An AutoML pipeline for the classification of noisy and clean electrocardiogram signals using Bayesian optimization

Lucia Billeci[1], Lorenzo Bachi[1], Chiara Podrecca[2], Maurizio Varanini[1], Pia Cincione[3]

1 Institute of Clinical Physiology, National Research Council of Italy (CNR-IFC), Pisa, Italy
2 Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy
3 Medical Department, University of Foggia, Foggia, Italy

Introduction
Electrocardiogram (ECG) can be affected by notorious sources of noise including powerline interference, muscle noise and motion artifacts. Although the numerous noise indices in literature, no single index emerges as the best. Our study introduces an AutoML pipeline that leverages multiple indices for ECG noise detection.

Methods
Our dataset resulted from merging multiple databases and each of the 155643 2 s long windows of signal was labeled as noise or clean. Twelve noise indices were considered, including two novel ones. The dataset was first split in 55% for training, 15% for validation and 30% for testing considering patients' separation. Then the training set was partially balanced by randomly undersampling the majority class (clean).

AutoGluon tool was used for the classification. Four different quality presets were tested. The model that provided the best performance (by balancing balanced accuracy – BA - and inference time), underwent Hyperparameters Bayesian Optimization (BO) and the optimized model was tested on the test set.

Results
On the validation set the Ensemble model of the Medium Preset demonstrated the highest BA (0.952), with the NeuralNetTorch also showing high BA (0.949) and a very low inference time (0.138 s), making it an optimal choice for BO. After optimization, the top three models showed improved BA on both validation and test, with slightly increased but still low inference times. Specifically, on the test set, post-BO accuracies ranged 0.952 to 0.948, with inference times between 0.56 to 0.10 s, an enhancement from pre-BO where accuracies ranged 0.946 to 0.944 with inference times between 2.64 to 0.13 s.

Conclusions
The ML approach proposed allows to accurately classifying ECG signals with a low inference time, suggesting the possibility to adopt this pipeline in a clinical scenario. Further research may be warranted to expand the database with more noisy and arrhythmic episodes.

model	score_test	score_val	pred_time_test	pred_time_val	fit_time
WeightedEnsemble_L2	0.951731	0.960149	0.563059	1.313317	1147.939857
NeuralNetTorch/e6f08692	0.950168	0.955674	0.221023	0.605468	846.816182
NeuralNetTorch/737716ad	0.948659	0.956789	0.108905	0.381413	90.524493

Figure 1: Performance of the three best models after Bayesian optimization on the test and validation sets.