Optimizing Machine Learning Models to Predict the Likelihood of Stroke With State-of-the-Art Accuracy

Chada, Saianiruddh (School: Rider High School)

Brain strokes are a prominent, life-threatening condition that significantly afflicts the world population, with 15 million people suffering from a stroke annually. A large percentage of stroke victims are left permanently disabled and many of the victims are left further susceptible to cardiovascular diseases and numerous other health problems. The goal of this project is to create a machine-learning model, which takes a patient's characteristics, such as age, gender, BMI, and whether they have hypertension and heart disease, and predicts a patient's risk of stroke with state-of-the-art accuracy. Creating the model, training the model with training data, testing with test data previously not seen by the model, and finally optimizing the model by tuning the hyperparameters are the major steps in this project. With this study, I was able to create a highly accurate machine-learning model for predicting the likelihood of stroke by tuning the number of layers, the number of neurons in each layer, the type of optimizer, and the number of iterations used for training. It was possible to achieve a state-of-the-art accuracy of 96.35% with my MLPClassifier neural network model, with 2 hidden layers, 200 neurons per layer, and an LBFGS optimizer, after being trained for 2000 iterations. A model with this level of accuracy has the potential to be used in a clinical setting for predicting the likelihood of stroke for a patient. This information can then be used by physicians to encourage their patients to modify the contributing risk factors.