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

Bismita Choudhury

Potential competing interests: No potential competing interests to declare.

The review is extensive, however, neither it contributes anything novel in terms of the performance of the ML models nor provides any insight into how can CVD risk be assessed efficiently.

The authors should consider the following points:

1. In the abstract, they wrote, “The UCI Heart dataset served as the training data for various ML models, including Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and K-Nearest Neighbors (KNN). To optimize model performance, Cross Validation (CV), normalization techniques, and hyperparameter tuning were employed. ” It is not clear how many Decision Trees, Random Forests, Support Vector Machines, and Artificial Neural Networks were used. From the method section, it seems a single classifier from each type was used, therefore, I would suggest correcting the abstract. Additionally, in the method and/or result sections, could not find anything on Artificial Neural Networks as mentioned in the abstract.
2. In the introduction section, they mentioned, “we aimed to develop an ML-based model to predict 10-year CVD risk”. But, they did not develop anything. They just evaluated and compared the performances of multiple ML algorithms.
3. In the related work section, 2.2. CVD risk prediction using ML, they mentioned “In our study, we attempt to address these limitations.” However, the limitations were not highlighted well. A few limitations were mentioned in the paragraph, “ Some studies were limited to specific populations or datasets, limiting the generalizability of their findings. Additionally, the interpretability of some ML models, particularly DL models, may pose challenges, which can hinder their clinical utility. The accuracy and usefulness of ML models are highly dependent on the quality and completeness of the input data, and their performance may not generalize well to different populations or settings .”, however, it is not clear if their experiments could solve at least one limitation.
4. It is mentioned, that 10-fold cross-validation was used for training and validation of the models. However, it is not clear, whether the results were the overall accuracy after cross-validation or the maximum accuracy obtained by the models. Then, again, the results were for training and testing during cross-validation or the separate testing set they split before conducting cross-validation. So the question is: the accuracy, precision, recall, F1, and ROC-AUC (Table 4 and Fig 5) were results for an 80:20 split or cross-validation on the 80% training set or 20% testing set after cross-validation?
5. In 3.2. Feature selection and Data Engineering, it is mentioned, “For feature selection and engineering, we examined all 76 variables in the dataset. However, we determined that only 14 variables were both relevant and complete for our analysis,…” It is not clear what feature selection and engineering methods were used to select those 14 features out of

76. Then, there are actually 13 features, the 14th column is for the target class.

6. In 3.2. Feature selection and Data Engineering, it is mentioned, “We considered variables with high positive or negative correlations (correlation>0.40 or correlation<-0.40) as potentially influential predictors of CVD risk.” It is now clear how and why 0.40 was set as the threshold.
7. How and why ExtraTreesClassifier was used "to reveal the features with the highest importance"? Was the correlation matrix not good enough? From Fig 3, it seems “sex” and “fbs” have correlation values < 0.40/-0.40, then why these features were part of the evaluation process?
8. There is a lack of clear explanation as to why Hyperparameter Optimization and Grid Search Cross Validation (GSCV) were used and how they helped in hyperparameter optimization. What are the optimized parameters after applying these methods? Is there any significant performance improvement?

There are a lot of unexplained points, mistakes, inconsistencies in writing, etc. Therefore, I would suggest the paper is not ready for publication yet.