An Ensemble of Machine Learning Models for Multilabel Classification of Cardiovascular Diseases by ECGs

Anastasia Bazhutina[1], Svyatoslav Khamzin[1], Mikhail Chmelevsky[1,2], Stepan Zubarev[1], Margarita Budanova[1], Werner Rainer[1]

[1]XSpline S.p.A, Bolzano, Italy
[2]Division of Cardiology, Fondazione Cardiocentro Ticino, Switzerland

The electrocardiogram (ECG) is an integral characteristic for evaluating cardiac electrophysiology. Therefore, automatic analysis of ECG signals, in particular, obtaining a diagnosis by ECG, is an important practical task for choosing a treatment strategy. Today, modern methods of machine learning and neural networks are used for automatic ECG classification. This study introduces two beat-to-beat classifiers for the automatic detection of cardiovascular disease by 12-lead ECG: the first detects a type of heart conduction disorder, and the second detects an infarction type.

For the training of classifiers, three open datasets were used: Chapman-Shaoxing, PTB-XL, and the dataset of Shandong Provincial Hospital. Each dataset contains 11 to 72 classes. Before the training stage, we perform a preprocessing of ECGs. Initially, all ECG records were filtered and noises in signals were deleted. Next, for each lead, only one QRS was chosen. For each QRS complex, we calculate statistical, time-domain, frequency features, and scalograms. For the first classifier (**10 classes**), we selected: normal ECG, LBBB, ILBBB, ventricular premature complex, left anterior fascicular block (FB), left posterior FB, IRBBB, RBBB, nonspecific intraventricular conduction disturbance, and ventricular preexcitation. For the second classifier (**5 classes**), we selected: non-infarcted ECGs, anteroseptal myocardial infarct (MI), lateral MI, inferior MI, and anterior MI. As a backbone of our solution, we implemented an ensemble of models: XGBoost for statistical features, an autoencoder+XGBoost for time-domain and frequency features, and ResNet for scalograms. The output result is a weighted prediction of ML models. The main advantage of this study is the extraction of all signal information in different ways and using it for efficient prediction in machine learning models.

In the table, the mean weighted values at validation samples for the 10 and 5 classes are presented. Current results may be improved by data balancing and weight settings in the ensemble.

Classes, n	ROC AUC	Accuracy	Specificity	Sensitivity	F1-score
10 (11024)	0.89	0.89	0.92	0.72	0.70
5 (10773)	0.92	0.86	0.88	0.80	0.76