



The Role of Big Data in Predicting Health Outcomes

Mwende Muthoni D.

Faculty of Medicine Kampala International University Uganda

ABSTRACT

By offering novel approaches for forecasting health outcomes, big data is transforming healthcare. The exponential growth of health-related data, obtained via biopharmaceuticals, wearable devices, and digital health records, has resulted in the creation of extensive analysis programs. These data-driven methodologies provide more precise forecasts, therefore assisting hospitals in resource allocation, tailoring patient treatment, and averting possible health emergencies. Nevertheless, the use of big data also brings about notable obstacles, such as issues about privacy, increasing processing complexity, and the need for specialised analytical methods. In this research, the significance of forecasting health outcomes, the difficulties of using big data in healthcare, and the possible prospects it offers are examined. This review analysed detailed case studies that demonstrate the efficacy of big data in forecasting patient health paths, enhancing healthcare administration, and assisting in preventative medicine.

Keywords: Big data, health outcomes, healthcare analytics, predictive modeling, machine learning.

INTRODUCTION

The rapid growth in biopharmaceutical and wearable devices and the digital records industry is contributing to the large amount of health-related data produced [1]. It has been noted that the trend of a larger volume of healthcare-related big data is perhaps converging into four well-known data aggregates - Volume, Veracity, Velocity, and Variety (4V's). These data aggregates are perceived to possess several characteristics in comparison to small healthcare data: documents, structured/unstructured text, images, video, audio recordings, and data posted regularly on social networks like Facebook, Twitter, etc [2]. Furthermore, the 4V's bring out another important property categorically related to healthcare: the value of big data, which maximizes efficiency and benefit for personalized medicine. Therefore, the myriad blend of 4V's and the unique property of the value of big data not only differentiates big data from other data but also addresses the significance of big data to healthcare [1, 2]. For healthcare facilities, operational and cost optimization are major agendas. Big data has the potential to provide answers on earlier unseen methods to visualize disease. The larger volumes of information also provide an opportunity to validate short-run findings in longitudinal databases. In this line, the requirement of big data lies in predicting health outcomes using health services: which is collecting and analyzing all fine-grained healthcare electrical digital data from millions of their patients. For example, using this big data to draw the path-lines of patients' health, forecast what's likely to happen next [3, 4].

IMPORTANCE OF PREDICTING HEALTH OUTCOMES

Predicting health outcomes is important because it impacts healthcare. For example, hospitals can use a prediction of health outcomes to decide on allocating resources, while patients can receive personalized care. The healthcare system may also adjust health management programs effectively and on time. However, predicting health outcomes can also help avoid trivial or critical errors that might cause physical injury, mental harm, or even result in fatal outcomes. After all, in our society of rapid health information dissemination, providing accurate health outcome predictions can effectively increase people's

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

awareness of health, lifestyle, genetic and other information, and encourage the public to take the personal initiative of taking preventive medicine, subjecting to related medical procedures, etc. All of which preach maintaining health and reducing stress on the health system [5, 6]. The imperfect availability of data may result in the prediction of outcomes being completely unpredictable. However, it is possible that the available data is sufficient but insufficient to capture and identify all the predictive abilities needed to make accurate predictions of the future. The big data in health research focuses on accurate prediction of predictor or dependent variables and the correlation between them. Also, the more data, the bigger the proof present in the results. The use of large-scale biological data in the form of big data would help improve health outcomes [7, 8].

CHALLENGES AND OPPORTUNITIES IN UTILIZING BIG DATA FOR HEALTH PREDICTIONS

Although estimates of potential opportunities abound, practical realization has long lagged, and previous commentaries have often described obstacles. Notions of freedom and privacy of medical data are central to individuals and society, so this shift must be approached carefully. Integrating the diverse sources of health and medical data creates a set of challenges that range from the computational to the normative. Analyzing "big data" involves quantitative and computational challenges. At the level of analysis, one must find appropriate methods for integrating diverse data, as both quality and methodological claims differ widely over medical, environmental, and diagnostic data. Furthermore, the sheer quantity, heterogeneity, and distributed nature of "big data" hamper efforts when securing systems and data against cyber threats and keeping unauthorized users from accessing or stealing protected health information. Curating the data against such threats is resource-intensive and is likely to slow real-time response following analysis. Nevertheless, the potential benefits are vast [9, 10]. Accurate prediction of future health and complex health trajectories requires broader perspectives. A host of factors - including one's age, prior medical history, genetics, health behaviors, environmental exposures, socioeconomic patterns, and healthcare access - interact over time to govern a person's individual health trajectory. Collectively, this array of influential factors is often called the "social determinants of health," and operationalizing an understanding of health that captures the multifactorial context of social determinants will enable more nuanced and powerful testable models rooted in broader evidence bases, which may also guide management strategies in a more systemic manner. Big data can empower us to predict the future of individual track records of these factors as well as the outcomes of that aggregate interaction. However, there are also significant ethical, legal, and social issues that must be addressed in order for these developments to benefit the population. The potential benefits are available only to individuals who have access to, can understand, and - if applicable - agree to the use of their digital trails' predictive power [11, 12].

METHODS AND TECHNIQUES FOR ANALYZING BIG DATA IN HEALTHCARE

The unique challenges and characteristics of healthcare big data require equally specific and tailored methods of analysis. Techniques like machine learning, data mining, and predictive modeling have been widely used to predict health outcomes based on healthcare big data. Data mining, a process that uses a variety of data analysis tools to discover patterns and relationships in data, has been widely used to analyze healthcare big data. In predictive modeling, classification, regression, clustering, and outlier detection methods are used to solve prediction problems. In recent years, machine learning methods have also gained an increasing role in the use of healthcare big data. Ensemble models, such as bagging and boosting, have been widely used to analyze healthcare big data problems [3, 13]. Techniques used to analyze big data require considerable resources, specialized knowledge, and time. Furthermore, the training and evaluation of machine learning models - Cross Validation in particular - impede their use in real-time scenario-based healthcare settings. Therefore, hybrid approaches, such as semi-automated learning, might show superior strengths. Additionally, consider that research problems in big data predictions are naturally complex, having high variance and bias and being considerably heterogeneous among individuals and organizations. Probably due to these complexities, most of the technology currently remains implemented for research purposes, within university settings and medical institutions. Some guidelines and recommendations have already been advised for the best performance and use of big data tools and procedures for real-world implementation analysis [3, 14].

CASE STUDIES AND APPLICATIONS OF BIG DATA IN PREDICTING HEALTH OUTCOMES

Using omic data and machine learning methods, we can predict disease outcomes, identify comorbidities, and reclassify complex phenotypes. We present a case study in prediabetes, where routine testing may aid

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in the prediction of health outcomes. A study relevant to clinical practice was developed by a physician who worked with predictive models of disease in patients with an inherited bone marrow failure syndrome. Genetic testing is not routine in these patients, as most of them do not have a clear mutation in current genes known to be involved in inherited bone marrow failure; however, the presence of a monosomy 7 in bone marrow cells might change the recommendation for clinical care (e.g. hematopoietic cell transplant versus drug therapy). Recently, big data analytics has also been utilized for prediction of optimal liposomal amphotericin B doses used in treatment of pediatric patients with cancer; one formulation is calculative and the other one is numerative. This research leveraged model stacking and LSTM models and assessed patients' whole blood gene expression and 30 immune analytes' responses to in vitro stimulation with either liposomal amphotericin B or endotoxin to approximate the child's critical illness (i.e., sepsis or septic shock) phenotype at time of clinical presentation. Furthermore, this work predicted the inherent production of IL-6 before endotoxin exposure and will contribute to the precision dosing of liposomal amphotericin B in children to improve their health. And finally, multiple types of 'big data' were used in a proof of concept project to predict complications in horses. Clinical pathologic data were sufficient to make good predictions and the addition of further data (machine learning feature) did not seem to improve accuracy. Overall, these case studies give insight into the use of big data in predicting health outcomes, with the setting ranging from problems that impact a broad range of patient populations down to specific clinical situations where "n = 1" predictions become evident [3, 15]. An example in which this relationship between hospital and community resources is used to predict outcomes is in Nassiri and colleagues. To predict which patients are at risk for opioid abuse hospital-wide and the likelihood by service, they used the social determinants from both [16, 17, 18, 19]. Further, similar to how Social Indicator Reports allow us to identify the county with a high incidence of diabetes and rank them for community intervention programs, we took ICD-10 codes and better related them to community data, showing for example that low adherence to screenings for diabetic retinopathy can be associated with characteristics of the population with prediabetes in your area. Another high level deduction I've made using community data is that we can use this to identify, for example, where additional hospital ED will be beneficial, while identifying what attributes they should emphasize in the ED settings for maximum effectiveness [16, 17, 18, 19].

CONCLUSION

Big data holds immense potential to transform healthcare by enabling more accurate and timely predictions of health outcomes. From improving resource allocation in hospitals to enhancing personalized care for patients, its applications are far-reaching. However, realizing this potential requires overcoming challenges related to data privacy, computational resources, and the complexity of analyzing heterogeneous datasets. Effective integration of big data into healthcare will demand a collaborative approach that addresses these technical, ethical, and practical challenges. As big data continues to evolve, its role in healthcare will become increasingly vital, providing the foundation for improved patient outcomes and a more efficient healthcare system.

REFERENCES

1. Shafqat S, Kishwer S, Rasool RU, Qadir J, Amjad T, Ahmad HF. Big data analytics enhanced healthcare systems: a review. *The Journal of Supercomputing*. 2020 Mar;76:1754-99. academia.edu
2. Son YS, Kwon KH. Utilization of smart devices and the evolution of customized healthcare services focusing on big data: a systematic review. *Mhealth*. 2024. nih.gov
3. Guo C, Chen J. Big data analytics in healthcare. In *Knowledge technology and systems: Toward establishing knowledge systems science 2023 Jun 14 (pp. 27-70)*. Singapore: Springer Nature Singapore. vt.edu
4. Zheng K, Cai S, Chua HR, Herschel M, Zhang M, Ooi BC. DyHealth: making neural networks dynamic for effective healthcare analytics. *Proceedings of the VLDB Endowment*. 2022 Aug 1;15(12):3445-58. ird.fr
5. Bertsimas D, Boussioux L, Cory-Wright R, Delarue A, Digalakis V, Jacquillat A, Kitane DL, Lukin G, Li M, Mingardi L, Nohadani O. From predictions to prescriptions: A data-driven response to COVID-19. *Health care management science*. 2021 Jun;24:253-72. springer.com
6. Stone K, Zwigelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: Towards a unified framework. *PLOS Digital Health*. 2022 Apr 14;1(4):e0000017. plos.org

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

7. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. *Nature medicine*. 2020. exlibrisgroup.com
8. Nathan R, Monk CT, Arlinghaus R, Adam T, Alós J, Assaf M, Baktoft H, Beardsworth CE, Bertram MG, Bijleveld AI, Brodin T. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science*. 2022 Feb 18;375(6582):eabg1780. rug.nl
9. Norori N, Hu Q, Aellen FM, Faraci FD et al. Addressing bias in big data and AI for health care: A call for open science. *Patterns*. 2021. cell.com
10. Maier-Hein L, Eisenmann M, Sarikaya D, März K, Collins T, Malpani A, Fallert J, Feussner H, Giannarou S, Mascagni P, Nakawala H. Surgical data science—from concepts toward clinical translation. *Medical image analysis*. 2022 Feb 1;76:102306. sciencedirect.com
11. Rehman A, Naz S, Razzak I. Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. *Multimedia Systems*. 2022. [[PDF](#)]
12. Huang H, Yao XA, Krisp JM, Jiang B. Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Computers, Environment and Urban Systems*. 2021 Nov 1;90:101712. researchgate.net
13. Khan ZF, Alotaibi SR. Applications of artificial intelligence and big data analytics in m-health: A healthcare system perspective. *Journal of healthcare engineering*. 2020. wiley.com
14. Mohamed A, Najafabadi MK, Wah YB, Zaman EA, Maskat R. The state of the art and taxonomy of big data analytics: view from new big data framework. *Artificial intelligence review*. 2020 Feb;53:989-1037. [[HTML](#)]
15. Dlamini Z, Francies FZ, Hull R, Marima R. Artificial intelligence (AI) and big data in cancer and precision oncology. *Computational and structural biotechnology journal*. 2020 Jan 1;18:2300-11. sciencedirect.com
16. Bakhtiari A, Pasha H, Kashefi F, Nasiri-Amiri F, Bakouei F. Factors affecting students' attitudes towards reproductive health in the north of Iran: Designing an educational program. *BMC Public Health*. 2023 Aug 16;23(1):1557. springer.com
17. Khalili M, Mirzazadeh A, Chegeni M, Abedi L, Rajaei L, Ardalan G, Haghdoost A, Nasiri N, Sharifi H. Prevalence of high-risk sexual behavior among Iranian young people: A systematic review and meta-analysis. *Children and Youth Services Review*. 2020 Dec 1;119:105526. [[HTML](#)]
18. Joseph Obiezu Chukwujekwu Ezeonwumelu, Uhama Kingsley Chukwuka, Ugwu Okechukwu Paul-Chima, Alum Esther Ugo, Ugwuanyi Anthony Chukwudi and Tambwe Patrick Rodrigue (2024). The Impact of Artificial Intelligence and Machine Learning on Pharmacy Practice. *RESEARCH INVENTION JOURNAL OF RESEARCH IN MEDICAL SCIENCES* 3(1):10-15.
19. Ugwu Okechukwu Paul-Chima, Alum Esther Ugo, Okon Michael Ben, Egba Simeon Ikechukwu and Uti Daniel Ejim. Integrating Wearable Health Monitoring Devices with IoT for Enhanced Personal Health Management: A Comprehensive Review. *Research Output Journal of Engineering and Scientific Research*. 2024 3(1):6-10.

CITATION: Mwendu Muthoni D. The Role of Big Data in Predicting Health Outcomes. Research Output Journal of Biological and Applied Science, 2024 3(3):36-39.