

# Clinical big data in healthcare: A survey of medical computers

Mothanna Almahmood<sup>1</sup>, Sayel Abualigah<sup>1</sup>, Omar N. Elayan<sup>1</sup>, Laith Abualigah<sup>2,3,4,5,6,9</sup>,  
Raed Abu Zitar<sup>7</sup>, Anas Ratib Alsoud<sup>3</sup> and Sathishkumar V. E.<sup>8</sup>

<sup>1</sup>Computer Information System Department, Jordan University of Science and Technology, Jordan.

<sup>2</sup>Computer Science Department, Prince Hussein Bin Abdullah Faculty for Information Technology, Al al-Bayt University, Mafrq 25113, Jordan.

<sup>3</sup>Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan.

<sup>4</sup>Faculty of Information Technology, Middle East University, Amman 11831, Jordan.

<sup>5</sup>Applied science research center, Applied science private university, Amman 11931, Jordan.

<sup>6</sup>School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia.

<sup>7</sup>Sorbonne Center of Artificial Intelligence, Sorbonne University-Abu Dhabi, Abu Dhabi, United Arab Emirate

<sup>8</sup>Department of Software Engineering, Jeonbuk National University, Jeonju, Republic of Korea

<sup>9</sup>aligah.2020@gmail.com and sathish@jbnu.ac.kr

**Abstract.** Big data is a massive amount of information, measurements, and observations, where it has the power to provide a solution to the impossibilities. Recently, it has become the most trending topic in the field of data analysis because of its amazing potentials in extracting the hidden facts. Which attracted various sectors all over the world to collect and analyze the big data in order to improve their services and introduce high valuable products. Specifically, in the healthcare industry, different sources generate big data such as; hospital records, medical records of patients, and results of medical examinations. This type of data is related to the population healthcare, and it requires analysis in order to extract valuable knowledge. Nowadays, with the available high-end computing solutions for big data analysis. It becomes easy for researchers to have solutions that improve the healthcare level of the population. The promising thinking to give new technologies, high services, and big profits for healthcare, can revolutionize the medical solutions and help the community in overcoming the impossible cases. This research discusses essential clinical big data matters related to the healthcare sector by introducing a clear definition and features of the clinical big data in healthcare and its process. Also, by presenting analytics, applications, benefits, challenges, and future of the clinical big data technologies in the healthcare sector. This survey aims to review state of the art for the application of the clinical big data in the healthcare sector, in which it would be an apparent reference, where authors can refer to in their future research.

**Keywords:** big data, clinical big data, healthcare, machine learning; survey.

## 1. Introduction

In fact, the definition of big data is not fixed yet, as it changes over and over, because of the continuous developments in its field. In general, it is defined as a large set of data features with specific properties that require the use of new mechanisms in order to store and analyse it successfully [1]. The generated datasets by healthcare sources are characterized by the main aspects that are well known for big data, which are known as the 5 V's; Value, Volume, Velocity, Variety, and Veracity [2,3]. This type of data is widely spreading, as it is generated by different sources such as healthcare systems, research outputs, health insurance companies, government agencies, etc. The importance of this type of data is because of its targeted use in many aspects that serve the community. Undoubtedly, this type of data is complex, as there is no chance for errors during the analysis process. Despite this, the researchers indicated that the process of storing and analysing this data would positively affect the medical and healthcare vacancies [4]. Given the world's progress in using this data and applying various researches to it, even now, it is being used well. For example, according to McKinsey's reports, the use of aspirin by those who at risk of heart diseases, along with their early screening, will reduce the total cost of their care services by more than \$ 30 billion [5]. The availability of healthcare big data enables rapid detection of patients at risk, and this requires early intervention in treatment, which reduces the future costs of healthcare services. The clinical evaluation process revealed many patients' cases by several phenomena, which resulted from the strong correlation between the systems of the human body. Therefore, the prediction process here had great interest from researchers by building approaches to understand the symptoms and reduce the possibility of diseases. One of the most popular applications that have served the community efficiently is the use of clinical evaluation big data to obtain a comprehensive view of different disease cases. There are some cases in which the patient feels gradually increasing pain, and here comes the role of modern medicine in dealing with this issue by conducting a description of many analyzes and categorizing many cases to take actions that limit its increase. The medical image is an important source of data, as it is widely used to diagnose several medical conditions, in addition to its usefulness in evaluating the treatment followed by the patient [6]. These photos are varied according to need but require large storage capacity for use later. Image processing requires accurate algorithms in cases of decision-making. For example, the researchers used a big and complex number medical images in classifying the cancerous cell of breast cancer patients, applying the methods of the convolutional neural network in the image processing, in order to find out and avoid what might accompany the pathological condition in the future [7]. Likewise, extracting medical signals requires high-speed big storage, as it constantly flows data from patient-related screens. It should also be noted that the temporal and spatial nature has a striking aspect in building prediction systems that monitor patients. The researchers confirm that, over time, a great deal of progress has been made in the field of medical signal processing, extraction of features, and classification by using machine learning algorithms and neural networks techniques, which greatly served clinical applications [8]. Genomic is the branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes. Genomic big data analysis is a major challenge in the field of biology, and its importance lies in developing actionable predictions and recommendations. Genome-Wide Association Studies (GWAS) in the Human Genome Project are expanding to explore the hidden linkages between genes and diseases, by doing several tests to build predictive patterns that identify patients at high risk, as is the case in diabetics Type (i) [9]. It is worth to say that the use of these high-density datasets needs novel big data approaches and analytics. Regardless of how much the healthcare systems spend, the results presented improve better. Referring to the global catastrophe that occurred due to the recent spread of the Coronavirus, big data technologies have made good progress in combating its spread through careful monitoring of all cases around the world [10]. These analytics of the healthcare big data contribute greatly to the process of discovering and exploring the expected solutions. Also, it helps design policies that improve the level of healthcare provided to the population around the world. More importantly, its benefits are especially reflected in saving many lives, reducing high costs, and protecting the entire human community. Perhaps this field is having difficulty obtaining benefits quickly, due to the challenges facing the use of analyzes. The researchers improve that in their studies about the most important challenges [11-13], where they have

completely focused on: the problem of segmented data, maintaining patient privacy, generalization using unspecified standards, and the lack of data sharing. Overcoming such challenges will have a major role in discovering new patterns and working on modern, highly effective applications. Furthermore, the future scope will focus on solving the problems that prevent the researchers from growing up the technological processes of this field.

## **2. Clinical big data: Definition**

The research that deals with the definition of big data in the field of health care are continuing due to the technological advances that we witness in this field. The researchers have conducted systematic research by looking at a number of publications and reflecting on the advantages attributed to this type of data, in order to provide an appropriate definition for it. They defined it by focusing on the size component, as a large set of data that varies in size and speed [14]. In a study on big data in the field of health care that have been published between 2010 and 2015, the researchers defined it as containing large data sets from health care institutions, in addition to requiring adequate storage, analysis, and visualization to make productive and positive decisions [15]. Given the importance of big data in the healthcare field, researchers from India have undertaken a study mainly aimed at providing them with a comprehensive meaning [16-18], where they considered the big data in healthcare as critical in its importance, as the use of modern technologies and their application to them, has brought incredible benefits to the entire society, which has been achieved in various health care institutions around the world [2].

## **3. Clinical big data: Analytics**

Big data analytics is a set of tasks that are performed to extract information that supports decision-making. In a study of the importance of big data analytics in healthcare, researchers explained that providing such advanced analytical tools that we are witnessing now, we can obtain predictive information and build accurate recommendation systems, that help increase profits in the health business sectors, and provide unusual solutions to various difficulties faced by global health systems [19,20].

The impact of big data on health care is closely related to the available analysis tools. In a study in which researchers provide a description of the importance of this data, they suggest using advanced analysis tools such as Hadoop, because of its impressive results that include answering the most important medical questions, which in turn address and solve the most critical outstanding medical problems [21]. In another study on data analyzes tools in healthcare, researchers suggested the use of modern technologies such as Hadoop, MapReduce, and Cloud Computations. Where they indicated that it will help to identify diseases and risks effectively, and will also enhance the process of effective decision making and development [22].

It was also known that the most critical process before beginning the analysis is to search for new sources of data. In light of this topic, the researchers summarize in a book about big data in the field of healthcare, that health care systems contain clinical source for collecting big data [23]. In fact, there are many approaches to analyzing clinical big data, each of which differs in its work and results. In a study on big data analyzes in healthcare and the standard used methodologies, the researchers summarized a primary methodology based on significant stages, where these stages were represented by identifying the need for analysis, developing suggestions, implementing steps, testing results, and validating results [24].

Also, in a study conducted by researchers with the aim of removing the barriers that most analysts encounter during their work, they find that it should focus on two main components; the computing platform for analytics-driven healthcare and applications that were built for data sharing [25]. Moreover, the researchers have recommended in the study [26], to use the latest big data analytical tools, with the aim of solving the real health problems in the world and reaching a high level of healthcare for the whole society.

#### **4. Clinical big data: Applications and benefits**

The goal of analyzing this type of big data is to obtain useful applications that serve the community and raise the level of health care services. In previous research that aimed on the most critical applications in this field, it was mentioned on the results of researchers in the study [27], that the analysis of these big data has produced efficient applications such as: identification of health risks, disease surveillance, telemedicine, predictive analyzes and Precise strategic layouts. Other researchers also clarified in another study about these applications that their interest is in providing services related to all patients' needs and addressing community health issues with high efficiency [28]. In the financial side, a study on big data in the field of health care and applications [29], the researchers pointed out that improving health care applications using modern big data technologies helps in changing the cost size for the better. With the proliferation of modern big data analytical tools, action must be taken, in which it gives the advantages of this type of data. And the researchers note in the study [30], that the analyzes based on this type of data, in addition to the effective use of the modern big data analysis tools, are undoubtedly able to reduce risks and resist problems related to society healthcare level.

#### **5. Clinical big data: Challenges**

From the beginning of the use of advanced data science technologies, and especially the emergence of big data technologies, it was followed by challenges that affected the goal of reaching the required range of knowledge. In a study on big data in healthcare and the challenges it faces, the researchers have found that the most critical challenges facing data analyzes are data collection, the cost of its creation, and protecting patient privacy [31]. The effectiveness of the results and their continuation in this field depend strongly on the process of exchanging data and sharing it widely because this process provides greater imposition to achieve better results. From this standpoint, it has been seen that the health institutions who monopolize their data increases the challenges facing researchers in this field. Researchers confirmed this to us in their study about the most important challenges that we face in this field [32], where they confirm that data privacy is a bad obstacle that we must avoid by sharing data on a large scale, to provide greater opportunities to obtain high-level healthcare for the community. Perhaps most importantly, the problem of mining in this data, as the big data of healthcare comes from multiple sources and spread in various forms such as structured, unstructured, and semi-structured data. In a study on big data in healthcare and the challenges it faces [33], the researchers point out that the sources and forms of this type of data greatly affect its privacy, if there is a need to share it. It is important to continue the process of interoperability of big data in the health care field, to extrapolate the needed results and find out the changes that occur in them with respect to the timestamp. In the study [34], the researchers have touched on the challenges we face in dealing with this type of data and stated that the lack of cooperation between researchers constituted a major challenge in enhancing the interoperability and reduced the opportunities for new developments research.

#### **6. Clinical big data: Future scope**

The hopes built on modern big data technologies are now seen as facts that have proven surprising results for us. Despite the strength of this big data, hopes are still open-ended for new developments. Logically, the increased availability of information gives more opportunities to obtain knowledge. This is what the researchers pointed out in their study on the future of big data in healthcare [35,36], where they praised future work on developing a detailed model of the human, which will enable us to strengthen our knowledge in the fields of disease. Remote patient monitoring is an essential and wonderful development, and that is what the researchers recommended in studies that focused on the future development of big data in the field of health care [37,38], where they indicated that the existing analyzes would be able to design a device used to monitor the patient remotely, and that will achieve unprecedented results that would help in improving the treatment process. In studying the hopes on big data in healthcare, the researchers note in the studies [39-41], that working on large groups of data in a participatory manner, will lead us to develop the ecosystem and raise the healthcare level of the community. Moreover, the researchers hope in their study about the future of society with the

development of big data technologies in this field [42], to grow up of visual exploration tools faster, so that this service becomes self and accessible to the community.

## 7. Conclusion

This study aimed to enhance the benefit aspects of clinical big data technologies on the development of the healthcare sector. It has presented the main concept of the clinical big data and its features, Volume, Variety, Velocity, Veracity, and Value. Then, it has described the process of generating a value from clinical big data. After describing the analytics of big data, it has discussed some technologies with benefits. Moreover, the challenges were proposed with details. Finally, the future scope of the clinical big data was described. Based on these literatures, the study concludes that the clinical big data technologies have a positive effect on the quality of the healthcare sector.

## References

- [1] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Libr. Rev.*, vol. 65, no. 3, pp. 122–135, 2016.
- [2] A. Hameed Shnain, H. Jasim Hadi, S. Hadishaheed, and A. Haji Ahmad, "Big Data and Five V'S Characteristics," *Int. J. Adv. Electron. Comput. Sci.*, no. 2, pp. 2393–2835, 2015.
- [3] Ishwarappa and J. Anuradha, "A brief introduction on big data 5Vs characteristics and hadoop technology," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 319–324, 2015.
- [4] R. Pastorino et al., "Benefits and challenges of Big Data in healthcare: An overview of the European initiatives," *Eur. J. Public Health*, vol. 29, pp. 23–27, 2019.
- [5] "The big-data revolution in US health care: Accelerating value and innovation | McKinsey." [Online]. Available: <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>. [Accessed: 26-Mar-2020].
- [6] H. B. Erdogan, "The Importance of Medical Imaging and Overview of Using 3D Printer," no. September, pp. 5–9, 2018.
- [7] R. Turkki et al., "Breast cancer outcome prediction with tumour tissue images and machine learning," *Breast Cancer Res. Treat.*, vol. 177, no. 1, pp. 41–52, 2019.
- [8] K. Pal, K. Mitra, A. Bit, S. Bhattacharyya, and A. Dey, "Medical Signal Processing in Biomedical and Clinical Applications," *J. Healthc. Eng.*, vol. 2018, p. 3932471, 2018.
- [9] "Genomics and the role of big data in personalizing the healthcare experience – O'Reilly." [Online]. Available: <https://www.oreilly.com/content/genomics-and-the-role-of-big-data-in-personalizing-the-healthcare-experience/>. [Accessed: 26-Mar-2020].
- [10] "Monitoring Coronavirus outbreak using Big Data," no. February, 2020.
- [11] S. White, "A review of big data in health care: challenges and opportunities," *Open Access Bioinformatics*, no. November, p. 13, 2014.
- [12] M. Adibuzzaman, P. DeLaurentis, J. Hill, and B. D. Benneyworth, "Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database," *AMIA ... Annu. Symp. proceedings. AMIA Symp.*, vol. 2017, pp. 384–392, 2017.
- [13] L. Hong et al., "Big Data in Health Care: What Is So Different About Was ist so anders am Neuroenhancement?" vol. 1, no. 2, pp. 122–135, 2018.
- [14] E. Baro, S. Degoul, R. Beuscart, and E. Chazard, "Toward a literature-driven definition of big data in healthcare," *Biomed Res. Int.*, vol. 2015, 2015.
- [15] S. SA, "Big Data in Healthcare Management: A Review of Literature," *Am. J. Theor. Appl. Bus.*, vol. 4, no. 2, p. 57, 2018.
- [16] P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare Big Data," no. October, pp. 72–100, 2018.
- [17] M. Chen, S. Mao, Y. Zhang, and V.C. Leung, "Big Data: Related Technologies, Challenges and Future Prospects," Springer Cham Heidelberg New York Dordrecht London, 2014

- [18] A. Bhadani, and D. Jothimani, "Big Data: Challenges, Opportunities and Realities," in M.K. Singh, and D.G. Kumar (Eds), *Effective Big Data Management and Opportunities for Implementation*, Pennsylvania, USA, IGI Global, pp. 01-24, 2016
- [19] P. Kamakshi, "Importance of Big data in Healthcare System," *Int. J. Appl. Eng. Res.*, vol. 13, no. 15, p. 4, 2018.
- [20] M. Ambigavathi and D. Sridharan, "Big Data Analytics in Healthcare," 2018 10th Int. Conf. Adv. Comput. ICoAC 2018, vol. 4, no. 1, pp. 269–276, 2018.
- [21] S. Kumar and M. Singh, "Big data analytics for healthcare industry: impact, applications, and tools," *Big Data Min. Anal.*, vol. 2, no. 1, pp. 48–57, 2018.
- [22] M. A. Alkhatib, A. Talaei-Khoei, and A. H. Ghapanchi, "Analysis of research in healthcare data analytics," *ACIS 2015 Proc. - 26th Australas. Conf. Inf. Syst.*, pp. 1–16, 2015.
- [23] J. M. Cavanillas, E. Curry, and W. Wahlster, "New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe," *New Horizons a Data-Driven Econ. A Roadmap Usage Exploit. Big Data Eur.*, pp. 1–303, 2016.
- [24] W. Raghupathi, V. Raghupathi, W. Raghupathi, and V. Raghupathi, "Big data analytics in healthcare: promise and potential to cite this version: HAL Id: hal-01663474 Big data analytics in healthcare: promise and potential," vol. 2, no. 1, 2017.
- [25] Intel, "White Paper Healthcare: Using Analytics to Improve Healthcare Outcomes.," 2015.
- [26] C. K. Reddy and C. C. Aggarwal, "Healthcare data analytics," *Healthc. Data Anal.*, pp. 1–727, 2015.
- [27] L. Elezabeth, V. P. Mishra, and J. Dsouza, "The Role of Big Data Mining in Healthcare Applications," 2018 7th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2018, no. August 2018, pp. 256–260, 2018.
- [28] H. Kumar and N. Singh, "Review paper on Big Data in healthcare informatics," *Int. Res. J. Eng. Technol.*, vol. 4, no. 2, pp. 197–201, 2017.
- [29] A. et al. Heinrich, "Big Data Technologies in Healthcare," *Big Data Technol. Healthc. Needs, Oppor. challenges*, p. 31, 2016.
- [30] J. K. Bains, "Big Data Analytics in Healthcare- Its Benefits, Phases, and Challenges," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 6, no. 4, pp. 430–435, 2016.
- [31] Q. K. Fatt and A. Ramadas, "The Usefulness and Challenges of Big Data in Healthcare," *J. Healthc. Commun.*, vol. 03, no. 02, pp. 1–4, 2018.
- [32] S. Patel and A. Patel, "A Big Data Revolution in Health Care Sector: Opportunities, Challenges and Technological Advancements," *Int. J. Inf. Sci. Tech.*, vol. 6, no. 1/2, pp. 155–162, 2016.
- [33] V. Carr, R. M., Orange, A., & Khungar, "乳鼠心肌提取 HHS Public Access," *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2016.
- [34] S. Salas-Vega, A. Haimann, and E. Mossialos, "Big data and health care: Challenges and opportunities for coordinated policy development in the EU," *Heal. Syst. Reform*, vol. 1, no. 4, pp. 285–300, 2015.
- [35] A. G. ALEXANDRU, I. M. RADU, and M.-L. BIZON, "Big Data in Healthcare - Opportunities and Challenges," *Inform. Econ.*, vol. 22, no. 2/2018, pp. 43–54, 2018.
- [36] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis, and future prospects," *J. Big Data*, vol. 6, no. 1, 2019.
- [37] C. Austin and F. Kusumoto, "The application of Big Data in medicine: current implications and future directions," *J. Interv. Card. Electrophysiol.*, vol. 47, no. 1, pp. 51–59, 2016.
- [38] GE & UCSF, "Big Data, Analytics & Artificial Intelligence the Future of Health Care is Here," p. 19, 2016.
- [39] P. Muthulakshmi and S. Udhayapriya, "a Survey on Big Data Issues and Challenges," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 6, pp. 1238–1244, 2018.
- [40] E. Summary, "Big data whitepaper," 2009.

- [41] A. Pashazadeh and N. J. Navimipour, "Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review," *J. Biomed. Inform.*, vol. 82, no. November 2017, pp. 47–62, 2018.
- [42] S. Mukherjee and R. Shaw, "Big Data- Concepts, Applications, Challenges, and Future Scope," *Int. J. Adv. Res. Comput. Commun. Eng. Vol.*, vol. 5, no. 2, pp. 66–74, 2016.