CAD-EDT: Using Meta-Ensemble Machine Learning to Power an Application Specializing in the Early Detection of Coronary Artery Disease Using Easily Accessible Risk Factors and Cardiac Clinical Data

Madan, Aryan (School: Roanoke Valley Governor's School for Science and Technology)

Coronary artery disease (CAD) is a condition with high global prevalence. At least 75% of deaths from CAD occur in low- and middle-income countries due to the lack of predictive expertise before a myocardial infarction (MI); this is the first symptom for 25% of patients. The goal of this study is to design a web application that utilizes meta-ensemble machine learning to advance the prediction of CAD. A 574-sample heart disease dataset consisting of patient-specific risk factors and electrocardiogram findings was used for experimentation. The data was cleaned, and features were scaled based on statistical distribution. 15 various machine learning algorithms were compared using their cross-validation accuracy scores. The top four models were hyperparameter-tuned using a Bayesian optimization. Two meta-ensemble algorithms, stacking and voting, were trained using the four tuned models as base learners. The meta-models were evaluated on a variety of performance metrics. The voting classifier achieved an accuracy of 94.78%, succeeding the lowest base learner by 4%. The web application allows one to enter a patient's data and receive a predicted class and probability for the disease. Getting an accurate prediction for CAD can be the difference between life and death for many; the severity of the disease increases after the first MI. The current rates of cardiovascular disease misdiagnosis by a general practitioner can be significantly improved with the help of this application. It will allow for a rapid, cost-effective, and non-invasive diagnosis of CAD, especially for those who have no prior knowledge of their condition.

Awards Won: