Classification of Cardiac Abnormalities Based on Varying Dimensions in ECGs by An Ensemble Machine Learning Model

Pu Jian[1], Yicheng Yao[1], Minfang Tang[1], Zhengling He[1], XianXiang Chen[1,*]

[1] University of Chinese Academy of Sciences, Beijing, China;

* Correspondence: cxxvac@163.com

The electrocardiogram (ECG) has been widely used to diagnose a variety of cardiac abnormalities. However, the automatic detection of cardiac abnormalities based on reduced-lead ECGs is difficult to demonstrate the early and correct diagnosis of cardiovascular diseases. We have developed an ensemble machine learning model to detect and classify cardiac abnormalities from either twelve-lead, six-lead, three-lead, and two-lead ECGs. The basic classifier models extract convolution features from ECG records by using a convolution neural network and combine bi-directional gated recurrent unit (GRU) for temporal aggregation of features. Especially, the self-attention (SA) is introduced to prevent the loss of information in a long-time sequence. Five folds cross-validation is used to produce five best validation models as the basic models. Meanwhile, a framework for joint iterative optimization of parameters and labels is proposed to solve the problem of noisy labels resulting from the preprocess of ECGs segmentation to fixed length. During the training procedure, we introduce a data augmentation scheme of oversampling to deal with the class imbalance issue. Finally, we ensemble models of different reduced-lead ECGs to predict cardiac abnormalities. By cross validating on the training set, it achieved an average score of 0.538 in the 12 leads test set, 0.531 in the 6 leads set, 0.527 in the 3 leads set, 0.524 in the 2 leads set, respectively. The name of Our team is F-team. Unfortunately, because the data type of binary output is set as float instead of bool, it cannot be recognized by the evaluation function. The score from the unofficial phase is -0.406. The true accuracy of our model has not been present.