01462

Recognising the inherent diversity in patients with infections and the challenge of fairness in AI systems

11. Other

11d. Digital health and infection (incl AI, data mining, informatics)

William Bolton [1], Timothy Rawson [1], Richard Wilson [1], Bernard Hernandez [1], David Antcliffe [2, 3], Pantelis Georgiou [1, 4], Alison Holmes [1, 5, 6]

[1]Centre for Antimicrobial Optimisation, Imperial College London - London (United Kingdom), [2]Department of Critical Care, Imperial College Healthcare NHS Trust - London (United Kingdom), [3]Faculty of Medicine, Imperial College London - London (United Kingdom), [4]Centre for Bio-inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London - London (United Kingdom), [5]National Institute for Health Research, Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London - London (United Kingdom), [6]Department of Infectious Diseases, Imperial College London - London (United Kingdom)

Background Research has shown a strong association between sensitive attributes (those not linkable or discriminatory) and significant health inequalities. For example, individuals with poor socioeconomic status have an increased risk of infection and poor outcomes, which has been emphasised during the COVID-19 pandemic. Thus, when developing artificial intelligence (AI) solutions, it is important to ensure they are un-biased through fairness metrics. Equalized odds (EO) can be considered the most relevant measure of fairness in this scenario given we want to acknowledge and ideally minimize false positives (i.e., predicting survival for patients who die) as well as obtain equal performance across sensitive attributes groups.

Methods A many-to-many long short-term memory recurrent neural network (LSTM-RNN) was developed that uses patient features from MIMIV-IV including lab test results and clinical parameters to predict mortality for patients receiving antibiotics within the intensive care unit. Predictions were generated for all individuals within an unseen test set. Results were then broken down by the sensitive attribute classes gender, socioeconomic status (i.e., insurance type), and ethnicity.

Results To attain EO the true positive rate (TPR) across groups within a sensitive attribute class and the false positive rate (FPR) across groups must be equal or at least similar, meaning the model has balanced performance across the sensitive attributes groups. EO was achieved for gender with a TPR of 0.82 and 0.79, and a FPR of 0.42 and 0.44, for males and females respectively. Reasonable results were obtained for socioeconomic status (Table 1), although the TPR and FPR showed greater variation across groups than seen for the sensitive attribute gender.

Performance across ethnicities was not very consistent though (Table 1), with model outputs particularly differing between those groups frequently and infrequently present in the dataset such as white and native American or Asian populations (Native FPR undefined due to no individual dying in the test set).

Conclusions The model demonstrated some EO fairness across genders, but ethnicity biases were present. Future research will investigate how to ensure datasets are representative and best mitigate these biases within AI models, particularly against minority groups, to obtain consistent performance across the intended patient population.

TPR and FPR for sensitive attributes groups

Table 1: True positive rate and false positive rate results across the sensitive attribute classes gender, socioeconomic status and ethnicity and their respective groups.

Sensitive attribute	Gender		Socioeconomic status			Ethnicity							
Groups	Male	Female	Medic-aid	Medi-care	Other	White	Black	Hispanic	Asian	Native	Other	Un-known
True positive rate	0.82	0.79	0.82	0.76	0.85	0.85	0.78	0.78	0.79	0.57	0.71	0.63
False positive rate	0.42	0.44	0.34	0.43	0.45	0.48	0.41	0.45	0.30	NaN	0.30	0.28

Keyword 1
Artificial intelligence (AI)
Keyword 2
Fairness
Keyword 3
Biases

Conflicts of interest

Do you have any conflicts of interest to declare?
I have no potential conflict of interest to report

Other support
William Bolton was supported by the UKRI CDT in AI for Healthcare http://ai4health.io (Grant No. P/S023283/1)