, Tortora G, Chang H, Choo KKR. Blockchain: A panacea for healthcare cloud-based data security and privacy? *IEEE Cloud Computing*. 2018;**5**(1):31-7. <https://doi.org/10.1109/mcc.2018.011791712>.

3. Hulsén T, Jamar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. *Front Med (Lausanne)*. 2019;**6**:34. [PubMed ID: 30881956]. [PubMed Central ID: PMC6405506]. <https://doi.org/10.3389/fmed.2019.00034>.
4. Dinov ID. Volume and value of big healthcare data. *J Med Stat Inform*. 2016;**4**. [PubMed ID: 26998309]. [PubMed Central ID: PMC4795481]. <https://doi.org/10.7243/2053-7662-43>.
5. Issa NT, Byers SW, Dakshanamurthy S. Big data: The next frontier for innovation in therapeutics and healthcare. *Expert Rev Clin Pharmacol*. 2014;**7**(3):293–8. [PubMed ID: 24702684]. [PubMed Central ID: PMC4448933]. <https://doi.org/10.1586/17512433.2014.905201>.
6. Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting big healthcare data. *Gigascience*. 2016;**5**:12. [PubMed ID: 26918190]. [PubMed Central ID: PMC4766610]. <https://doi.org/10.1186/s13742-016-0117-6>.
7. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. *Biomed Res Int*. 2015;**2015**:370194. [PubMed ID: 26229957]. [PubMed Central ID: PMC4503556]. <https://doi.org/10.1155/2015/370194>.
8. Frohlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: Data science enabling personalized medicine. *BMC Med*. 2018;**16**(1):150. [PubMed ID: 30145981]. [PubMed Central ID: PMC6109989]. <https://doi.org/10.1186/s12916-018-1122-7>.
9. Zhang S, Bamakan SMH, Qu Q, Li S. Learning for personalized medicine: A comprehensive review from a deep learning perspective. *IEEE Rev Biomed Eng*. 2019;**12**:194–208. [PubMed ID: 30106692]. <https://doi.org/10.1109/RBME.2018.2864254>.
10. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: Review, opportunities and challenges. *Brief Bioinform*. 2018;**19**(6):1236–46. [PubMed ID: 28481991]. [PubMed Central ID: PMC6455466]. <https://doi.org/10.1093/bib/bbx044>.
11. Iyengar A, Kundu A, Pallis G. Healthcare informatics and privacy. *IEEE Internet Computing*. 2018;**22**(2):29–31. <https://doi.org/10.1109/mic.2018.022021660>.
12. Harper EM. The economic value of health care data. *Nurs Adm Q*. 2013;**37**(2):105–8. [PubMed ID: 23454988]. <https://doi.org/10.1097/NAQ.0b013e318286db0d>.