International Journal of Advances in Medicine
Laishram IS. Int J Adv Med. 2022 Dec;9(12):1239-1241
http://www.ijmedicine.com

pISSN 2349-3925 | eISSN 2349-3933

DOI: https://dx.doi.org/10.18203/2349-3933.ijam20223030

Review Article

Data bias in precision medicine

Indira Singha Laishram*

Westcliff University, Irvine, California, USA

Received: 18 October 2022
Accepted: 14 November 2022

***Correspondence:**
Dr. Indira Singha Laishram,
E-mail: i.laishram.239@westcliff.edu

ABSTRACT

Precision medicine is poised to increasingly improve health outcomes for more people in the near future. In contrast to the more traditional reactive methods of disease treatment, precision medicine is a customizable treatment and disease prevention approach that is tailored for the individual. Artificial intelligence (AI) using sophisticated algorithms and machine learning (ML) tools powers these precision medicine processes. These algorithms analyze big data collected from multiple sources over the past decades to aid physicians to make data-backed critical clinical decisions. However, studies have shown that unintentional biases in the source data and in the process can affect these precision medicine efforts.

Keywords: Precision medicine, AI, ML, Data bias

INTRODUCTION

In contrast to a more traditional one-size-fits-all symptom-driven approach, precision medicine has prospects of customization of treatment for patients.[1,2] Integration of the enormous amount of health data accumulated over the past century, environment and lifestyle makes this possible.[3,4] Big data analytics can examine patient health data, data from the omics-genomics, transcriptomics, proteomics, pharmacogenomics[5], and possibly data from the microbiome[6] to inform clinical decision making.[7] Furthermore, the incorporation of data from the omics can extend the benefits of precision medicine from treatment to disease prevention and timely diagnosis.[8]

Big data analytics uses techniques and tools based on data mining, web mining, statistical analytics, etc., to derive health insights from the information contained in the data.[9] Analytics of big data from patients as well as healthy individuals integrates bioinformatics, medical imaging, sensor informatics, medical informatics as well as the health informatics. Much of this data is complex, heterogenous, structured as well as the semi-structured or unstructured.[10]

Analytics of medical big data helps identify clusters and correlation between datasets, and develop predictive models to provide robust data-backed decision-making capabilities. The processing of this big data needs powerful computing capacity that employs cloud computing with powerful multi core central processing units (CPUs), graphics processing units (GPUs), field-programmable gate arrays (FPGAs) with parallel processing systems thus laying the ground for AI and self-learning algorithms for ML.[11]

CHARACTERISTICS OF BIG DATA

There is no consensus on the definition of big data. It was earlier described as data exhibiting three Vs (Volume-referring to the size of data in terabytes, Velocity-referring to the speed of collection and processing, and variety-referring to the types of this data including structured, unstructured, heterogenous, etc.). Over time other factors have been included to now total to 6Vs. The new Vs being Variability – referring to seasonal changes in health and progression of disease, and veracity-referring to the reliability of the data.[12,13] As one of the Vs other researchers have value-referring to the relevance of the data in clinical decision-making virtuosity-to ensure the foundations of equity and justice in each of the Vs.[11,14] The characteristics of big data are shown in Table 1.

Table 1: Definitions of big data characteristics.

Definitions	Data characteristics
Original definitions	
Volume	The amount of data
Velocity	Speed of collection and processing
Variety	Types of data including structured, unstructured, heterogenous, etc.
Recent additions	
Variability	Referring to seasonal changes in health and progression of disease
Veracity	Reliability of the data
Virtuosity	Equity and justice in each of the Vs

BIAS IN AI

Successful adoption of AI in medicine depends on the trust patients and physicians have on the AI systems' training algorithms to produce reliable output. Yet it is the output that is sometimes unreliable due to algorithmic bias[15] and bias in the dataset.[16] Given the consequence that an incorrect medical decision can have, it is critical that medical AI algorithms are unbiased and are trained with unbiased data.

BIAS IN DATA

The AI system's output is shaped by the data that it learns from. The most common data bias is seen where the data does not correctly represent the population. For example, AI algorithms that have trained on data from skin images of mostly white candidates frequently misreport readings from darker skinned patients. A study reported potential racial bias in pulse oximetry measurements placing black patients at a higher risk of occult hypoxia.[17]

An oximeter is a device that detects oxygen-saturation of hemoglobin in the blood by reading the color of the blood. This it does by shining an infra-red and a red light through a finger. Studies have found an overestimation of saturation levels for non-white people thus leading to adverse patient outcomes in this group of patients.[18]

Convoluted neural networks (CNNs) that are able to classify skin lesions with high accuracy are frequently trained with sample images from white individuals where an estimated 5%- 10% of the datasets are samples from black patients. In the clinical setting, this can lead to misdiagnoses by these algorithms and exacerbate an already higher mortality rate from skin cancer among black patients.[16]

Yet another example of data bias is seen in data sets in oncology where patients with superior performance status in clinical trials would have an over-representation or, patients with restricted access to care having an under-representation thus leading to bias due to class imbalance.[19]

ALGORITHMIC BIAS

Algorithmic bias is one that exacerbates prevailing inequities in socioeconomic status, race, ethnic background, gender, etc., and further intensifies them. The bias in precision medicine introduced by bias in data on which algorithms train and learn have been discussed here. However, there are biases due to algorithms themselves. For example, algorithms have been trained to predict healthcare costs rather than disease. However, health spending of black patients with the same illness is lesser than that of white patients. The algorithm erroneously infers black patients to be healthier than just as sick white patients.[19]

DISCUSSION

The hope of precision medicine is that it will deliver customized medication to patients and reduce errors in treatment.[2] However, precision medicine depends on the data that its AI tools have trained on. It is thought that any existing bias found in society will be found in the data that is collected from it. This biased data in turn will be fed into ML algorithms thus manifesting in the outputs used in clinical decision-making. It is imperative that all stakeholders help to address inaccuracies in these high-stakes decision-making processes. Through these algorithms, precision medicine promises fewer clinical errors with increasing positive patient outcomes. This promise can be delivered to individuals of all races and socio-economic status by removing the biases inherent in the source data and by developing improved algorithms. It has been reported that over 80% of genomics data sets are made up of individuals from European ancestry.[20] This poses a serious data diversity and representation issue in precision medicine.[21] Studies have shown that mitigation methods for racial bias in data can increase fairness, avoid inequities in diagnosis and medical care.[22]

CONCLUSION

With increased data accessibility and enormous improvements in computational capacity, deep learning models are already impacting precision medicine. However, bias in data raises serious concerns due to the negative impact it may have in clinical decision making. AI decision support tools that have trained on biased data pose health risks to patients. Such data informed clinical errors can be mitigated by removing the bias in data. Revisiting the source of data and removing fields that could cause bias and building fair algorithms that can evaluate data fairness could be the first steps in advancing the field of precision medicine in the right direction.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Ashley EA. Towards precision medicine. Nature Rev Genetics. 2016;17(9):507-22.
2. Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. BMC Med Informatics Decision Making. 2018;18(1):1-15.
3. Schaefer GO, Tai ES, Sun S. Precision Medicine and Big Data. Asian Bioethics Rev. 2019;11(3):275-88.
4. Mesko B. The role of artificial intelligence in precision medicine. Expert Rev Precision Med Drug Develop. 2017;2(5):239-41.
5. Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The need for multi-omics biomarker signatures in precision medicine. Int J Molecular Sci. 2019;20(19):4781.
6. Cammarota G, Ianiro G, Ahern A, Carbone C, Temko A, Claesson MJ et al. Gut microbiome, big data and machine learning to promote precision medicine for cancer. Nature Rev Gastroenterol Hepatol. 2020;17(10):635-48.
7. Leff DR, Yang GZ. Big data for precision medicine. Engineering. 2015;1(3):277-9.
8. Chen R, Snyder M. Promise of personalized omics to precision medicine. Wiley Interdisciplinary Rev: Systems Biol Med. 2013;5(1):73-82.
9. Kornelia B, Ślęzak A. The use of Big Data Analytics in healthcare. J Big Data. 2022;9(1).
10. Cirillo D, Valencia A. Big data analytics for personalized medicine. Curr Opinion Biotechnol. 2022;58:161-7.
11. Ristevski B, Chen M. Big data analytics in medicine and healthcare. J Integrative Bioinformatics. 2018;15(3).
12. Mayer-Schönberger V, Ingelsson E. Big Data and medicine: a big deal? J Internal Med. 2018;283(5):418-29.
13. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE J Biomed Heal Informatics. 2018;19(4):1193-208.
14. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Heal. 2019;9(2).
15. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns. 2021;2(10):100347.
16. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Eng J Med. 2020;383(25):2477-8.
17. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. J Perinatol. 2022;42(1):79-85.
18. Tasci E, Zhuge Y, Camphausen K, Krauze AV. Bias and Class Imbalance in Oncologic Data-Towards Inclusive and Transferrable AI in Large Scale Oncology Data Sets. Cancers. 2022;14(12):2897.
19. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53.
20. Martin AR, Stroud RE II, Abebe T, Akena D, Alemayehu M, Atwoli L et al. Increasing diversity in genomics requires investment in equitable partnerships and capacity building. Nature Genetics. 2022;54(6):740-5.
21. Lee SS, Fullerton SM, McMahon CE, Bentz M, Saperstein A, Jeske M et al. Targeting Representation: Interpreting Calls for Diversity in Precision Medicine Research. Yale J Biol Med. 2022;95(3):317-26.
22. Huang J, Galal G, Etemadi M, Vaidyanathan M. Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. JMIR Med Informatics. 2022;10(5).

Cite this article as: Laishram IS. Data bias in precision medicine. Int J Adv Med 2022;9:1239-41.