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

Lin Yang

Potential competing interests: No potential competing interests to declare.

This article discusses the performance evaluation of various cardiovascular disease prediction models, including both traditional models and machine learning models, through a comprehensive review of methods, algorithms, and datasets used in different studies. In the evaluation of traditional models, it is observed that they exhibit limited generalizability across different racial populations. As a solution, the researchers propose the use of machine learning to overcome these limitations. The article compares multiple machine learning algorithms, including Adaboost, Decision Trees, K-Nearest Neighbors, Logistic Regression, Random Forests, Support Vector Machines, and XGBoost, using metrics such as accuracy, precision, recallsensitivity, F1 score, and AUC-ROC to assess model performance.

Detals of deficiencies for this article:

1.Model Interpretability:Despite concluding that Adaboost, LR, RF, SVM, and XGB models perform well in the article, these machine learning models often involve numerous parameters and complex hierarchical structures, making the interpretation of model decisions challenging.
2.External Validation:The article only uses a single dataset for validating machine learning performance. This approach may result in models performing well on the current dataset but poorly on other datasets. Therefore, a better practice would involve additional external datasets to validate the models' generalization abilities, ensuring model performance stability across different datasets.

3.It only introduce the models, lacking the exactly explanation how these features related to the CVD works in this models. And relatively speaking, the amount of data appear to be sparse comparing with the large project they want to do. That ma y leads to an inaccurate result.