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

Wenxiao Jia¹, Xian Xu¹, Xiao Xu¹, Yuyao Sun¹, Xiaoshuang Liu¹

¹Ping 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 develop an end-to-end deep neural network model to classify 27 scored clinical diagnosis from Electrocardiogram (ECG). The model is a modified version of Squeeze and Excitation (SE) network, which can explicitly model channel-interdependencies within modules and selectivity enhance useful features and suppress less useful ones. The SE network and ResNet are integrated into a deep neural network, which is called SE-ResNet. We use the 1 dimensional convolutional filter to extract the features among the different 12-lead ECG channels and the convolutional network is a standard 34-layers ResNet. The SE network is able to obtain the channel attention to explore the spatial enhancement. Finally, we also concatenate some statistical features from the ECGs and the deep features from the SE-ResNet 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 name is PAlab and the best model performance of our submission is 0.653 in the hidden testing set during official phase.

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 algorithm that automatically identify the cardiac abnormality present in each 12-lead ECG recording by using the challenge ECG data.

In this study, we developed a DNN model, more specifically SE-ResNet, to identify 27 scored rhythm classes from raw 12-lead ECG using four different training dataset consisting of 43101 patients. During official phase, the best model performance in the internal validation data set offline and in the hidden testing set online is 0.682, 0.653 respectively. 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

Our research is based on the data from PhysioNet/Computing in Cardiology Challenge 2020. The ECG data were included from four distinct resources consisting of 43101 patients. The first source is from the China Physiological Signal Challenge in 2018 (CPSC2018), which consists of two sets of 6877 and 3453 of 12-ECG recordings lasting from 6 seconds to 60 seconds and was sampled at 500 Hz. The second source set is from St Petersburg INCART 12-lead Arrhythmia Database from 32 Holter records. Each record is 30 minutes long, and was sampled at 257 Hz. The third source from the Physikalisch Technische Bundesanstalt (PTB) comprises two public databases: the PTB Diagnostic ECG Database and the PTB-XL [9]. The first PTB database contains 549 records and was sampled at 1000 Hz. The PTB-XL contains 21837 clinical 12-lead ECGs of 10 second length with a sampling frequency of 500 Hz. The fourth source is a Georgia database from South-eastern United States, including 10344 clinical 12-lead ECGs of 10 second length with a sampling frequency of 500 Hz. Figure 1. shows a typical example of 12-lead ECG data with atrial fibrillation and nonspecific intraventricular conduction disorder.

![Figure 1. Typical example of raw 12-lead ECG with atrial fibrillation.](image)

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(Fs=500Hz). For convenient model training with non-identical length of ECG recordings, each recording has been truncated into length of 10 seconds. If the length of original recording was less than 10 seconds, we padded the recording into 10 seconds by assigning zero values at the beginning period.

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 2). In order to improve the efficiency of traditional CNN, the design includes 17 sequential skip connections. In each block, the same operations were performed. As shown in Figure 2, the modules consist of 1-dimensional convolutional (1D Conv) layers, batch normalization (BN) layers, rectified linear units (ReLU) of activation layer and SELayer.

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 2. The overall structure of the neural network architecture with repeated 34 1D convolutions with skip connections.](image)

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 SElayer(Squeeze-and-Excitation block)[11].

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:

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

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 emphasised opposed to one-hot activation:

$$s = F_s(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1z)) \quad (2)$$

where $\delta$ refers to the ReLU function, $W_1 \in \mathbb{R}^{C \times C}$ and $W_2 \in \mathbb{R}^{C \times c}$, $r = 16$. 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:

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

where $\hat{x} = [\hat{x}_1, \hat{x}_2, ..., \hat{x}_c]$ 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 $s_c$.

To produce a prediction, a fully connected layer was used to transform the outputs from the convolution and pooling layers to a 27x1 vector of numerical values, which corresponded to the outputs for each channel. A sigmoid function was used to represent these values as a probability by normalizing them between 0 and 1.

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.

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.

### 2.3. 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 than confusing other classes. We also assessed prediction performance with area under the receiver operating characteristics curve (AUROC), area under the precision recall curve (AUPRC) et al.

### 3. Result

For this year’s Challenge, organizer 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 prediction performance for different methods for multiple diagnosis prediction are illustrated in Table 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>Challenge Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rsenet50</td>
<td>0.938</td>
<td>0.629</td>
</tr>
<tr>
<td>Rsenet50+lgb</td>
<td>0.951</td>
<td>0.650</td>
</tr>
<tr>
<td>Resnet34</td>
<td>0.959</td>
<td>0.670</td>
</tr>
<tr>
<td>SE-Resnet</td>
<td>0.967</td>
<td>0.682</td>
</tr>
</tbody>
</table>

In order to evaluate the model performance comprehensively, we also calculated multiple metrics by comparing the labels and outputs in validation set. Table 2 shows the result of our model (SE-Resnet).

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUROC</td>
<td>0.967</td>
</tr>
<tr>
<td>AUPRC</td>
<td>0.660</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.435</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.601</td>
</tr>
<tr>
<td>Fbeta-measure</td>
<td>0.645</td>
</tr>
<tr>
<td>Gbeta-measure</td>
<td>0.384</td>
</tr>
<tr>
<td>Challenge metric</td>
<td>0.682</td>
</tr>
</tbody>
</table>

### 4. Discussion

We proposed a new deep learning model named SE-
ResNet to identify 27 clinical diagnosis from 12-lead ECGs. The model showed good classification performance on the PhysioNet/Computing in Cardiology Challenge 2020. Specifically, SE-ResNet achieved high and stable performance measured by challenge metric on both testing set offline (0.682) and online (0.653) respectively. 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 [12]. 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-ResNet 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


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