Automatic 12-lead ECG Classification Using Deep Neural Networks

Wenjie Cai, Shuaicong Hu, Jingying Yang, Jianjian Cao

University of Shanghai for Science and Technology, Shanghai, China

Abstract

ECG is the most commonly used diagnostic tool for identifying cardiovascular disease. However, manual interpretation of ECG is inefficient and requires medical practitioners with a lot of training. In this work we proposed two deep learning models to classify ECG automatically. One model had a hybrid architecture of convolutional neural network and recurrent neural network. The other model contained deep residual neural networks. The output layer of both models was activated by a sigmoid function to get classification results. We manually located all the premature beats in each ECG recording and selected 10 s segments which contained at least one premature beat as training samples. Recordings without premature beats were randomly split into 10 s segments. The models were then trained on these ECG segments for 30 epochs with an optimizer of Adam. After training, the model performance was evaluated on the hidden validation set and test set maintained by the challenge organizers. Our team, nebula, achieved a challenge validation score of 0.526, and full test score of 0.109, but was not ranked due to omissions in the submission. The results show potential application value in automatically classifying 12-lead ECG.

1. Introduction

Cardiovascular diseases (CVD) are the number one cause of death in the world, killing more than 17 million lives each year [1]. Electrocardiogram (ECG), reflecting the electrical activity of the heart, is the preferred method for screening and diagnosing CVD. The standard ECG has 12 leads and provides more diagnostic information than single lead ECG. Six of the leads are called “limb leads” which show information of electrical activity transmission on the coronal plane. The other six leads are called “chest leads” which show electrical transmission in the transverse plane. The doctor makes a diagnostic conclusion by checking the ECG beat by beat and lead by lead. The ECG interpretation process is time consuming and tedious, and it is prone to errors. Computerized interpretation of ECG based on expert systems can reduce the workloads but it was reported to have a 5.8% higher error rate than average cardiologists [2]. So more advanced algorithms are required for automated ECG interpretation.

Recently, deep learning has achieved great success in computer vision, natural language processing and speech recognition. With this cutting edge technique, researchers have explored many methods for automatic ECG classification [3-9]. These methods mainly involve convolutional neural networks (CNN), recurrent neural networks (RNN), or a combination of both. Hannun *et al.* developed a deep residual neural network to classify single-lead ECG into 12 classes [3]. Their model got more accurate results than average cardiologists. Faust *et al.* proposed an LSTM model to detect atrial fibrillation and achieved 98.51% accuracy [8]. Xiong *et al.* used 21-layer convolutional recurrent neural network to classify single lead ECGs in the 2017 PhysioNet/CinC Challenge and got $F1$ score of 0.82, which is among the best scores [9]. However, there are still few studies on the classification of 12-lead ECG. This may be due to the lack of appropriate database of 12-lead ECG. The PhysioNet/Computing in Cardiology Challenge 2020 provides more than 43,000 ECG recordings with diagnostic labels [10]. This study aims to develop an automated method for classifying cardiac abnormalities from 12-lead ECGs.

2. Methods

2.1. Data preprocessing

Data values of all recordings were divided by their corresponding amplitude resolutions with the unit of mV. Then all samples were resampled to 500 Hz with fast Fourier transformation. Each lead of every recording was subtracted by its mean value. There are some abnormal spikes [11] with the values greater than 20 mV in the recordings from the dataset of China Physiological Signal Challenge in 2018 (CPSC2018). These spikes were examined and replaced with normal values next to them.

2.2. Data relabeling

Some labels were considered as the same diagnosis according to scoring algorithm provided by the challenge organizer. Complete right bundle branch block (CRBBB) and right bundle branch block (RBBB) were merged as

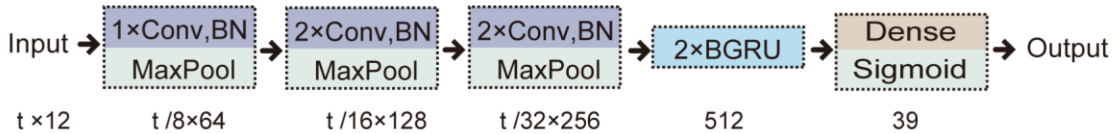

Figure 1. The architecture of Model 1. The tensor dimensions of layer's output are shown and t denotes the samples of the input ECG.

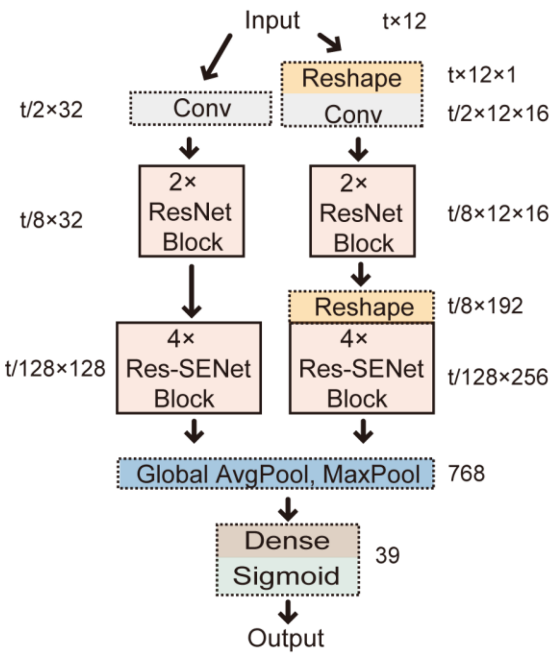

Figure 2. The architecture of Model 2. The tensor dimensions of layer's output are shown and t denotes the samples of the input ECG.

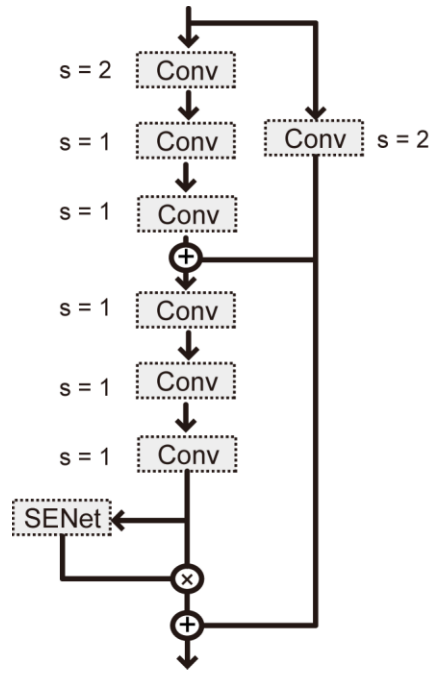

Figure 3. The architecture of Res-SENet. The letter s means strides.

RBBB. Premature atrial contraction (PAC) and supraventricular premature beats (SVPB) were merged as PAC. Ventricular ectopics (VEB) belongs to unscored label, but it has the same medical meaning with Premature ventricular contractions (PVC) and ventricular premature beats (VPB). So these three labels were merged together as PVC. For each unscored label, if the total number of ECG samples with this label was less than 300, we removed this label from all recordings.

2.3. Premature beats locating

For PAC or PVC, the premature beat may only occur once in a very long recording. So we manually located all the premature beats and used segments that contained at least one premature beat for training.

2.4. Feature extraction

The feature extraction process was carried out according to the sample classifier which provided by the challenge organizer. Briefly, all the ECG recordings were band-pass filtered between 0.5-15 Hz to remove baseline wandering and some noise. Then ECG peaks were detected based on Pan-Tomkins algorithm [12]. The statistical features about the peak values and peak intervals including mean, median, standard deviation, variance, skewness, kurtosis were calculated. We extracted features from lead I, II, III and got 12 features from each lead. Counting age and gender, each recording generated a total of 38 characteristics.

2.5. Deep learning models architecture

Two models were proposed in this study. The first model consists of five CNN layers, three max pooling layers, two stacked bidirectional GRU layers and two fully connected layers (Fig. 1). Sigmoid was used as the activation function in the last layer and Rectified Linear Unit (ReLU) was used elsewhere. The second model had two parallel residual neural networks and each used the residual neural network (ResNet) block as a basic block

(Fig. 2). Each parallel Residual neural network contains 1 CNN layer and 6 ResNet blocks, and the last 4 blocks used Squeeze-and-Excitation networks (SENet) to pay more attention to meaningful feature channels (Fig. 3). The outputs of ResNet blocks were compressed by using a global average pooling layer and a global max pooling layer simultaneously. The last layer used a fully connected layer with sigmoid as activation function. For both models, their input dimensions were not fixed at a certain number, so both models accepted ECG data with variable lengths.

2.6. Model training

The challenge data were randomly shuffled and 80% of the data were put into the training set and the remaining were used as our own test set. Our models were trained with the training set using 5-fold cross validation strategy. Although our models accepted data with various length, data with fixed length can take the most advantage of parallel processing power of GPU and reduce a lot of training time. Thus, we used 10 s segments for training in this study. If a recording was labelled as PAC or PVC, the segments that contained at least one premature beat were used. For other recordings, the segments were randomly chosen. These segments were further processed for data augmentation on the fly. Data augmentation techniques included adding random Gaussian noise, combining a random sinusoidal signal [13] and shifting random baseline. The models were trained using Adam algorithm with the learning rate set between 1e-3 and 1e-4. Total epochs were set at 30.

2.7. XGBoost classifier

Extreme gradient boosting (XGBoost) is an optimized decision trees based gradient boosting framework [14]. We trained 24 XGBoost classifiers to predict all the scored labels with normal features and deep features. Specifically, deep features included the output of the second last layer of the first model and the output of the last layer of the second model.

2.8. Model inference

The ECG recordings were pre-processed as described at section 2.1. Then the data were fed into the deep learning models. The average values of two models' outputs were used to make a classification with threshold set at 0.5. If XGBoost classifiers were used, normal features and deep features were fed into 24 separate XGBoost classifiers to make predictions for 24 scored labels.

3. Results

After models training, we evaluated their performance on our own test set. As shown in Table 1, the best method was the ensemble model which combined the decisions of model 1 and model 2. It had a challenge score of 0.560. XGBoost classifier had a challenge score of 0.546 which was lower than that of model 2 and ensemble model.

Table 1. Performance of proposed models on our own test set.

Methods	Challenge Score
Model 1	0.534
Model 2	0.558
Ensemble	0.560
XGBoost	0.546

Table 2. Performance of proposed models on the official validation set.

Methods	Challenge Score
Ensemble	0.526
XGBoost	0.517

Table 3. Performance of the ensemble model on the official test set.

Test set	Challenge Score
Database 1	0.736
Database 2	0.086
Database 3	0.052
All	0.109

The models' performance was further evaluated on the hidden validation set. As shown in Table 2, the ensemble model got a challenge score of 0.526 and XGBoost classifier had a challenge score of 0.517.

Finally, the ensemble model was evaluated on the official full test set. As shown in Table 3, the model received challenge scores of 0.736, 0.086 and 0.052 from three test databases respectively. And our team got the final challenge score of 0.109.

4. Discussion and Conclusions

The results shown in Section 3 indicate that the model with deeper layers is more effective in classifying ECG abnormality than the model with shallower layers. Model 1 is simple and contains 5 CNN layers with a small receptive field. It can't extract and identify complex features. However, this simple model runs fast and can be used as a baseline model. Model 2 has two parallel deep residual neural networks and each has 37 CNN layers. One major difference between the two parallel deep residual neural networks lies in different convolution kernels and different filters. So they have variant receptive fields.

Another difference lies in the first two ResNet blocks of both branches. One branch uses 1 dimensional CNN layers, whereas another branch uses 2 dimensional CNN layers. The 2 dimensional CNN layer is designed to make sure that the same kernel walks through each lead and extract features common to different leads. The ResNet like structure has shortcuts that jump over different layers, which can avoid the problem of vanishing gradients during training and make full use of the features extracted by different convolutional layers. Furthermore, SENet, which won the first place in ILSVRC 2017 classification challenge [15], recalibrates channel-wise feature importance and makes the model more effective. So model 2 got much better performance than model 1.

Although XGBoost classifiers had made use of age, sex, peak related features and deep features, their performance was not as good as model 2. One possible reason is that these features are not good enough to make a classifier with high quality. Further features about frequency, HRV and morphology may be required. The other reason is that we used the default parameters to train XGBoost classifiers, and these default parameters were not optimal, which might lead to overfitting.

There are several limitations in our study. Firstly, an ablation study has not been carried out. Our proposed models are not optimal and could be further tuned. Secondly, we trained our models on 10 s segments and evaluated them on various long samples. It may attenuate the overall performance. Thirdly, the generalization of our models needs to be improved since they had a big difference in performance on different databases of the official test set.

In conclusion, the presented deep learning models showed potential application value in automatically classifying 12-lead ECG.

Acknowledgments

This work was supported by University of Shanghai for Science and Technology under Science and Technology Development Project.

References

[1] Virani SS, Alonso A, Benjamin EJ, *et al*. Heart Disease and Stroke Statistics-2020 Update: A report from the American Heart Association. Circulation, 2020;141(9):e1-e458.

[2] Rautaharju PM. Eyewitness to history: Landmarks in the development of computerized electrocardiography. Journal of Electrocardiology, 2016; 49(1):1-6.

[3] Hannun AY, Rajpurkar P, Haghpanahi M, *et al*. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 2019; 25(1):65-69.

[4] Wang G, Zhang C, Liu Y, *et al*. A global and updatable ECG beat classification system based on recurrent neural networks and active learning. Information Sciences, 2019; 501:523-542.

[5] Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine, 2018; 96:189-202.

[6] Sannino G, De Pietro G. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems, 2018; 86:446-455.

[7] Al Rahhal MM, Bazi Y, Al Zuair M, *et al*. Convolutional neural networks for electrocardiogram classification. Journal of Medical & Biological Engineering, 2018; 38:1014-1025.

[8] Faust O, Shenfield A, Kareem M, *et al*. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals, Computers in Biology and Medicine, 2018; 102:327-335.

[9] Xiong Z, Nash MP, Cheng E, *et al*. ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network, Physiological Measurement, 2018; 39(9):094006.

[10] Perez Alday EA, Gu A, Shah A, *et al*. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge. Physiological Measurement, 2020 (Under Review).

[11] Cai W, Hu D. QRS complex detection using novel deep learning Neural Networks. IEEE Access, 2020; 8:97082-97089.

[12] Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering, 1985; 32(3):230-236.

[13] Tan JH, Hagiwara Y, Pang W, *et al*. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals, Computers in Biology and Medicine, 2018; 94:19-26.

[14] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016; San Francisco, California, USA.

[15] Hu J, Shen L, Albanie S, *et al*. Squeeze-and-excitation networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020;42(8):2011-2023.

Address for correspondence:

Wenjie Cai
School of Medical Instrument and Food Engineering,
University of Shanghai for Science and Technology,
516 Jungong Road, Yangpu Distric, Shanghai, China
wjcai@usst.edu.cn