BMJ Health & Care Informatics

Operationalising fairness in medical algorithms

Sonali Parbhoo,[1] Judy Wawira Gichoya,[2] Leo Anthony Celi ,[3,4] Miguel Ángel Armengol de la Hoz ,[5] for MIT Critical Data

To cite: Parbhoo S, Wawira Gichoya J, Celi LA, *et al*. Operationalising fairness in medical algorithms. *BMJ Health Care Inform* 2022;**29**:e100617. doi:10.1136/bmjhci-2022-100617

Received 20 May 2022
Accepted 24 May 2022

[1]Harvard Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA
[2]Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia, USA
[3]Laboratory for Computational Physiology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts, USA
[4]Division of Pulmonary Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[5]Big Data Department, Regional Ministry of Health of Southern Spain, Sevilla, Spain

Correspondence to
Dr Leo Anthony Celi;
LCeli@mit.edu

The world is abuzz with applications of machine learning and data science in almost every field: commerce, transportation, banking, and more recently, healthcare. Breakthroughs in these areas are a result of newly created algorithms, improved computing power and, most importantly, the availability of bigger and increasingly reliable data with which to train these algorithms. For healthcare specifically, machine learning is at the juncture of moving from the pages of conference proceedings to clinical implementation at the bedside. Yet, succeeding in this endeavour requires synthesising insights from both the algorithmic perspective as well as the healthcare domain to ensure that the unique characteristics of machine learning methods can be leveraged to maximise benefits and minimise risks.

While progress has recently been made in establishing certain guidelines or best practices for the development of machine learning models for healthcare as well as protocols for the regulation of such models, these guidelines and protocols tend to overlook important considerations such as fairness, bias and unintended disparate impact.[1 2] Nevertheless, it is widely recognised in other domains that many of the machine learning models and tools may have discriminatory effect by inadvertently encoding and perpetuating societal biases.[3]

In this special issue, we highlight that machine learning algorithms should not be focused solely on accuracy but should be evaluated with respect to how they might impact disparities in patient outcomes. Our special issue aims to bring together the growing community of healthcare practitioners, social scientists, policymakers, engineers and computer scientists to design and discuss practical solutions to address algorithmic fairness and accountability. We invited papers that explore ways to reduce machine learning bias in healthcare or explain how to create algorithms that specifically alleviate inequalities.

To prevent artificial intelligence (AI) from encoding the disparities that exist, algorithms should predict an outcome as if the world were fair. If designed well, AI may even provide a way to audit and improve the way care is being delivered across populations. There is growing community momentum towards not just detecting bias but operationalising fairness, but this is a monumental task. Some of the encouraging developments that we have seen have been incorporating patients' voices in AI. Patient engagement is crucial if algorithms are to truly benefit everyone.

The papers in this special issue cover a variety of topics that addressed the objectives laid out in the call, these were:

- Identifying Undercompensated Groups Defined by Multiple Attributes in Risk Adjustment[4]
- A Proposal for Developing a Platform That Evaluates Algorithmic Equity and Accuracy[5]
- Can medical algorithms be fair? Three ethical quandaries and one dilemma[6]
- Resampling to Address Inequities in Predictive Modeling of Suicide Deaths[7]
- Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation[8]
- Operationalizing fairness in medical AI adoption: Detection of early Alzheimer's Disease with 2D CNN[9]
- Global disparity bias in ophthalmology artificial intelligence applications[10]
- Investigating for bias in healthcare algorithms: A sex stratified analysis of supervised machine learning models in liver disease prediction[11]

It has been more than 5 years since the ProPublica investigative report on machine bias was published. The report detailed how a software used in judicial courts across the

USA to inform decisions around parole was prejudiced against black people. Everything we have achieved since then has always been geared towards understanding how difficult it is to prevent AI from perpetuating societal biases in algorithms.

There is a long road ahead before we can leverage the zettabytes of data that are routinely collected in the process of care. We should not only invest in storage and compute technologies, federated learning platforms, GPTs, GRUs and NFTs. Machine learning in healthcare is not just about predicting something for the sake of prediction. The most important task is to augment our capacity to make decisions, and that requires understanding how those decisions are made.

Twitter Judy Wawira Gichoya @judywawira, Leo Anthony Celi @MITCriticalData and Miguel Ángel Armengol de la Hoz @miguearmengol

Contributors Initial conceptions and design—SP, JWG, LAC and MAAdlH. Drafting of the paper—SP, JWG, LAC and MAAdlH. Critical revision of the paper for important intellectual content—SP, JWG, LAC and MAAdlH.

Funding LAC is funded by the National Institute of Health through the NIBIB R01 EB017205.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; internally peer reviewed.

ORCID iDs
Leo Anthony Celi http://orcid.org/0000-0001-6712-6626
Miguel Ángel Armengol de la Hoz http://orcid.org/0000-0002-7012-2973

REFERENCES

1 Wawira Gichoya J, McCoy LG, Celi LA, *et al*. Equity in essence: a call for operationalising fairness in machine learning for healthcare. *BMJ Health Care Inform* 2021;28:e100289.
2 McCoy LG, Banja JD, Ghassemi M, *et al*. Ensuring machine learning for healthcare works for all. *BMJ Health Care Inform* 2020;27:e100237.
3 Sarkar R, Martin C, Mattie H, *et al*. Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. *Lancet Digit Health* 2021;3:e241–9.
4 Zink A, Rose S. Identifying undercompensated groups defined by multiple attributes in risk adjustment. *BMJ Health Care Inform* 2021;28:e100414.
5 Cerrato P, Halamka J, Pencina M. A proposal for developing a platform that evaluates algorithmic equity and accuracy. *BMJ Health Care Inform* 2022;29:e100423.
6 Bærøe K, Gundersen T, Henden E, *et al*. Can medical algorithms be fair? three ethical quandaries and one dilemma. *BMJ Health Care Inform* 2022;29:e100445.
7 Reeves M, Bhat HS, Goldman-Mellor S. Resampling to address inequities in predictive modeling of suicide deaths. *BMJ Health Care Inform* 2022;29:e100456.
8 Foryciarz A, Pfohl SR, Patel B, *et al*. Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation. *BMJ Health Care Inform* 2022;29:e100460.
9 Heising L, Angelopoulos S. Operationalising fairness in medical AI adoption: detection of early Alzheimer's disease with 2D CNN. *BMJ Health Care Inform* 2022;29.
10 Nakayama LF, Kras A, Ribeiro LZ, *et al*. Global disparity bias in ophthalmology artificial intelligence applications. *BMJ Health Care Inform* 2022;29:e100470.
11 Straw I, Wu H. Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. *BMJ Health Care Inform* 2022;29.