**Binary classification of AF/Normal 20 sec ECG samples**

* Dataset (30 min ECG signal):

- 30-min ECG signal downloaded from PhysioNet

- There are 23 ECG signals are labeled as AF and 18 ECG signals are labeled as Normal.

* Samples (20 sec ECG):

- non-overlapped 20 sec ECG samples extracted from labeled (AF/Normal) 30 min ECG (each patient). All 20 sec ECG samples preserve the same labeling as the 30-min ECG which they are extracted from.

- All 20 sec ECG samples are randomly mixed to perform non-patient specific modeling.

* Data splitting:

- Training data: 1000 AF & Normal samples.

- Test data (the rest): 3140 AF & 2240 Normal samples.

* Feature engineering:

- Features: MSE & MSF of 20 sec ECG samples.

- Feature scaling: Apply min-max normalization to each feature separately. The minimum and maximum scales are based on only training data and then apply to test data.

* Training:

- Machine learning technologies applied: linear SVM, rbf SVM.

- Model selection is performed via 10-fold cross-validation on training data to find the model with the smallest validation error. Note that in each experiment of cross-validation, balanced learning set (for generating the classifier) and validation set (for estimating the validation error of the classifier) are applied.

- Validation error = (#false positive + # false negative)/# validation samples.

* Test:

The trained model would be applied to each test sample and record the prediction accuracy of the two classes (AF/Normal).

Test results:

Classification accuracy (linear SVM)

|  |  |  |
| --- | --- | --- |
|  | Training set | Test set |
| AF | 0.55 | 0.55 |
| Normal | 0.90 | 0.90 |

Classification accuracy (rbf SVM)

|  |  |  |
| --- | --- | --- |
|  | Training set | Test set |
| AF | 0.65 | 0.62 |
| Normal | 0.87 | 0.85 |