18 March 2025 · CC-BY 4.0

Qeios

Peer Review

Review of: "Fair Diagnosis: Leveraging Causal Modeling to Mitigate Medical Bias"

Sami Zhioua[1,2]

1. CS, Ecole Polytechnique, France; 2. University of Qatar, Qatar

The aim of the paper is clear: assessing discrimination more accurately by isolating the direct effect from the indirect effect of S on \hat{Y}. However, I am not sure I understood very well the goal of the adversarial part of the work. The aim of that part should be better motivated and explained. In particular, is the goal to just assess discrimination or to pretrain the data to reduce discrimination based on the proposed metric (DF) ?

On the formalization aspect of the paper, I would say that Definition 3 is wrong. Defining the indirect effect as the difference between the total effect and the direct effect does not always hold. As far as I know, it holds only when the structural equations are linear. Otherwise, it is wrong to assume it. Also, as it is defined, the direct effect should be more accurately called the Controlled Direct Effect as opposed to the most common definition: the Natural Direct Effect (NDE).

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/FLFIM5