SYSC4415 Group 11

Assignment 3 Report

By: Justin Huang 101073186
Mingzhi Chen 101036533
Rebecca Li 101150053

April 11th, 2023

How did you split the data into train/test/etc.

The train_test_split function from sklearn.model_selection was utilized to split the training in a 80/20 ratio, in which 80% was used for training and 20% was used for validation. Both sets have a balanced distribution of classes. The test set was provided separately, was not used in the training process and it was not split.

What ML approach you used

A combination of three machine learning algorithms was used, which includes a CNN (convolutional neural network), an XGBoost classifier, and a RandomForest classifier. The CNN combines a 1D model for processing signal data and a 2D model for processing signal images. Both the XGBoost and RandomForest classifiers process selected extracted features.

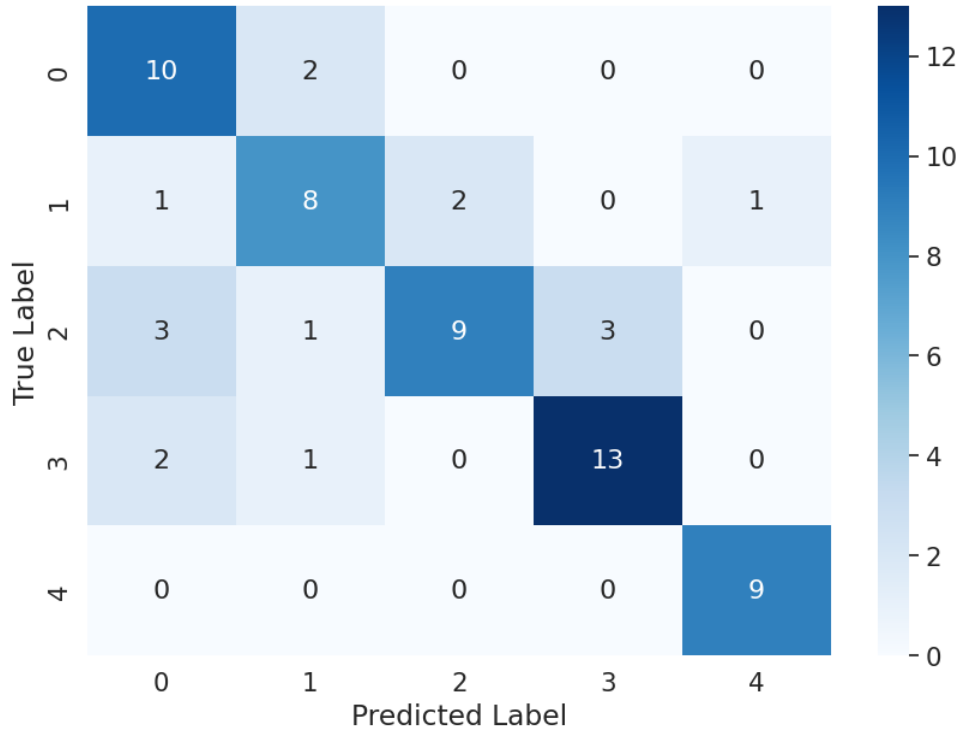
What features you used

The CNN model uses un-altered signals and images as training features. XGBoost and RandomForest classifiers used most relevant features from the extracted_features file. Features from extracted_features included statistical measures and other information derived from the raw signal data.

What method of combining models did you use?

Predictions from each model were recorded and saved to individual csv files. Each model's val test accuracy was also recorded to be used in the ensemble combining process. Ensemble prediction was calculated by using a weighted approach based on the val set performance of each model, the results were then saved in a .csv file.

Confusion matrix



PR-curves for 1-vs-rest for each class (this can go on a separate page if it doesn't fit)