

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

Kim Gaik Tay<sup>1</sup>

<sup>1</sup> Universiti Tun Hussein Onn Malaysia

**Potential competing interests:** No potential competing interests to declare.

Authors compared seven machine learning (ML) algorithms (Adaboost(AdB), Decision Trees (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forests (RF), Support Vector Machines (SVM), and XGBoost (XGB) to predict 10 year risk of Cardiovascular disease.

ML has been an active area of research for many years. Therefore, the novelty of this manuscript may depend on the specific criteria set by the editor for publication. If the revised paper is accepted for publication, there are some concerns that need to be addressed as follows:

1. **Revise abstract:** As it seems indicates this manuscript is a survey manuscript. An abstract should be structured with a clear problem statement, objectives, methodology and results.
2. **Define acronym:** All acronyms such as FRS, QRISK, ASCVD, DHP, GSCV, RSCV should be defined for the first time they are used in the manuscript to ensure clarity for readers.
3. **Remove unused techniques:** If ANN and NBC were not used in the study, they should be removed from Figure 1 to avoid confusion.
4. **Inconsistent variables:** The number of independent variables listed in Figure 3 should match the number and names of variables mentioned in Section 3.2. Any discrepancies should be corrected.
5. **Abbreviations for variables:** Provide abbreviations for variables listed in Figure 2 such as chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, and ST depression induced by exercise relative to rest.
6. **Accuracy comparison:** Since accuracy of AdB, DT, KNN, RF, SGM and XGB in the current study is lower than that in previous studies (Table 2), then authors should discuss potential reasons for this discrepancy and whether it impacts the significance of their findings.
7. **Accuracy statement:** The statement "The LR, SVM, and XGB models demonstrated similar performance with Accuracy ranging from 0.8525 to 0.8689" is wrong as accuracy for LR, SVM, and XGB are all 0.8525. Correct it.
8. **Tables 5-6:** Provide proper citations and discussion for Tables 5 and 6, addressing why KNN, SVM, and XGB show decreased accuracy after GSSV and why there is no change after RSCV compared to Table 4.
9. **Figure 6:** Remove the extra XGB bar from Figure 6 to ensure accuracy and clarity in the visual representation.
10. **AUC\_ROC values:** Discuss why the AUC\_ROC values in this study (ranging from 0.6959 to 0.8734) do not compete with the results in Table 2 (ranging from 0.864 to 0.927). Analyze potential factors contributing to these differences.

11. **Research gap:** What are the differences between this study and the previous study?