Examining Big Data in Medicine:
Applications, Challenges and Benefits

Matthew N. O. Sadiku, PhD
Kazeem B. Olanrewaju, PhD
Sarhan M. Musa, PhD
Roy G. Perry College of Engineering,
Prairie View A & M University, Prairie View, TX, USA

Abstract

Advances in technology have created and will continue to create changes in medical science and practice. In recent years, the medical community has seen massive amount of information, which is collectively known as "big data." Literatures from various sources were reviewed to explore the application of data analytics in medicine, the benefits and challenges associated with its use. With recent advances in digital technology, the ability to gather and examine big data has become far cheaper and faster. With the promise of data-driven knowledge and intelligent decision-making, big-data projects are going on in every branch of medicine. However, the process of engaging stakeholders to benefits from the information provided by big data remain a major issue. This paper presents how big data and big data analytics are used in medicine.

Keywords: Big Data, Big Data Analytics, Medicine

Introduction

Technology plays a fundamental role in every aspect of the modern life. It is revolutionizing our understanding and treatment of disease and transforming how medicine is practiced. Modern technologies (such as smart phone and Internet) are generating enormous amounts of data, created daily at an incredible rate from activities including social medial, sensors, and monitoring tools. It is needless to say that data by themselves are useless. To be useful, data must be analyzed, interpreted, and acted on.

Data is an important source of knowledge, insights, and value. The ability to transform data into useful knowledge will disrupt many areas in medicine. Data is transforming every aspect of our world, including medicine.

Big data can be described as the rapidly increasing size of available data, the speed with which it is produced, and the ways in which it is represented. The bigness of data is not its absolute size, but the task of deriving wisdom from it. Generating large, complex data and learning from it has become routine in commercial enterprises (Wakahia, 2019). Industries such as Google, Netflix, and Amazon that have harnessed big data to provide accurate and personalized real time information. Facebook, Twitter and other social media generate a variety of data and provide a view into the locations, health behaviors, and social interactions of users. Information from vaccine supply database has been of tremendous resources in optimizing the distribution of vaccine for preventable disease in country like Nigeria (Hirsh Bar Gai, Graybill, Voevodsky, & Shittu, 2018; Shittu, Harnly, Whitaker, & Miller, 2016). Compared with industry, the situation is different with medicine partly because medical data is complex and contains sensitive, private, personal information. The medical community is slowly adopting big data approaches because they have realized the practical purposes of such approaches. However, the integration of new approaches will require new thinking on the part of medical authorities (Krumholz, 2014) .

Medicine is a profession that deals with a lot of information. For example, the vast amount of information that is continuously generated from patient medical experience remains untapped. A new thinking requires not wasting this potential source of knowledge. There are three primary sources of big data in medicine. The first source includes electronic medical records (EMRs), claims and billing data, imaging data, pharmacy data, past medical history, active medical problems, immunizations, etc. Imaging data may be considered to the largest in volume. The second source is omics, including genomics, proteomics, and metabolomics. The third is patient portals, i.e. data from mobile device (such as smartphones), wearables, and Internet activities, sensors, and monitoring tools (Bauer,). This allows patients to access personal health information stored in a health care organization's medical database/

Review Methodology

This review was approached by extracting information from various sources to examine the application of big data in medicine, its benefits and challenges. First, a detail search of peer review journals, books, conference proceedings paper and online published literatures was conducted based on a broad range of key terms such as big data, big data analytics, and medicine. Ten databases were explored including, ACM Digital Library, Computer Source, Engineering Village (also INSPEC), IEEE Engineering, JSTOR, Wiley Online Library (InterScience), ProQuest Databases (Dissertations and Theses), Google Scholar, Social Science Databases and Science Direct. Lastly, information was also extracted from other peer reviewed journals such

as Sage Publication Journals, Taylor and Francis Journals, Health Affairs, and Journal of Scientific and Engineering research. The search process unearthed 16 peer reviewed journals articles, 3 books, 1 conference proceeding paper and 2 online literature published from 2014 to 2019. The information mined and harnessed from these sources will be utilized in this paper to uncovered and examine big data analytics in medicine. Thus, detail description of the nature and characteristics of big data will be presented. Application of data analytics in various branches of medicine such as precision medicine, cardiovascular care, biomedical research, psychiatry, medical data, and medical imagine will be elucidated. Key benefits along with other auxiliary dividends will be discussed. Challenges associated with data analytics in medicine will be highlighted and expounded.

Big data characteristics

Big data refers to huge volume of data so large that it is difficult to process using traditional techniques. Big data is sometimes expensive data. Although there are various definitions of big data, it is commonly described to include six V's, which are illustrated in Figure 1 (Ristevski, B. and Chen, M., 2018) and explained as follows (Sadiku, M. N. O., 2019; Sadiku, M. N. O., 2019; Sadiku, M.N.O., Tembely, M., and Musa, S. M., 2016).

- *Volume:* Vast volumes of data produced, typically over one petabyte in volume. In healthcare, we have large volumes of data coming in EMRs alone.
- *Velocity:* The speed at which a data user can access and analyze the data. Velocity is needed in healthcare since it allows healthcare providers to exchange and use data in a timely manner.
- *Variety:* This refers to structured, semi-structured, and unstructured data. This comes from diverse sources of data in healthcare including multimedia, social media, and financial transactions**.**
- *Veracity:* This refers to the quality of data produced. Healthcare data must be relevant, reliable, and error-free because life and death decisions depend on accurate information.
- *Value*: This is the most important aspect of the big data. It denotes the value derived from the analysis of the existing data. The value for big data in healthcare today is largely limited to research.
- *Variability*: This regards about consistency of the data over time
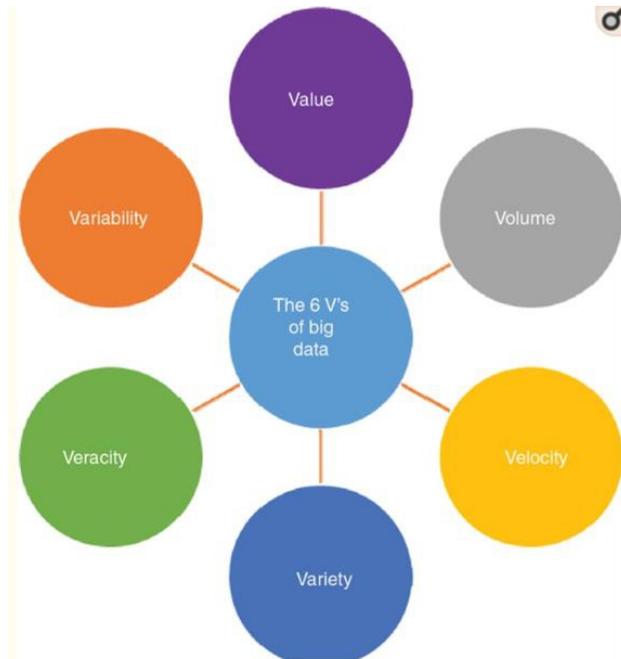
Figure 1. Big data's 6V (Ristevski, B. and Chen, M., 2018).

There are three key areas where big data is different from conventional analyses of data samples (Mayer-Schonberger & and Ingelsson, 2018): (1) Data are captured more comprehensively relative to the phenomenon under study; (2) Data are often analyzed using machine learning tools rather than traditional statistical methods; (3) The purpose of the analyses of data is no longer simply answering existing questions. Because big data approaches differ so fundamentally from small data approaches, research processes and mindsets need to change.

Big data in medicine uses large volumes of medical information to look for trends. It refers to these various large and complex data, which they are difficult to analyze with traditional software or hardware. It requires network connectivity with other locations, physicians, databased, and scientists. Big data analytics in medicine covers integration and analysis of large amount of complex heterogeneous data such as biomedical data, experimental data, electronic health records data, and social media data. It includes data quality control, analysis, modeling, interpretation and validation. The steps of big data analysis are shown in Figure 2(Thiyagarajan, 2018).
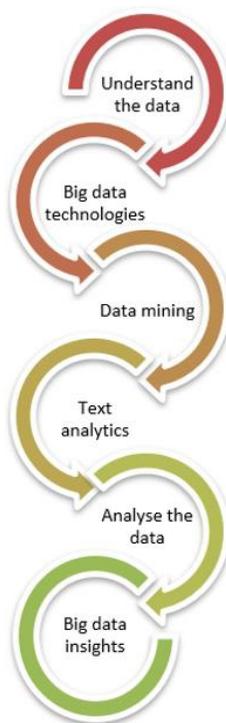
Figure 2. The steps of Big data analysis(Thiyagarajan, 2018).

Applications in medicine

Healthcare organizations in many countries have developed several models of healthcare information systems. These models for personalized, predictive, participatory and preventive medicine are based on using of electronic health records (EHRs) and huge amounts of complex biomedical data, which are inherently in heterogeneous formats and can be structured, semi-structured or unstructured. Integrating such diverse data makes big data analytics to combine several disciplines, such as bioinformatics, medical imaging, sensor informatics, medical informatics, health informatics, and computational biomedicine (Ristevski, B. and Chen, M., 2018).

The processing of the big data in medicine can be accelerated by using supercomputers, cloud computing and powerful central processing units (CPUs), graphics processing units (GPU) and field-programmable gate arrays (FPGAs) with parallel processing methods.

Applications that use big data must address multitasking, data visualization, and cognitive processing. Software platforms that can handle big data in medicine include Apache Hadoop MapReduce that use massive parallel processing.

Different areas where big data are applied in medicine include precision medicine (a.k.a. personalized medicine), clinical medicine,

translational medicine, general medicine, genomic medicine, cardiology, psychiatry, biomedical research, clinical research, medical laboratory, nephrology, radiography, drug discovery, vaccine supply chain, biomarker development, and basic research in cancer, rare diseases, neurodegeneration, and diabetes. Some of these applications are shown in Figure 3 (Bychenkov, 2015) and discussed as follows:

- *Precision Medicine*: This can be regarded as an approach to provide the right treatments to the right patients at the right time. It replaces the imprecise one-size-fits-all medicine. Precision medicine allows individually tailored medicine and enables patients to receive care best matched to their specific health condition. It combines comprehensive data collected about an individual's genetics, environment, and lifestyle (such as smoking and heart disease), to advance disease understanding, aid drug discovery, and ensure delivery of appropriate therapies. Combining big data and personalized or precision medicine holds great promise for the future of medicine. Big data holds promise transform clinical decision making toward precision medicine. Precision medicine powered by big data will also reduce costs(Hopp, Li, & and Wang, 2018; Huang, Mulyasasmita, & and Rajagopal, 2016).

- *Cardiovascular Care*: Cardiovascular diseases are well known to be heterogeneous in nature. Applying big data analytic to cardiovascular care will translate into better care at a lower cost. Although there are several risk models related to cardiovascular conditions, big data analytics may yield more powerful prediction of outcomes ranging from mortality to patient-reported outcomes(Shah & and Rumsfeld, 2017). Big data has tremendous potential to improve cardiovascular quality and outcomes of care.

- *Biomedical Research*: Big data is radically transforming biomedical research. The main issue in biomedical research is how to extract knowledge from big data. In the era of digital biomedicine, biomedical scientists are confronting with dealing with ever larger sets of data. Big data in biomedicine has tremendous potential to transform biomedicine, healthcare, drug discovery, and development (Costa, 2014). It is also expediting biomedical discovery. In view of the complexity of biomedical big data, there is an urgent need to produce bioinformatics professionals who are capable of processing, analyzing, and interpreting big data.

- *Psychiatry:* The growth of big data in psychiatry will provide unprecedented opportunities for exploration. Psychiatrists will increasingly have to evaluate results from research studies and commercial analytical products that are based on big data. Big data

from clinical, administrative, and imaging will increase understanding of existing and new questions in psychiatry (Monteith, 2015).

- *Medical Data:* The electronic medical records replace the conventional paper-based records. Adequately utilized, these reservoirs of data can be a source of knowledge to fuel a learning health care system. The vast amount of data is continuously generated from patients' experiences with health remains largely untapped. A hallmark of big data is combining different data sources and types. A premise of big data is that data can be the source material for vast amounts of medical research that can ultimately meet the needs of future patients, clinicians, and other health professionals (Krumholz, 2014). Big data analytics involves analyses of electronic medical records (EMRs) that are in structured and unstructured data formats.

- *Vaccine Supply Chain Database:* Application of big data analytics to the healthcare system, especially for most developing countries, can significantly address the challenges of accessibility and availability of drugs and vaccines for preventable diseases. These issues are most often than not connected with poor transportation and logistics systems in the delivery of these vaccines to the communities. Data on vaccine supply chain was mined, harnessed and coupled into a simulation model to create a scenario that integrates locations while minimizing cost, improving accessibility and optimizing storage facilities (Hirsh Bar Gai et al., 2018; Shittu et al., 2016).

- *Medical Imaging:* This is a non-invasive technique of visualizing the anatomy of human body. Among big data types, imaging data can be considered the largest in volume. Diagnosis of major diseases is incomplete without medical imaging. For example, medical images help in early detection, diagnosis, and prognosis of neurological disorders and neurodegenerative diseases. Big data analytics is applied to medical image processing (Nair & and Ganesh, 2016).
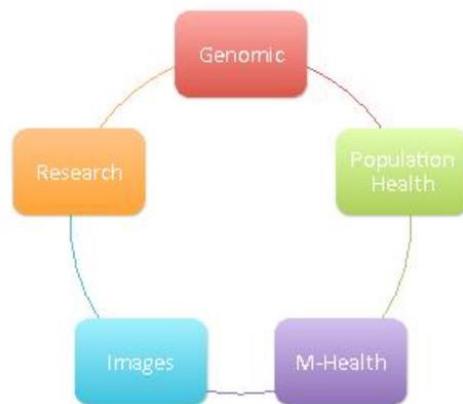
Figure 3. Some applications of big data in medicine (Bychenkov, 2015).

Benefits

Big data has the potential to provide unparalleled opportunities to answer questions that cannot be answered by clinical trials. There is a benefit from not being blind to the information or dependent on a physician to interpret it for you. The new knowledge provided by big data analytics techniques should benefits to the patients, clinicians, and health policy makers. Big data promises to connect physicians, nurses, radiologists, laboratory technologists, researchers, pharmaceutical, and healthcare companies to personalized patient experiences everywhere. Other benefits include the following (Ijaz, 2018):

- *Change:* Big data analytics has the potential to drive real change in clinical practice, from personalized therapy to electronic health record mining.
- *Intelligence*: Big data has the ability to combine data from numerous sources and provide "intelligence" not derived from any single data source. It allows the prediction of patterns, trends, and outcomes.
- *Reduced Costs:* By using predictive analysis, healthcare expenses can be reduced.
- *Eliminating Human Errors:* Making mistakes is inevitable and human. For example, medical professionals can prescribe wrong medicine. Big Data can dramatically decrease the likelihood of errors.

Challenges

Although big data promises great benefits for medical research, using big data in medicine is difficult. Some challenges must be addressed in order for big data to have wide acceptance in the medical community. A major challenge around big data has to do with how one engages others to benefit

from that information. In order for big-data projects to succeed, commitment and participation of physicians is required. There is an urgent need to train thoroughly clinical investigators in data science. Some the other challenges include (Bauer, ; Hulsen T. et al., 2019).

- *Ethics and privacy:* The main issues in big data in medicine are security and privacy of the patients as well as ownership, stewardship, and governance of data. These issues are important when gathering and accessing data in health care. The use of big data for medical research poses unprecedented ethical challenges. These information privacy laws are a major hurdle that prevents comprehensive data use medicine. Big data projects are often distributed across nations with different laws, making issues of privacy, and consent complicated solving challenges associated with data privacy, and security.
- *Collaboration:* Every step in the acquisition, processing, cleaning, analysis, and interpretation of big data is difficult because projects require the active collaboration of people with a wide range of expertise such as physicians, statisticians, biologists, software engineers, mechanical engineers, network security experts, and project managers.
- *Risks:* Different risks should to be carefully considered and monitored when looking at new sources of big data.
- *Standardization:* There is need for standardization of data content, format, and clinical definitions. Standardization of samples allows for data generated from one individual to be used in other related studies.
- *Intellectual Property* (IP): IP protection, mainly provided by copyrights and patents, is the pillar of national research policies. It helps in effectively translating innovation by commercialization. In the absence of such protection, companies will be unwilling to invest in the development of diagnostic tests or treatments. However, the operation of the IP system is being fundamentally changed by new data driven techniques.
- *Noise*: Handling noisiness and incompleteness of EHRs are still challenging. More data often means more noise in which any true signals drown.

Conclusion

Today, medical data in the world is exploding. Big data seems to promise a revolution, sweeping away conventional approaches to medical science and practice. Big data analytic may seem like a new disruptive technology and the current clinical practice may need to undergo change to incorporate the new big data based technical solutions. Training in data

science will need to become a part of medical curricula so that clinicians will be more comfortable with the evidence that is generated by the new tools. Big data is poised to have enormous impact on the medicine, especially on personalized medicine.

Data analytics has received tremendous applause in adopting robust computational software with algorithms carefully structured and designed to mine data pertinent to quotidian human needs such as health, food supply, logistics and transportation, to mention but a few. It is evident that big data analytics resourcefulness to dispense better healthcare deliverables is not only limited to core medical practitioners but also lend its usefulness to all other stakeholders in the field. Therefore, data analytics is a tool that can be adopted to assess patient's health from pretreatment check-up, prognosis, diagnosis, therapy and the delivery of vaccine and drug for post-treatment care. The future of medicine will definitely be enabled by big data. More information about the use of big data in medicine can be found in books in (Lytras & and Papadopoulou, 2018; Suganya, Rajaram, & and Abdullah, 2018; Wang, Li, & and Perrizo, 2105).

References:
1. Bauer, M.Big data, technology, and the changing future of medicine. Retrieved from https://www.medicographia.com/2018/02/big-data-technology-and-the-changing-future-of-medicine/
2. Bychenkov, K. (2015). Big data in medical image processing. Retrieved from http://0x1.tv/Big_data_in_medical_image_processing_(Konstantin_Bychenkov,_SECR-2015)
3. Costa, F. F. (2014). Big data in biomedicine. Drug Discovery Today, 19(4), 433-440.
4. Hirsh Bar Gai, D., Graybill, Z., Voevodsky, P., & Shittu, E. (2018). Evaluating scenarios of locations and capacities for vaccine storage in nigeria doi:https://doi.org/10.1016/j.vaccine.2018.04.072
5. Hopp, J. W., Li, J., & and Wang, G. (2018). Big data and the precision medicine revolution. Production and Operations Management, 27(9), 1647-1664.
6. Huang, B. E., Mulyasasmita, W., & and Rajagopal, G. (2016). The path from big data to precision medicine. Expert Review of Precision Medicine and Drug Development, 1(2), 129-143.
7. Hulsen T. et al. (2019). From big data to precision medicine. Frontiers in Medicine, 6
8. Ijaz, R. (2018). 5 ways big data is transforming the medical field. Retrieved from https://www.smartdatacollective.com/5-ways-big-data-is-transforming-the-medical-field/

9. Krumholz, H. M. (2014). Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system. Health Affairs, 33(7), 1163-1170.
10. Lytras, M. D., & and Papadopoulou, P. (Eds.). (2018). Applying big data analytics in bioinformatics and medicine IGI Global.
11. Mayer-Schonberger, V., & and Ingelsson, E. (2018). Big data and medicine: A big deal? Journal of Internal Medicine, 283(5), 418-429.
12. Monteith, e. a. (2015). Big data are coming to psychiatry: A general introduction. International Journal of Bipolar Disorders, 3(21)
13. Nair, S. S. K., & and Ganesh, N. (2016). An exploratory study on big data processing: A case study from a biomedical informatics. Paper presented at the Proceedings of the 3rd MEC International Conference on Big Data and Smart City,
14. Ristevski, B. and Chen, M. (2018). Big data analytics in medicine and healthcare. Journal of Integrative Bioinformatics, 15(3)
15. Sadiku, M. N. O. (2019). Emerging internet-based technologies (1st ed.). Boca Raton, FL: CRC Press.
16. Sadiku, M.N.O., Tembely, M., and Musa, S. M.,. (2016). Big data: An introduction for engineers. Journal of Scientific and Engineering Research, 3(2), 106-108.
17. Shah, R. U., & and Rumsfeld, J. S. (2017). Big data in cardiology. European Heart Journal, 38(24), 1865-1867.
18. Shittu, E., Harnly, M., Whitaker, S., & Miller, R. (2016). Reorganizing nigeria's vaccine supply chain reduces need for additional storage facilities, but more storage is required. Health Affairs, 35(2), 293-300. doi:10.1377/hlthaff.2015.1328
19. Suganya, R., Rajaram, S., & and Abdullah, A. S. (2018). Big data in medical image processing. Boca Raton, FL: CRC Press.
20. Thiyagarajan, A. (2018). Big data usage in medical field. ACTA Scientific Medical Sciences, 2(4), 32-34.
21. Wakahia, S. K. (2019). Big data analytics and electronic resource usage in academic libraries: A case study of a private university in kenya.15(13), 344-359. doi:10.19044/esj.2019.v15n13p344
22. Wang, B., Li, R., & and Perrizo, W. (Eds.). (2015). Big data analytics in bioinformatics and healthcare IGI Global.