

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

Teresa Magalhães¹

¹ Universidade Nova de Lisboa

Potential competing interests: No potential competing interests to declare.

After reading carefully the paper submitted, I found some important points that I consider relevant to the purpose of the paper.

Starting with the objective of the paper "to develop an ML-based model to predict 10-year CVD risk". Why ten years? Where in the development of the model the ten-year variable is set?

What and how the Cardiovascular Diseases are treated in the paper? Cardiovascular disease is a set of diseases that affect the heart and blood vessels, too different between them. The introduction doesn't explain what the cardiovascular diseases are and how you manage it.

The Authors don't also refer any risk predictor factors to the CVD. This leads to another reflection on the way you selected the 14 variables. What was the support evidence to select them?

Also it must be very well described what are the origin of the dataset. What is UCI Heart Disease? How the data from the patients were collected? Where? Are we including patients from hospitals? Or outpatients? Are we including all kind of patients? Is omitted also the characterization of the population. Without this I can't understand what was the data support for the algorithm.