Understanding Bias in Artificial Intelligence Algorithms Intended for Clinical Care

ANIKA ACHARI
University of Calgary

BACKGROUND

Artificial intelligence (AI) based tools utilize the prowess of smart algorithms and fast computational abilities to obtain meaningful outcomes from large datasets (Norori et al., 2021). Clinical data, which is routinely collected in healthcare settings forms a huge repository of knowledge (Norori et al., 2021). Solutions to a myriad of medical problems exist within the folds of such large and complex datasets (Réda et al., 2020). To access the information within these large datasets, AI based tools are being developed and deployed. Patterns that exist within these intricate datasets are not easily discernable to humans and can only be revealed through intelligently crafted machine learning (ML) algorithms.

ML algorithms utilize statistical methods to make predictions using large databases (Norori et al., 2021). Data is continuously added to these large databases which increases the complexity of recognizing patterns using existing algorithms. A major shortcoming of clinical datasets is that the data may be erroneous or have meaningless entries, leading to what is termed as 'noisy' data (Norori et al., 2021). In addition to the aforementioned shortcomings, it is imperative to understand that the data collection process plays a significant role in the introduction of bias in AI algorithms. Racial and social inequalities lead to disparities in healthcare settings, which subsequently impacts the data that is collected for the development of AI algorithms. In this paper we aim to address: (a) existing biases in healthcare, (b) how bias has impacted the development of AI algorithms, and (c) whether it is practical to eliminate inherent racial and socioeconomic differences when developing AI algorithms for clinical care.

EXISTING BIASES IN THE HEALTHCARE SYSTEM

Healthcare practitioners are expected to provide equal and unbiased treatment to all patients based on their clinical needs. However, implicit, and explicit biases exist and play a role when it comes to providing clinical care to patients. Factors such as race, sex and gender, sexual identity, socio-economic status, education level, physical appearance, disability, and age may play a role in the prejudice experienced in healthcare settings (Norori et al., 2021; Dovidio et al., 2016). A study reported that black men were less

likely to be recommended for coronary by-pass surgery than their white counterparts (Dovidio et al., 2016). The physician's decision was influenced by the perception that black men were less educated and therefore, were less likely to follow the required physical activity guidelines following surgical intervention (Dovidio et al., 2016). Black and white racial disparities have also been recorded in the context of prescribing pain management medication (Burgess et al., 2014). Reportedly, black patients are prescribed lower doses of pain medication (Burgess et al., 2014) and this differential treatment has also been observed for black children seeking similar remedies (Goyal et al., 2015).

While the prejudice harbored by medical practitioners impacts the therapeutic care received, the discrepancy in socio-economic status (SES) also plays an undeniable role in the access to healthcare resources. Currently, novel practices in health are easily communicated to the masses through digital technologies (Davenport & Kalakota, 2019). However, it should be noted that digital technologies for health purposes are mostly used by educated, young, and urban patients. Internet access may not be reasonably priced and may not even be available in rural areas (Tieu et al., 2015). Hence, SES is a primary determinant in owning digital tools such as cellular phones and computers. In addition to owning such devices, it is also essential to utilize these devices to gather health-related information and education (Tieu et al., 2015). During the COVID-19 pandemic, healthcare delivery experienced a rapid progression to digital health technologies (Obermeyer et al., 2019). Despite the advantages conferred by these technologies, it was observed that low SES populations lacked the same level of access to such technologies compared to their high SES counterparts (Obermeyer et al., 2019). Low SES populations consisted of patients with low English-language proficiency, less education, and low-income backgrounds (Obermeyer et al., 2019). In summary, either people do not have access to digital devices or do not know how to access information regarding the medical issues that plague them.

BIAS IN AI ALGORITHMS
Ideally, the use of advanced AI algorithms in medicine for diagnosis should lead to bias-free clinical care, as it is hoped that the power of artificial intelligence will transcend the prejudices of humans. However, biases have been introduced into AI algorithms as a result of the implicit prejudice that remains rooted in the cognitive behavior of humans. The primary reason for the introduction of this bias is that humans are responsible for the development of these algorithms. Prejudiced assumptions can be made during the algorithm development process where developers may unconsciously or consciously introduce the bias.

AI/ML algorithms can help to interpret the data obtained from diagnostics tests and offer advice on accurate clinical interventions for all patients equally without racial bias. However, a recent study reported that algorithms assigned a lower risk score to black people than to white people,

although both demographics faced similar medical issues (Obermeyer et al., 2019). As a result, black people were less likely to be referred to the programs that provide more-personalized care. The algorithm that resulted in this erroneous outcome was designed to consider the healthcare costs accrued by the patient over a period of one year (Obermeyer et al., 2019). The algorithms took a simplistic approach of associating healthcare costs with healthcare needs, and it failed to account for the fact that black people generally face more chronic issues such as diabetes, anaemia, kidney failure and high blood pressure (Obermeyer et al., 2019). Further investigation into hospital data revealed that the care provided to black people cost an average of US $1,800 less per year than the care given to a white person with the same number of chronic health problems (Obermeyer et al., 2019). The reasons for this discrepancy were attributed to either distrust of the healthcare system and/or direct racial discrimination by healthcare providers (Obermeyer et al., 2019). The serious consequence of this bias was that in general black people had to be sicker than white people before being transferred for supplementary medical care (Obermeyer et al., 2019).

Furthermore, AI algorithms rely heavily on good, clean data to train the system (Norori et al., 2021). AI algorithms are trained to utilize features from the dataset and find patterns (Norori et al., 2021). If a particular dataset contains bias, then the algorithm will learn to utilize this biased data and generate outcomes (Norori et al., 2021). For example, the most significant challenge that plagues large clinical datasets is that they do not always represent the diversity that exists in the human population (Tarver, 2021). Clinical datasets primarily represent the white populations and there is a scarcity observed in the enrollment of African American or even Latino communities (Tarver, 2021; Oh et al., 2015). Many African Americans have a deep distrust for the medical community partly due to the Tuskegee Syphilis Study which is a prime example of systemic racism that is pervasive in healthcare settings across the United States of America (Tarver, 2021). Although awareness regarding the unethical use of select demographics in clinical research has increased following the Tuskegee Syphilis Study, a fear of participating in medical research still exists within the African American community (Tarver, 2021). In 1989, when Los Angeles was experiencing a measles epidemic, a new vaccine was introduced and tested on 900 Latino and African American babies. Parents were not informed that this new vaccine was not yet approved for use in the United States by the Centers for Disease Control and Prevention (Tarver, 2021; Cimons, 2019).

Ensuring diversity during the data collection process remains difficult, given that people may not always be willing to participate in research studies for a number of reasons. For example, it has been reported that less educated older African American men are less likely to participate in clinical research (Webb et al., 2019), possibly due to their personal adverse experiences or their awareness of the historical exploitation of certain minorities in healthcare settings (Webb et al., 2019). Alternatively,

if African Americans have participated in a research study in the past, they are likely to do so again (Webb et al., 2019). Their willingness to participate again, may be the result of positive experiences they have had with the healthcare system and increased awareness of the importance of maintaining overall health (Webb et al., 2019). It has been suggested that increasing the enrollment of African Americans in clinical research will require unique tactics such as a higher compensation for participation, increased emphasis on trust-building, and utilizing easy-accessible venues (Webb et al., 2019).

The aforementioned lack of participation from African American and other ethnic minorities in clinical research has led to a lack of diversity in datasets. This lack of diversity means that algorithms are being trained on datasets that only represent a small cohort of the human population and therefore, the AI outcomes may be biased (Norori et al., 2021). Application of such AI analysis for diagnosis in clinical settings may be pertinent only for a small cohort of human populations, although they may be erroneously applied to all groups. For example, patients of lower socioeconomic strata may either access the healthcare system less or may receive fewer diagnostic tests and medications for chronic diseases due to innate bias (Signorello et al., 2014; Ejike et al., 2021; Ledford, 2019). Precision medicine recognizes that individuals may differ in their response to medication based on their genetic make-up, physiology, social, economic, and environmental factors. A combination of fewer diagnostic tests and a lack of participation in research-based initiatives, will result in insufficient health-related data for a certain group of individuals. This compromises the ability to provide precision health solutions that are necessary, considering the unique make-up of every individual.

IS IT REALISTIC TO COMPLETELY ELIMINATE DIVISIVE FACTORS FROM AI ALGORITHMS?

As we continue developing AI algorithms to improve clinical care, it is important to be cognizant that bias has a strong potential to adversely impact the outcomes generated by these algorithms. We are moving towards an age in which healthcare solutions are dominated by precision medicine, which recognizes individual differences and considers them when providing treatment. To be able to accurately predict treatment, it is imperative that AI algorithms for clinical care be trained with data which exhibits the genetic variety that exists in the human population which is further complicated by social, economic, and demographic factors.

Given the existence of variability in human populations, it would be an overly simplistic approach if we were to remove questions regarding age, gender, and ethnicity from the data collection process in order to address bias. In fact, if we are to progress towards finding solutions for the different diseases and chronic health conditions that plague humans, it is important to understand and consciously include diverse populations when developing datasets that are used to train AI algorithms. While bias is extremely

harmful in clinical care settings, it is important that we remain mindful of the inherent diversity that forms the core of human populations and work towards finding solutions that are uniquely suitable to each individual.

References

Burgess, D. J., Nelson, D. R., Gravely, A., Bair, M. J., Kerns, R. D., Higgins, D. M., Van Ryn, M., Farmer, M. M., & Partin, M. R. (2014). Racial Differences in Prescription of Opioid Analgesics for Chronic Noncancer Pain in a National Sample of Veterans. *The Journal of Pain*, *15*(4), 447–455. https://doi.org/10.1016/j.jpain.2013.12.010

Cimons, M. (2019, March 5). CDC Says It Erred in Measles Study - Los Angeles Times. *Los Angeles Times*. https://www.latimes.com/archives/la-xpm-1996-06-17-mn-15871-story.html#:~:text=A%20government%2Dsponsored%20study%20of,federal%20health%20officials%20said%20Sunday.

Davenport, T. H., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Dovidio, J. F., Eggly, S., Albrecht, T. L., Hagiwara, N., & Penner, L. A. (2016). Racial biases in medicine and healthcare disparities. *TPM - Testing, Psychometrics, Methodology in Applied Psychology*, *23*(4), 489–510. https://doi.org/10.4473/tpm23.4.5

Ejike, C., Woo, H. Y., Galiatsatos, P., Paulin, L. M., Krishnan, J. A., Cooper, C. B., Couper, D., Kanner, R. E., Bowler, R. P., Hoffman, E. A., Comellas, A. P., Criner, G. J., Barr, R. G., Martinez, F. J., Han, M. K., Martinez, C. a. R., Ortega, V. E., Parekh, T. M., Christenson, S. A., . . . Hansel, N. N. (2021). Contribution of Individual and Neighborhood Factors to Racial Disparities in Respiratory Outcomes. *American Journal of Respiratory and Critical Care Medicine*, *203*(8), 987–997. https://doi.org/10.1164/rccm.202002-0253oc

Goyal, M. K., Kuppermann, N., Cleary, S. P., Teach, S. J., & Chamberlain, J. M. (2015). Racial Disparities in Pain Management of Children With Appendicitis in Emergency Departments. *JAMA Pediatrics*, *169*(11), 996. https://doi.org/10.1001/jamapediatrics.2015.1915

Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. *Nature*, *574*(7780), 608–609. https://doi.org/10.1038/d41586-019-03228-6

Norori, N., Qiyang, H., Aellen, F. M., Faraci, F., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

Obermeyer, Z., Powers, B. W., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Oh, S. S., Galanter, J., Thakur, N., Pino-Yanes, M., Barcelo, N. E., White, M. J., De Bruin, D. M., Greenblatt, R. M., Bibbins-Domingo, K.,

Wu, A. H., Borrell, L. N., Gunter, C., Powe, N. R., & Burchard, E. G. (2015). Diversity in Clinical and Biomedical Research: A Promise Yet to Be Fulfilled. *PLOS Medicine*, *12*(12), e1001918. https://doi.org/10.1371/journal.pmed.1001918

Réda, C., Kaufmann, E., & Delahaye-Duriez, A. (2020). Machine learning applications in drug development. *Computational and Structural Biotechnology Journal*, *18*, 241–252. https://doi.org/10.1016/j.csbj.2019.12.006

Signorello, L. B., Cohen, S. S., Williams, D. R., Munro, H. M., Hargreaves, M. K., & Blot, W. J. (2014). Socioeconomic Status, Race, and Mortality: A Prospective Cohort Study. *American Journal of Public Health*, *104*(12), e98–e107. https://doi.org/10.2105/ajph.2014.302156

Tarver, M. E. (2021). Race and Ethnicity in Real-World Data Sources: Considerations for Medical Device Regulatory Efforts. *Journal of Primary Care & Community Health*, *12*, 215013272199404. https://doi.org/10.1177/2150132721994040

Tieu, L., Sarkar, U., Schillinger, D., Ralston, J. D., Ratanawongsa, N., Pasick, R. J., & Lyles, C. R. (2015). Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. *Journal of Medical Internet Research*, *17*(12), e275. https://doi.org/10.2196/jmir.4847

Webb, F. J., Khubchandani, J., Striley, C. W., & Cottler, L. B. (2018). Correction to: Black–White Differences in Willingness to Participate and Perceptions About Health Research: Results from the Population-Based HealthStreet Study. *Journal of Immigrant and Minority Health*, *21*(2), 306. https://doi.org/10.1007/s10903-018-0747-0