Automatic Detection and Classification of 12-lead ECGs Using a Deep Neural Network

Wenxiao Jia [1], Xian Xu [1], Xiao Xu [1], Yuyao Sun[1], Xiaoshuang Liu[1]

[1]Ping An Health Technology, Beijing China

Abstract

The objective of the PhysioNet/Computing in Cardiology Challenge 2020 is to identify clinical diagnoses from 12-lead ECG recordings. We developed an end-to-end deep neural network model to classify 27 scored clinical diagnosis from Electrocardiogram *(ECG). The Squeeze and Excitation (SE) layer, which can explicitly model channel-interdependencies within modules and selectively enhance useful features and suppress less useful ones, and ResNet are integrated into a deep neural network, which is called SE-ResNet34 in our paper. We use the one- dimensional convolution to extract the features among different 12-lead ECG channels and the convolution network is a standard 34-layers ResNet. Finally, we also concatenate some demographic features from the ECGs and the deep features from the SE-ResNet34 to identify clinical diagnosis. The evaluation metrics is calculated, which assigns different weights to different classes, according to the similarity between different classes. Our team named PALab ranked 10 out of 41 teams in the official ranking and achieved a challenge validation score of 0.653 and full test score of 0. 359.*

If confirmed in clinical settings, this approach could reduce the rate of misdiagnosed computerized ECG interpretations and improve the efficiency of expert human ECG interpretation.

1. Introduction

The 12-lead ECG plays a critical role in the clinical diagnoses, including varieties of arrhythmias and other cardiac abnormalities. The early detection and classification of cardiac abnormalities can tremendously increase the chance of successful treatment.

Automatic detection and classification of cardiac abnormalities can assist physicians in the diagnosis of the growing number of ECG recorded. Over the last decade, there have been increasing numbers of attempts to identify 12-lead ECG clinical diagnosis, mainly including a variety of traditional machine learning methods, requiring extensive data pre-processing, feature engineering or handcrafted rules[1-2]. However, substantial algorithms have gained more attention by a specific class of models known as deep neural networks (DNN) in the past five years. Many of these algorithms have the potential for more accurate identification of cardiac abnormalities[3-6]. DNN consists of multiple processing layers, with each layer being able to learn increasingly abstract, higher-level representations of the input data relevant to specific tasks, which makes them particularly well suited to interpret ECG. Therefore, some authors integrate features from domain knowledge into DNN model to obtain a better accuracy and interpretation [7-8].

The lack of appropriate data and well-defined evaluation has limited generalizability of automatic interpretation algorithm for 12-lead ECGs [9]. Much of the previous work used data from single, small, or relatively homogeneous datasets, which is limited by the small number of patients and rhythm episodes present in the dataset. These models perform well on the training set, but often fail to perform well on the external test set. The PhysioNet/Computing in Cardiology Challenge 2020 provides an opportunity to address this problem by providing data from a wide set of sources. This paper attempt to design a robust model that automatically identify the cardiac abnormality in each 12-lead ECG recording by using the challenge ECG data [10].

In this study, we developed a DNN model, more specifically SE-ResNet34, to identify 27 scored clinical diagnosis from raw 12-lead ECG using four different training dataset consisting of 43101 patients. In summary, we demonstrate that end-to-end deep learning approach can classify a broad range of distinct arrhythmias from 12-lead ECGs. If confirmed in clinical settings, we believe that this approach has the potential to improve the accuracy, efficiency, and scalability of ECG interpretation.

2. Methods

2.1. Datasets Reprocessing

Our research is based on the data from PhysioNet/Computing in Cardiology Challenge 2020. Figure 1. shows a typical example of 12-lead ECG data with atrial fibrillation (AF). As we can see, AF is an abnormal heart rhythm (arrhythmia) characterized by the

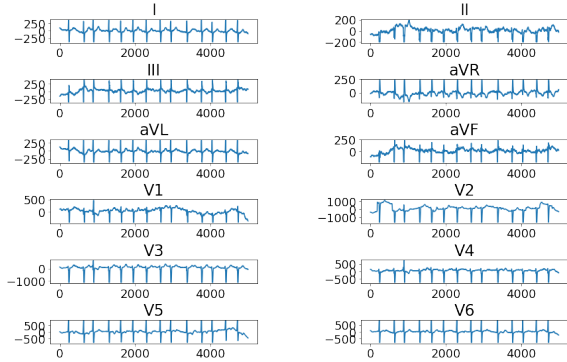rapid and irregular beating of the atrial chambers of the heart.

Figure 1. Typical example of raw 12-lead ECG with atrial fibrillation.

We ignored the INCART dataset which only has 74 recordings of 30 minutes. In the remained datasets, each recording has an uncertain length ranging from 6 to 60 seconds, which is sampled at 500Hz. For convenient model training with non-identical length of ECG recordings, each recoding has been truncated into length of 10 seconds (5000 sample points). In addition, if the data sampling frequency is not 500Hz, we should resample the data into 500Hz. As showed Figure 2, if the length of original recording was less than 5000 points, we padded the recording into 5000 points by assigning zero values in the end of data. If the length of original recording was longer than 5000 points, we truncated the recording into 5000 points by discarding values in the end of data.
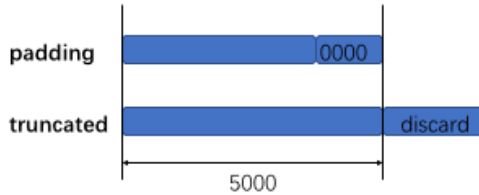

Figure 2. Preprocessing of different data lengths

A typical ECG heartbeat is characterized by a recurrent sequence of waves including P, QRS and T waves which represent the depolarization of the atria and ventricles, followed by repolarization of the ventricles. In this study, the ECG signals have not been filtered in the pre-processing stage because of two main factors. On one hand, in this database, all the 12 leads recordings are used and the amount of the data is large. Using the original data rather than filtering them, the computation cost is significantly reduced. On the other hand, as mentioned above, most of arrhythmia detection algorithms have applied filtering to process ECG signals [1-2]. However, in this study, the proposed model shows good performance in anti-noise interference due to no filtering, demonstrating a potential for practical applications.

In this study, PVC and VPB, CRBBB and RBBB，PAC and SVPB are classified as the same type of diagnosis.

2.2. Model development

A 34-layer ResNet was developed for the ECG classification task (Figure 3). In order to improve the efficiency of traditional CNN, the design includes 17 sequential skip connections [11]. In each block, the same operations were performed. As shown in Figure 2, the modules consist of one-dimensional convolutional (1D Conv) layers, batch normalization (BN) layers, rectified linear units (Relu) of activation layer and SE layer.

The convolution layer was the major feature learning component of the CNN, which involved a 7x1 filter with trainable weights that sliding across the signals to extract features from the waveform. When significant features were detected, the filters were activated by adapting the weights. Providing the labelled data, the model was able to learn the significant features that represented different diagnosis classes.
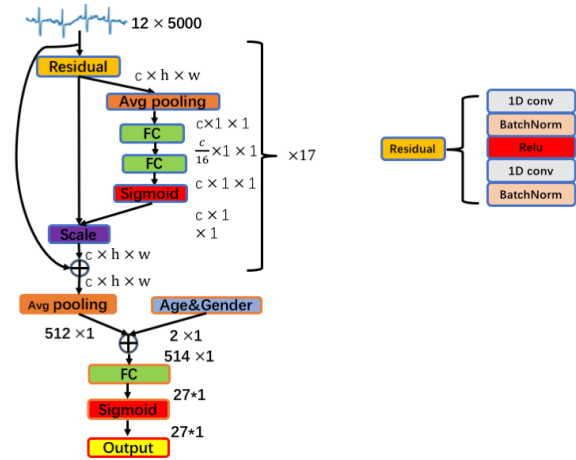

Figure 3. The overall structure of the neural network architecture with repeated 34 1D convolution with skip connections

Each learned filter was operated with a local receptive field, so that each unit of the transformation output was unable to exploit contextual information outside of this region. In order to tackle the issue of exploiting channel dependencies, we added SE layer (Squeeze-and-Excitation block)[12].

Firstly, we squeezed global spatial information into a channel descriptor by using global average pooling to generate channel-wise statistics. Formally, a statistic $z \in \mathbb{R}^C$ was generated by shrinking $\mathcal{U}$ through spatial dimensions $H \times W$, where the c-th element of z is calculated by :

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{w} u_c(i,j) \qquad (1)$$

Here $\mathcal{U} = [u_1, u_2, ..., u_c]$ was the output of previous layer, $\mathcal{U} \in \mathbb{R}^{H \times W \times C}$. The transformation output $\mathcal{U}$ could be interpreted as a collection of the local descriptors, which were expressive for the whole signal.

Secondly, to make use of the information aggregated in the squeeze operation and fully capture channel-wise dependencies, we employed a simple gating mechanism with a sigmoid activation. It could learn a non-mutually-exclusive relationship since we would like to ensure that multiple channels were allowed to be emphasized opposed to one-hot activation:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (2)$$

where δ refers to the Relu function, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$, $r = 16$. To limit model complexity and enhance generalization, we parameterized the gating mechanism by forming a bottleneck with two fully connected (FC) layers around the non-linearity, i.e. a dimensionality-reduction layer with parameters W_1 with reduction ratio r, a ReLU and then a dimensionality increasing layer with parameters w_2. The final output of the block was obtained by rescaling the transformation output $\mathcal{U}$ with the activations:

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (3)$$

where $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_c]$ and $F_{scale}(u_c, s_c)$ refers to channel-wise multiplication between the feature map $u_c \in \mathbb{R}^{H \times W}$ and the scalar s_c.

In order to integrate the demographic information, we also concatenate the age and sex with the features from convolution layers. Following this, a fully connected layer was used to transform the features to a 27x1 vector of numerical values, which corresponded to the outputs for each class. A sigmoid function was used to represent these values as a probability by normalizing them between 0 and 1. The network took segments of length of 10 seconds as input, and produced a prediction for each segment. The model output was a probability for each class, and the predicted class was the one which probability greater than threshold.

We randomly divided the model development data into two parts: 80% for model training, and 20% for internal validation. The validation set was used to tune the parameters. We implemented all the models with Pytorch 1.1, and trained them on machines with NVIDIA TESLA P100 GPUs. The adaptive momentum estimation (Adam) optimizer, with a learning rate of 0.001, was used to optimize the network parameters.

2.3.   Baseline Model

In order to compare model performance for different parameters setting, we also designed several baseline models. Inspired by the concept that the more network layers, the better model performance, we designed the ResNet50 model, which is same as ResNet34 model but with 50 layers. In addition, we also utilize LightGBM classifiers to identify clinical diagnosis. Firstly, we extracted features from ResNet50 model, including the last layer and penultimate layer of the network layer. Then we concatenated these features as the final features. Finally, these features were feed into the input for the LightGBM classifiers.

2.4.   Evaluation metrics

The challenge metric awards partial credit to misdiagnoses that result in similar outcomes as the true diagnosis as judged by cardiologists. It originates from the institution that some misdiagnoses are more harmful than others and should be scored accordingly. Moreover, it reflects the fact that confusing some classes is much less harmful that confusing other classes. Figure 4 shows the Reward matrix W for the diagnoses scored in the Challenge.
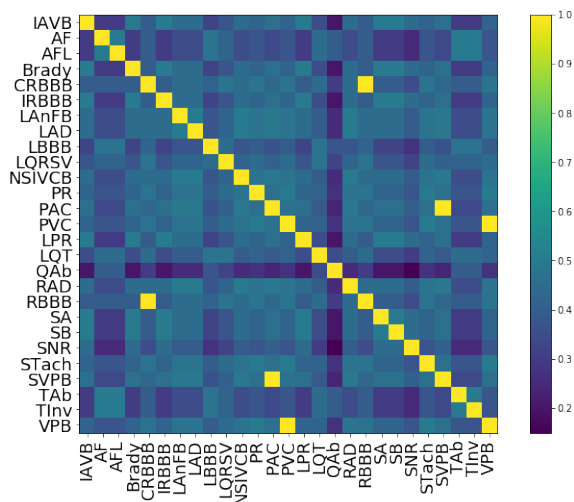

Figure 4. The Reward matrix W for the diagnoses scored in the Challenge, where columns are the true label and columns and rows are the predicted label

3.   Result

For this year's Challenge, organizers developed a new scoring metric that awards partial credit to misdiagnoses that result in similar outcomes or treatments as the true diagnoses as judged by our cardiologists. The predicted performance of different methods on intermediate validation set are illustrated in Table 1. As we can see, the **S**E-Resnet34 shows model outperformance (0.653) compared with other methods. In addition, the SE layer is critical for clinical diagnosis of 12-lead ECG. Table 2. shows the final testing set result of our model (SE-Resnet34) in the official phase, where the challenge score is 0.359 on full test set.

Table 1. The performance of different methods on intermediate validation set.

Model	AUC	Challenge Score
Rsenet50	0.938	0.594
Rsenet50+lgb	0.951	0.624
Resnet34	0.959	0.630
SE-Resnet34	**0.967**	**0.653**

Table 2. The model performance on the testing set in the official phase

Model	Challenge Score
Test Database 1 Score	0.836
Test Database 2 Score	0.623
Test Database 3 Score	0.144
Full Test Set Score	**0.359**

4.   Discussion

We proposed a new deep learning model named SE-ResNet34 to identify 27 clinical diagnosis from 12-lead ECGs. The model showed perfect classification performance on the PhysioNet/Computing in Cardiology Challenge 2020. Specifically, SE-ResNet34 achieved high and stable performance score measured by challenge metric (0.359) in the full test set. Considering the fact that manual interpretation of the electrocardiogram is time-consuming, and requires skilled personnel with a high degree of training, this model can easily be applied to assist the cardiology doctors to identify the clinical diagnosis.

It can be inferred that the challenge score is related to the incidence rates of each class, which is also proved in our previous study [13]. Therefore, given a set of clinical diagnosis with different incidence rate, we can adjust the cut-off value to dichotomize their predicted likelihoods for better challenge score. Considering the incidence rates significantly change over different classes, using different cutoffs may have a better potential to lead to better total utility score, which has been validated in our validation set. However, it is a little improvement in hidden testing set, which is probably overfitted in validation set and needs further confirmation in the future.

Although SE-ResNet34 model can provide accurate classification of 12-lead ECGs, it still has some limitations. The generalization and stability of the proposed model needs to be systematically evaluated with more data in clinical reality.

References

[1] Kropf M, Hayn D, Schreier G. ECG classification based on time and frequency domain features using random forests. 2017 Computing in Cardiology (CinC). IEEE, 2017: 1-4.

[2] Smíšek R, Hejč J, Ronzhina M, et al. SVM based ECG classification using rhythm and morphology features, cluster analysis and multilevel noise estimation. 2017 Computing in Cardiology (CinC). IEEE, 2017: 1-4.

[3] Awni, Y, Hannun et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. J. Nature Medicine, 2019.

[4] Attia Z I , Kapa S , Lopez-Jimenez F , et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature medicine, 2019, 25(1):70-74.

[5] Attia Z I, Noseworthy P A, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet, 2019, 394(10201): 861-867.

[6] Raghunath S, Cerna A E U, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. Nature Medicine, 2020: 1-6.

[7] Hong S , Xiao C , Ma T , et al. MINA: Multilevel knowledge-guided attention for modeling electrocardiography signals. Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19. 2019.

[8] Li X, Qian B, Wei J, et al. Domain knowledge guided deep atrial fibrillation classification and its visual interpretation. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 129-138.

[9] Strodthoff N , Wagner P , Schaeffter T , et al. Deep learning for ECG analysis: benchmarks and insights from PTB-XL. 2020.

[10] Erick A. Perez Alday, Annie Gu, Amit Shah, Chad Robichaux, An-Kwok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, Qiao Li, Ashish Sharma, Gari D. Clifford, Matthew A. Reyna. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiol. Meas.

[11] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[12] Hu J , Shen L , Albanie S , et al. Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[13] Li X, Xu X, Xie F, et al. A Time-phased machine learning model for real-time prediction of sepsis in critical care. Critical Care Medicine, 2020.

Address for correspondence: No.1-3 Xinyuan South Road, Chaoyang District, Beijing, China
My Name: Wenxiao Jia
My E-mail address:jiawenxiao502@pingan.com.cn