Peer Review File

Reviewer A

Comment 1: "Machine learning (ML), a technique that offers predictions based on existing data..." This definition seems rather light, machine learning can designate a family of methods, which can be supervised or unsupervised, for AF detection, we generally use classification methods (supervised), the authors should therefore be more explicit about these methods.

Reply 1: Thank you. We revised this section to be more explicit.

Changes in text: Lines 33-34

Comment 2: The only reference provided concerning the methods, deals with deep learning, this is not representative of all the machine learning methods used. For example, feature selection is one example of ML methods that can be applied in AF detection. See this reference for an example with original datasets: https://pubmed.ncbi.nlm.nih.gov/33947379/.

Reply 2: Thank you. We have added a more broad reference that discusses many types of machine learning in regards to electrocardiogram data.

Changes in text: The new source is citation 2 from the Journal of Electrocardiology.

Comment 3: Another aspect that is not discussed in this paper concerns the data processing, the authors only talk about ECG data, but it is common to treat the problems using RR interval time series, that requires a lot of data preprocessing. This this literature review for example: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246533/

Reply 3: The authors focused on ECG data because our commentary about bias pertains to the most used ECG databases. We clarified this in the text.

Changes in text: lines 46-47

Comment 4: Concerning the racial paradox evoked by the authors, the authors talk about this issue without really proposing ways to solve it, otherwise. The inclusion of socio-demographic variables in prediction models should be discussed from a technical point of view, and what would be the limitations?

Reply 4: Thank you. We have added text to clarify this

Changes in text: 95-104

Reviewer B: This is an interesting viewpoint that focuses on the racial bias in prediction models using AI. Indeed, as every model, AI-based models reflect the racial and socio-economical, and cultural pattern of the population of the dataset used for development and validation. It is not a problem only of AI-based models, but of all kinds of models, but deserves to be discussed. Some comments:

Comment 1: The title suggests that the authors will discuss several aspects of Machine Learning methods in health. Indeed, there are many questions and problems, as the interpretability of the results, the absence of test of their validity in the real world, the possibility of inducing errors, etc. Since these multiples themes are not discussed here, I would suggest changing the title to reflect the text, mostly based on the racial question
Reply 1: Thank you, the title was adjusted accordingly.
Changes in text: title

Comment 2: Screening with ECG for AFib is now recommended in some guidelines: see ESC2020 guidelines for AFib. "Opportunistic screening for AF by pulse taking or ECG rhythm strip is recommended in patients ≥65 years of age. (IB) and Systematic ECG screening should be considered to detect AF in individuals aged ≥75 years, or those at high risk of stroke (IIa B).
Reply 2: Thank you. I've corrected the text accordingly.
Changes in text: 33-34

Comment 3: "Hospitals should validate...". Oh, this reflects a hospital-centric view of health services not compatible with current clinical practice in the world. Maybe you should change to "Health systems should validate"
Reply 3: Thank you. I've corrected the text accordingly.
Changes in text: 101-102