Review of: "Implementing Machine Learning to predict the 10-year risk of Cardiovascular Disease"

Hamza Elhaj

Potential competing interests: No potential competing interests to declare.

After reading the following piece of work, I enjoyed reading this paper and found some exciting knowledge in this work. However, some points missed some clarifications in this work.

First, the authors explained that the study predicts a 10-year risk of CVD. However, there is no explanation for why ten years. How does it be different for either five years or 15 years? More proof is needed for this point.

Secondly, I understand that this paper does not try to treat CVD; instead, it tries to predict the happening of CVD. However, the article failed to explain the risks associated with this disease. The literature section lacks some clarity around this point.

Third, the study explains the shortcomings of the traditional models to predict CVD as a lack of generalization and indicates that ML models also have this issue. Therefore, the study fails to clarify how ML techniques would overcome this issue. Besides, it lacks the explanation of the ML capabilities to overcome the traditional issues.

Fourth, the study lacks clarity around the criteria behind selecting the 14 variables among the 76 variables. What are the reasons behind that!!

Fifth, despite. The declaration of the reason behind the transformation of the study from a multi-variable study to a binary study simplifies the work. This does not count as a scientific reason to do that. A scientific basis needs to be provided with more explanation around this point.

Sixth, this study uses publicly available data to generate the model. It is preferred to compare the study results with other or similar studies that used the same data to create prediction models.

Seventh, more description needs to be provided around the origin of the dataset, sampling and from which ethnicity since the authors emphasized that the traditional models are impacted by ethnicity, and it works better on certain ethnicities. Also, what are the inclusion and exclusion criteria for this sample data? More explanation needs to be included in this study.