

Combination of Ensemble Learning Based Models to Classify 12-Lead ECGs

Mohammed H Baydoun⁽¹⁾, Lise E Safatly⁽²⁾, Hassan M Ghaziri⁽¹⁾, Ali El-Hajj⁽²⁾

⁽¹⁾ Beirut Research and Innovation Center, Beirut, Lebanon

⁽²⁾ American University of Beirut, Beirut, Lebanon

This work deals with the Physionet 2020 challenge which targets Classification of 12-lead ECGs. The challenge deals with a 9-class multi-label output that should be predicted for each input. The labels can be normal and others such as Atrial Fibrillation and several cardiac abnormalities.

The provided data includes the 12-lead ECG record in addition to the age and gender of the patient besides the diagnosis. The preliminary approach relies on detecting the R-peaks information of each of the 12 leads using the Pan-Tompkins algorithm followed by combining the output of each lead into a more accurate R-peaks localizer. Afterwards, each of the P, Q, S and T peaks are detected based on their vicinity to the found R peak. This is then followed by extracting several features that utilize time, peaks and statistical information such as the mean RR interval, mean R peak value, standard deviation of these and others for each found peak. This leads to a feature vector for each lead. Then, the median and the standard deviation of these vectors are found and combined with all the feature vectors to provide a feature vector of more than 3000 features. This can be utilized as an input to a classification system to obtain the output. The work utilizes 9 different binary classifiers with one for every output. The classification systems are all ensemble models and include Adaboost, GentleBoost and RUSBoost classifiers. This leads to 9 binary outputs that are combined to yield the output.

The proposed algorithm was checked both locally by utilizing 5-fold cross-validation and on the initial test set. The scoring function is the geometric mean of two scores. The first is the class-weighted F-score and the second is a generalization of the Jaccard measure (G-score). We provide the results in the shown tables.

Initial cross-validation results

Utilized Score	Value
F-score	0.81
G-score	0.60
Geometric Mean	0.697

Initial test results

Utilized Score	Value
F-score	0.79
G-score	0.57
Geometric Mean	0.671