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

Premaladha Jayaraman[1]

1 SASTRA Deemed University

Potential competing interests: No potential competing interests to declare.

1. Stated that the dataset has 76 variables, were examined, but only 14 relevant and complete variables were chosen for analysis. how it is chosen .Justify feature selection process.

2. plot correlation for all 76 variables to confirm the negative influence.

3. The proposed work is claimed as binary classification, but it has class 0 to class 4, Clarify. Are these Class 1 (sub classes represents Staging or severity??)

4. Why the Extra Trees Classifier is applied to choose the top four features. why not other classifiers??, And is this a real classifier that aims to choose contributing feature selection? clarify.

5. Share the number of data samples used in training and justify the cross fold validation value as 10. Why not it can use range between 2-4 for faster convergence?Justify

6. Include more details on the network or method you employed in feature selection process.

7. Expected the justification that why specific optimizer are choosen for implementation over other advanced optimizer and how that influences the results.

8. Brief, how you handle both discrete and continuous features at the same instance in the feature extraction. what is the threshold for continuous features.

9. A confusion matrix of the classified samples and no.of misclassified sample is expected to add more visibility to the proposed work.