Question 4

Ian Dover

January 2023

1 Prediction

Using B-Spline basis functions and fPCA, feature engineer the data for prediction on the test set:

Four steps were performed for classification of the ECG data:

1. B-spline coefficients were determined for each ECG signal.
2. The coefficients were dimensionally reduced using PCA.
3. XGBoost was trained on RandomizedSearchCV with 5 folds and 20 iterations.
4. Then best model had the performance metrics determined on the test set.

ROC-AUC of the final model:

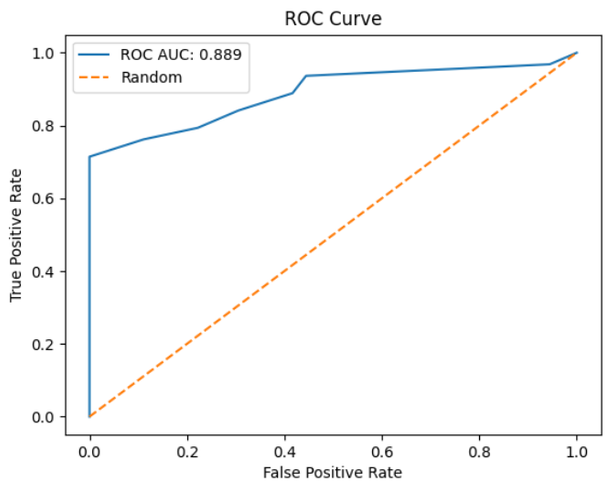

Figure 1: Receiver Operating Characteristics of the Model on the Test Set

Accuracy of the model at multiple probability thresholds:

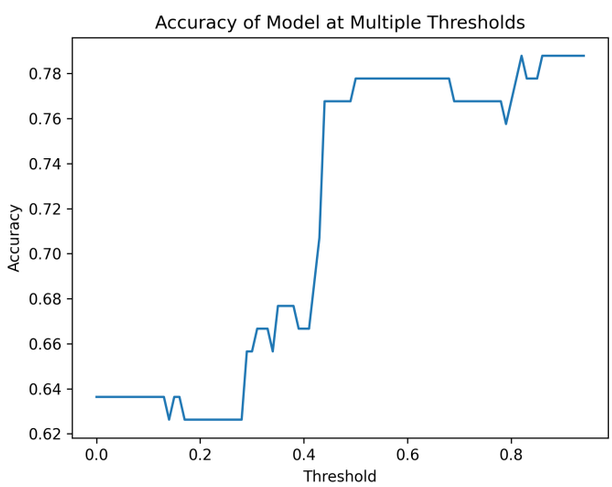

Figure 2: Accuracy of the Model at Multiple Probability Thresholds on the Test Set

F1 of the model at multiple probability thresholds:

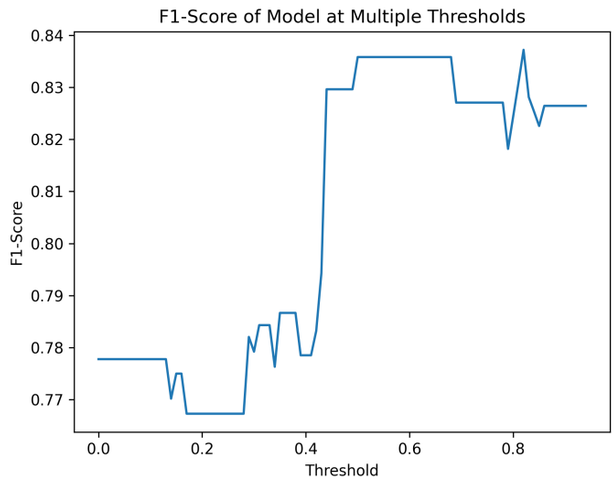

Figure 3: F1-Score of the Model at Multiple Probability Thresholds on the Test Set

Based on the results of Figure (2) and Figure (3), the threshold of 0.9 was selected for the probability threshold of the model to maximize overall accuracy. This threshold is then used to derive the confusion matrix. Confusion matrix of the final model:

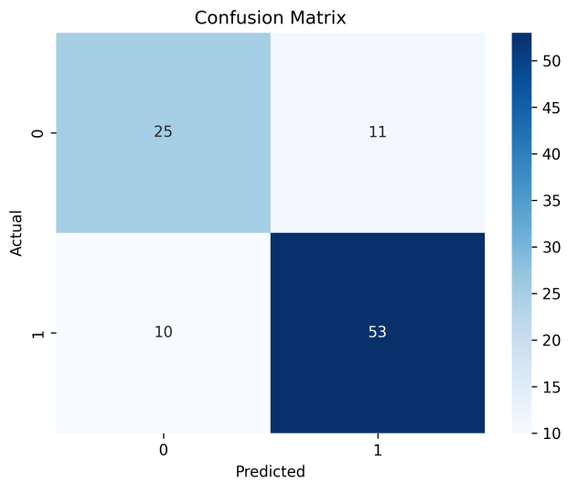

Figure 4: Confusion Matrix of the Model on the Test Set