Classification of Cardiac Abnormalities by An Ensemble Machine Learning Model

Minfang Tang[1,2], Chenshuo Wang[1,2], Ruishi Zhou[1,2], Yicheng Yao[1,2], Pu Jian[1,2], Xiuying Mou[1,2], Zhen Fang [1,2,*]

[1] University of Chinese Academy of Sciences, Beijing, China; 1293227181@qq.com;312353235@qq.com; zhouruishi17@mails.ucas.ac.cn;yaoyicheng19@mails.ucas.ac.cn; 3195794910@qq.com; shadow_mou@126.com.
[2] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

* Correspondence: zfang@mail.ie.ac.cn.

The electrocardiogram (ECG) has been widely used to diagnose a variety of cardiac abnormalities. The automatic detection of cardiac abnormalities plays an important role in the early and correct diagnosis of cardiovascular diseases. We have developed a stacking-based integrated machine learning model to detect and classify cardiac abnormalities. The basic classifier models extract convolution features from 12-lead ECG records by using a convolution neural network and combine bi-directional gated recurrent unit (GRU) for temporal aggregation of features. Five folds cross-validation is used to produce five best validation models as the basic models. Meanwhile, we use a logistic regression classifier (LR) as the secondary classifier of the model. The time-frequency domain analysis results and sample entropy of heart rate variability (HRV) are calculated and set together with the predicted classification probabilities of the basic models as the inputs of the LR to implement the classification task. During the training procedure, we introduce a simple data augmentation scheme of oversampling to deal with the class imbalance issue. And it achieved a F_beta score of 0.807 and a G_beta score of 0.579 in the test set. We have submitted a total of 2 entries during the unofficial phase of the Challenge, and received the best F_beta score is 0.802 and G_beta score is 0.590.