



FAIR: a recipe for ensuring fairness in healthcare artificial intelligence

Takashi Yoshiura¹ · Shigeru Kiryu²

Published online: 4 August 2023

© The Author(s) under exclusive licence to Japan Radiological Society 2023

The issue of the fairness in automated decision-making algorithms has been actively discussed since the mid-2010s. Cases of biased algorithms causing various ethical issues have been reported [1, 2]. These problems could promote and further perpetuate inequality and discrimination in society [3]. In their review article, Ueda et al. [4] focused on artificial intelligence (AI) in healthcare, particularly in radiology, and defined fairness as “development and deployment of unbiased AI that provides accurate diagnoses and treatment for all patients regardless of their social status or ethnic differences.” They proposed a list of recommendations to achieve fairness in healthcare AI called the Fairness of Artificial Intelligence Recommendations (FAIR) principles.

The authors categorized the biases in healthcare AI at different levels, from development to clinical use, into four types: data biases, which is introduced when collecting and organizing the AI training data; algorithmic biases, which is introduced during the development and implementation of AI; clinician interaction-related biases, which is introduced when the physicians use AI for patient care; and patient interaction-related biases, which is introduced between AI and patients receiving AI-driven care. Each of these biases causes problems that hamper patient equality in receiving high-quality medical care. Moreover, the problems can be complicated and difficult to address because of the inherent characteristics of AI: (i) healthcare AI is trained using a large amount of personal information, (ii) AI decision-making processes are often unclear and unexplainable to

humans, and (iii) unfamiliarity with AI among physicians and patients could cause overreliance on or the unreasonable refusal of AI.

To mitigate the abovementioned biases, countermeasures at their corresponding levels, such as collecting diverse and representative training data, conducting regular audits and validation of AI algorithms, and educating physicians and patients, are indicated. Additionally, maintenance of privacy and security of the training data, clarification of stakeholder roles and responsibilities, transparency of AI algorithms, and explanatory visualization of the AI decision-making processes are presented as important factors to ensure fairness in healthcare AI.

To summarize the strategies to ensure fairness in healthcare AI, the authors presented the FAIR principles comprising 10 recommendations. This statement is intended to be applicable not only to AI in radiology but also to healthcare AI systems in general. Thus, it should be shared throughout various healthcare fields.

We believe that AI-based systems will provide powerful assistance in clinical radiology. Furthermore, in the near future, AI may efficiently and reliably reshape healthcare to offer substantial benefits to the patients. The achievement of fairness in healthcare AI would lead to improved healthcare.

Declarations

Conflict of interest TY received a research grant from Shin Nippon Biomedical Laboratories, Ltd.; SK received a research grant from Canon Medical Systems.

This is a commentary to DOI: <https://doi.org/10.1007/s11604-023-01474-3>.

✉ Takashi Yoshiura
yoshiura@m3.kufm.kagoshima-u.ac.jp

¹ Department of Radiology, Kagoshima University Graduate School of Medical and Dental Sciences, 8-35-1 Sakuragaoka, Kagoshima 8908544, Japan

² Department of Radiology, International University of Health and Welfare Narita Hospital, 852 Hatakeda, Narita 286-0124, Japan

References

1. Angwin J, Larson J, Mattu S, Kirchner L [Internet]. Machine bias; 2016. Available from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed 17 July 2023.

2. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–53.
3. Barocas S, Selbst AD. Big Data's disparate impact. *Cal L Rev*. 2014;104:671–731.
4. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, et al. Fairness of artificial intelligence in healthcare: review and

recommendations. *Jpn J Radiol*. 2023. <https://doi.org/10.1007/s11604-023-01474-3>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.