ECG classification combining conventional signal analysis and tree-based machine learning algorithms

Martin Kropf[1], Martin Baumgartner[2], Sai Veeranki[3], Lukas Haider[2], Dieter Hayn[2], Günter Schreier[2]

1. Department of Internal Medicine and Cardiology, Charité University Medicine, Campus Virchow-Klinikum, Berlin, Germany
2. AIT Austrian Institute of Technology, Graz, Austria
3. Institute of Neural Engineering, Technical University of Graz, Graz, Austria

Introduction:

Electrocardiograms (ECG) represent a noninvasive and indispensable tool in clinical diagnosis which indicates various cardiac abnormalities. ECG signals are typically interpreted by experienced clinicians, who analyze the raw ECG signals and interpret features derived by signal processing algorithms. In recent years, research also focused on automated ECG interpretation utilizing recent advances in machine learning. This paper presents a combined approach with conventional signal analysis and tree-based machine-learning algorithms by the team *easyG* for competing in the Computing in Cardiology Challenge 2021.

Methods:

We split all available ECGs into a 5-fold cross-validation scheme. The ECGs were normalized, resampled to a 500 Hz sampling rate and trimmed to a maximum length of 60 seconds.
To classify ECG recordings into the 27 classes as defined by the challenge, we developed a MATLAB-based signal processing unit, which was combined with models implemented in Python. To handle the multi-label problem, we used one-hot-encoding and trained a one vs. rest classifier. We tested several base classifiers, e.g. random forests and gradient boosting machines.

Result:

During the unofficial phase, we achieved a challenge score of 0.17 on the hidden test set for the 12-lead dataset using the random forest as base classifier. Our average internal challenge score on the 5-fold cross-validation scheme was 0.61 using a gradient boosting machine as base classifier.

Discussion:

Our results indicate that our feature-based approach is able to detect various cardiac abnormalities. In the future we will also evaluate how well our approach works with 6, 3 or 2 channels. Furthermore, we will employ model stacking, to build a meta-classifier based on several base-learners. We assume that the performance will be better, if several complementary learners are combined. Another idea is to apply weights based on the frequency of samples per class, as the distribution of classes is severely unbalanced.