A Novel Approach to the Diagnosis of Heart Disease Using Machine Learning and Deep Neural Networks

Ankireddy, Sahithi (School: James B. Conant High School)

Heart disease is the leading cause of death worldwide. Currently, 33% of cases are misdiagnosed. The use of Artificial Intelligence could reduce the chance of error leading to possible earlier diagnoses which could be the difference between life and death for some. The objective of this project was to develop accurate Machine Learning (ML) and Deep Neural Network (DNN) algorithms to create a application for assisted heart disease diagnosis. My project compared the ML model with the DNN and the most accurate model is used for the application. A series of scoring methods such as confusion matrices, Matthews Correlation Coefficient and ROC AUC scores/graphs were used to determine which model, out of well-known ML algorithms, performs the best under the dataset. Random Forest (RF) Classifier performed the highest out of the KNN, SVM and the Naive-Bayes models. All the ensemble classification methods in sci-kit learn were also evaluated to reconfirm that RF Classifier was the best. For the DNN, the Keras Sequential Model was used, running on a backend of TensorFlow. A dataset provided by the the Cleveland Clinic Foundation was used to train the models. The dataset was split into a training and validation set, and the models were built based on various optimization techniques and an GridSearch algorithm. Cross-validation, a resampling procedure, was applied to further estimate the skill of the model. The application, running on Flask, and utilizing Bootstrap was developed using the RF ML model, as it performed higher than the DNN. Required data inputs for patients are read to determine an accurate diagnosis. This tool can possibly be a great contribution to the cardiology field as it can be used by medical care professionals to assist them in more accurate diagnoses.