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

Rafael Leon[1]

1 Universidad de Alcalá de Henares

Data used in this research is clearly insufficient: 303 cases and 14 variables that seem to be very similar to heart.csv dataset of Tensorflow web site. And this is an example dataset, not for scientific research purposes. It is not possible to derive any valid conclusion from these data.

Machine Learning models like Support Vector Machines, Random Forest ML, …, have a short introductory description, but authors should explain how they have configured and implemented them with detail.

As far as I can see on the Jupyter notebook, models have a very basic configuration, those tests must be repeated with more configurations for every model. Authors should consider using fewer Machine Learning strategies, with a deeper and broader focus on the ones chosen.

Authors have used Grid Search Cross Validation and Randomized Search Cross Validation for models hyperparameter tuning, but it seems that the experiments carried out are not extensive enough to reach any conclusion.