Multi-Label Classification of 12-lead ECGs by Using Residual CNN and Class-Wise Attention

Yang Liu[1], Kuanquan Wang[1], Yongfeng Yuan[1], Qince Li[1,2*], Yacong Li[1], Yongpeng Xu[3], Henggui Zhang[2,4,5]

[1] School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin, China
[2] Peng Cheng Laboratory, Shenzhen, China
[3] Yongjia County Public Security Bureau, Wenzhou, China
[4] School of Physics and Astronomy, The University of Manchester, Manchester, UK
[5] Key Laboratory of Medical Electrophysiology of Ministry of Education and Medical Electrophysiological Key Laboratory of Sichuan Province, Institute of Cardiovascular Research, Southwest Medical University, Luzhou, China

Abstract

Cardiovascular diseases have become the leading cause of illness and death worldwide. Due to their chronic nature, early screening and follow-up management will effectively improve the prevention and treatment of cardiovascular diseases, where automatic electrocardiogram (ECG) classification will play an important role. In this work, we take part in the 2020 PhysioNet - CinC Challenge (in the team ECGMaster) and propose a novel multi-label classifier of 12-lead ECG recordings which combines a residual convolutional network (residual CNN) with a class-wise attention mechanism. To deal with the problem of data imbalance between classes, we utilize a novel weighted focal loss in the training of our models. Our models were trained and tested in a 5-fold cross validation on the training data with resulting scores of 0.5501 ± 0.0223 according to the challenge metric, demonstrating a promising method for the classification of ECGs. We note that we were unable to score and rank our model on the official test data, the results were obtained on the training set only and may be over-optimistic.

1. Introduction

Cardiovascular diseases, on behalf of a number of abnormal heart conditions, have been regard as the leading cause of morbidity and mortality around the world [1]. The standard 12-lead electrocardiogram (ECG) is the most commonly used diagnostic tool for the cardiovascular diseases. Up to now, the treatment of ECG still relies on manual interpretation, which is time-consuming and error-prone. In addition, the miniaturization and portability of ECG devices have been inducing the rapid growth of recording volume. In consequence, automatic detection and classification of ECG abnormalities is very necessary to reduce the working pressure of physicians and improve the accuracy of diagnosis.

Although the study on algorithms for automatic ECG analysis has been going on for several decades, it is still a challenging work to accurately detect and identify each abnormality in an ECG recording. The difficulty mainly lies in the diversity (up to hundreds) of abnormal types and the high variability of ECG signal. The ECG signal measured on the body surface is a reflection of the electrical activities of millions of individual cells in the heart, thus conveys diagnostic information about the condition of the heart. However, besides the cardiac electrical activity, many other factors can affect the morphology of the ECG waveform, including anatomical structure [2] and conductivity of the body [3], respiration [4] and skeletal muscle movement [5], electrode placement [6], and electromagnetic noise in the environment [7]. All these factors bring difficulties to the interpretation of electrocardiogram, and even skilled experts can differ significantly in their interpretation of certain ECG records.

Because of the huge complexity of ECG classification, it is impractical to directly translate human knowledge into algorithms. As a result, data-driven methods, especially machine learning methods, have become the mainstream in this field in recent years. A machine learning algorithm for ECG classification typically consists of three steps: preprocessing, feature extraction and classification. In the preprocessing, noise suppression and data transformation methods are usually applied to reduce noise interference and enhance the recognisability

of the patterns [8, 9]. Then, features are extracted from the preproessed data and are used for the final classification. Traditionally, the feature extraction and classification are implemented separately with different methods [10, 11]. However, the deep learning method incorporates these two parts into a unified end-to-end neural network [12].

In this study, we try to deal with the problem of multi-label classification of 12-lead ECG based on the dataset of PhysioNet/Computing in Cardiology Challenge 2020 [13], which incorporates ECG data from six databases collected by 4 different institutions. The public available training set includes a total of 43101 ECG recordings with 111 types of labels of which 24 types are for scored. Each record is labelled as one or several categories, with the distribution of label numbers shown in Figure 1. In addition, this dataset is extremely imbalanced with 20846 samples for the biggest category (normal sinus rhythm, NSR) and only 188 samples for the smallest category (premature ventricular contraction, PVC). Therefore, this dataset provides rich data, and at the same time effectively reflects the challenges of automatic ECG classification in practice, so it is of great research value. In the following sections, we will describe in detail our approaches to these challenges along with the experimental results.

2. Methods

In this work, we adopt the deep learning methodology to deal with the problem of multi-label classification of 12-lead ECG. The raw ECG signal is first preprocessed to remove baseline wanders and some other noises. Then, a novel neural network combining a residual CNN and a class-wise attention mechanism is proposed to extract features automatically and detect the existence of each type of heart condition. In addition, in order to address the data imbalance problem, we propose a novel weighted focal loss for the model training.
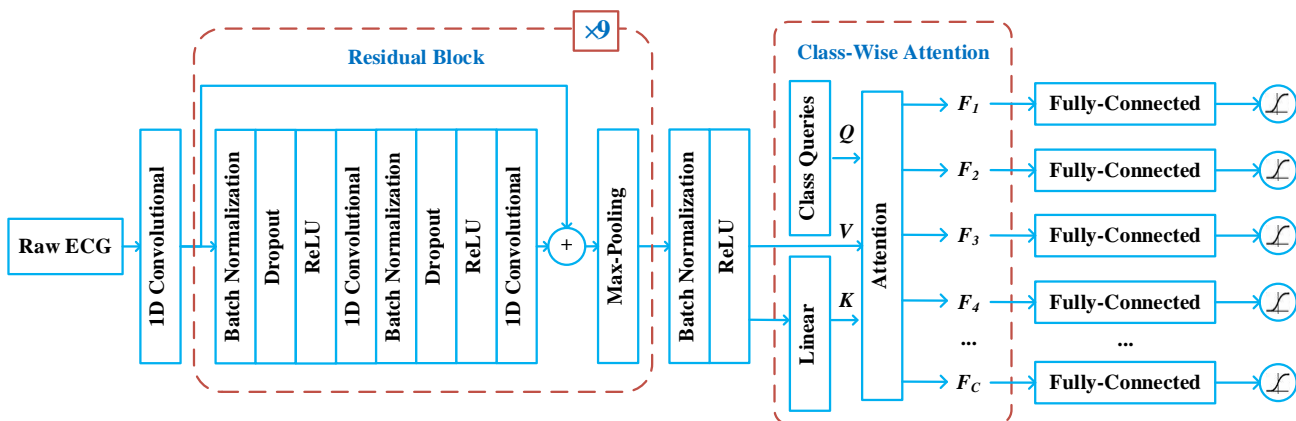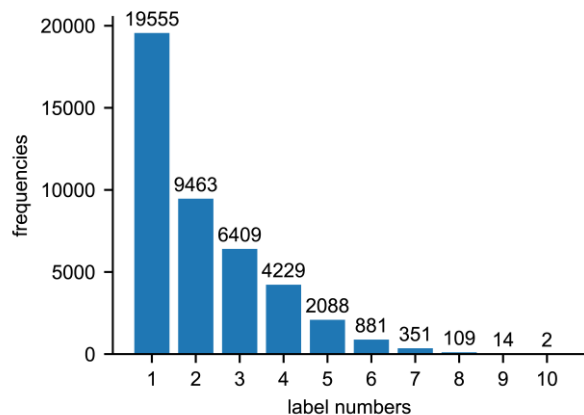

Figure 1. The distribution of label numbers of the recordings in the challenge dataset.

2.1. Preprocessing

In the preprocessing stage, the raw ECG signals are first scaled into the same amplitude resolution (1000/mV) and resampled to 250Hz. Then, we utilize moving average filter with a window size of 250 to estimate the baseline wander and subtract it from the signal. The other noises, such as muscle movement and environmental noises, are suppressed by wavelet denoising based on 3-level 'db4' wavelet transform and soft-thresholding. For easy of batch processing during the model training, the recordings are truncated or padded with zeros to the same length (60 seconds in our experiments) in a random manner as stated in our previous study [12].

2.2. Network architecture

The structure of our network is constituted by three parts: residual CNN, class-wise attention and fully-connected layers, as shown in Figure 2. The residual

Figure 2. The proposed network structure for multi-label classification of ECG.

CNN part consisting of 9 residual blocks is used to extract features from the input ECG signal. Each residual block contains 2 convolutional layers each with three other assistant layers (batch normalization, dropout and ReLU activation). The output of the second convolutional layer is merged with the block input by addition and down-sampled by a max-pooling layer to half of the input length. Each convolutional layer in the first block contains 32 filters each with a kernel size of 32. As the network goes deep, for every three residual blocks, the filter number grows by 32, while the kernel size is reduced by half. In consequence, the last (or 9th) block has 96 filters with kernel size of 8 in each of its convolutional layer.

The residual CNN output a feature map describing the features of different part of input ECG signal. To make a diagnostic on the whole signal, we need to aggregate the local features into a global feature vector. In this network, we utilize an attention mechanism to do this work. As our model aims to deal with multi-label classification, and different parts of an ECG signal may contains features relevant to different classes, the attention mechanism in our network is designed to be class-wise to harmonize the divergence of attentions between different classes. We will present the class-wise attention in the next subsection. The class-wise attention part generates an individual feature vector for each class, which is then fed into a fully connected layer with sigmoid as the activation function to predict the probability of the corresponding condition in the input signal.

2.3.   Class-wise attention

The class-wise attention is based on the scaled dot-product attention proposed by [14] whose input includes three parts: query vectors (Q), key vectors (K) and value vectors (V). The output of the attention layer is a set of feature vectors (F) which is one-to-one with the queries. In the setting of multi-label ECG classification, we use the one-hot encoding of the classes as Q. The K is linearly transformed from the local features in the extracted feature map, and have the same last dimension as the Q. The local features themselves are directly used as V. Therefore, we get the class-wise attention as follows:

$$F = softmax(Q(W^T V^T + b))V \qquad (1)$$

where W is the weights of linear transformation, b is the bias. Based on this mechanism, the network can learn specific attention for each class and avoid the conflicts between different classes.

2.4.   Weighted focal loss

One of the main challenges to train an effective ECG

classifier is the data imbalance of the training set. When training a DNN model with extremely imbalanced dataset, the instance-rich classes are prone to dominate the training procedure, leading to significantly worse performance for instance-scarce classes. To handle this challenge, data resampling and loss reweighting are the mostly common used strategies. In this work, we adopt a loss reweighting strategy by extending the focal loss method, an influential approach in field of pattern recognition [15]. We assign different weights to the final activations during the loss computation according to the frequencies of the corresponding classes. The loss function is as follows:

$$L(p_i) = \begin{cases} -\alpha_{i,0} p_i{}^r \log(1 - p_i) & \text{if } y_i = 0 \\ -\alpha_{i,1}(1 - p_i)^r \log(p_i) & \text{otherwise} \end{cases} \qquad (2)$$

$$\alpha_{i,0} = \left(\frac{N}{N - N_i}\right)^b \qquad (3)$$

$$\alpha_{i,1} = \left(\frac{N}{N_i}\right)^b \qquad (4)$$

where p_i is the predicted probability for class i, $r > 0$ is the focusing parameter of the focal loss, N is the total sample number of the training set, N_i is the sample size of class i, $b \in [0, 1]$ is a tunable parameter to control the balancing degree. The bigger the value of b, the more balance will be reached between positive and negative examples for each category. When $b = 1$, the total weights of positive samples are equal to that of negative sample, but this may lead to too many false positive predictions. In our experiments, we set b to 0.5 for a compromise between specificity and sensitivity.

3.   Results

We evaluate our model on the challenge dataset in a 5-fold cross validation. The network is implemented based on the Keras framework, and trained on a workstation with one CPU running at 3.5 GHz, an NVIDIA Quadro k6000 GPU, and 64 Gb of memory. The metrics of the model performance includes the area under the receiver operating characteristic ($AUROC$), the area under the precision-recall curve ($AUPRC$), F_2 score, G_2 score and a specialized metric for the challenge (challenge metric). The detailed definitions of these metrics can be found in the challenge paper [13]. The experimental results are shown in Table 1.

From the results, we can see that the $AUROC$ of the models reached quite high scores, while the $AUPRC$ measures are much lower, indicating that the imbalance of the dataset still have a significant impact on the results of the classification. We note that, though the mean score of our model on the challenge metric is higher than the

top ranked score of the challenge, these results were obtained on the training set only and may be over-optimistic. The scores of F_2, G_2 and the challenge metric are also far from ideal.

Table 1. The experimental results of the proposed classifier in a 5-fold cross validation. (Ave: average, Std: standard deviation)

Fold	AUROC	AUPRC	F_2	G_2	challenge metric
1	0.9504	0.5729	0.5845	0.3330	0.5625
2	0.9470	0.5528	0.5689	0.3214	0.5538
3	0.9491	0.5540	0.5781	0.3258	0.5780
4	0.9461	0.5500	0.5648	0.3054	0.5220
5	0.9502	0.5779	0.5863	0.3343	0.5341
Ave	0.9486	0.5615	0.5765	0.3240	0.5501
Std	0.0019	0.0129	0.0095	0.1116	0.0223

4.   Conclusions

We took part in the 2020 PhysioNet - CinC Challenge (in the team *ECGMaster*) and proposed a novel neural network for the problem of multi-label classification of 12-lead ECG signals. The class-wise attention enable network to learning an individual attention for each class and avoid the conflicts between classes. We also extended the focal loss function by add different weights to positive and negative predictions of each class according to their frequencies. The experimental results show that the model achieved a high score of *AUROC*, but scores of *AUPRC*, F_2 and G_2 were still unsatisfactory, where the data imbalance is one of the main reasons and needs further study.

References

[1] P. Kligfield *et al.*, "Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology," *Journal of the American College of Cardiology,* vol. 49, no. 10, pp. 1109-1127, 2007.

[2] A. Mincholé, E. Zacur, R. Ariga, V. Grau, and B. Rodriguez, "MRI-Based Computational Torso/Biventricular Multiscale Models to Investigate the Impact of Anatomical Variability on the ECG QRS Complex," *Frontiers in Physiology,* vol. 10, 2019.

[3] D. U. J. Keller, F. M. Weber, G. Seemann, and O. Dössel, "Ranking the influence of tissue conductivities on forward-calculated ECGs," *IEEE Trans Biomed Eng,* vol. 57, no. 7, pp. 1568-1576, 2010.

[4] J. T. Flaherty *et al.*, "Influence of respiration on recording cardiac potentials: Isopotential surface-mapping and vectorcardiographic studies∗," *The American Journal of Cardiology,* vol. 20, no. 1, pp. 21-28, 1967.

[5] J. S. Paul, M. R. Reddy, and V. J. Kumar, "A transform domain SVD falter for suppression of muscle noise artefacts in exercise ECG's," *IEEE Trans Biomed Eng,* vol. 47, no. 5, pp. 654-663, 2000.

[6] M. Bickerton and A. Pooler, "Misplaced ECG electrodes and the need for continuing training," *British Journal of Cardiac Nursing,* vol. 14, no. 3, pp. 123-132, 2019.

[7] J. C. Huhta, "60-Hz interference in electrocardiograph," *IEEE Trans. Biomed. Eng,* vol. 20, 1973.

[8] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Comput. Biol. Med.,* vol. 93, pp. 84-92, Feb. 2018.

[9] R. He *et al.*, "Automatic detection of atrial fibrillation based on continuous wavelet transform and 2d convolutional neural networks," *Front. Physiol.,* vol. 9, pp. 1206-, Aug. 2018.

[10] R. J. Martis, U. R. Acharya, K. Mandana, A. K. Ray, and C. Chakraborty, "Cardiac decision making using higher order spectra," *Biomed. Signal Process. Control,* vol. 8, no. 2, pp. 193-203, 2013.

[11] Y. Liu, K. Wang, Q. Li, R. He, and H. Zhang, "Diagnosis of AF based on time and frequency features by using a hierarchical classifier," in *2017 Computing in Cardiology Conference*, 2017.

[12] Y. Liu *et al.*, "Automatic detection of ECG abnormalities by using an ensemble of deep residual networks with attention," in *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting*: Springer, 2019, pp. 88-95.

[13] E. A. Alday, Perez *et al.*, "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Meas,* 2020.

[14] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998-6008.

[15] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. PP, no. 99, pp. 2999-3007, 2017.

Address for correspondence:

Qince Li - E-mail: qinceli@hit.edu.cn
Xinjishu Building 901,
Harbin Institute of Technology,
Xidazhi Street, Nangang District,
Harbin, China