A Novel Multi-Scale Convolutional Neural Network for Arrhythmia Classification on Reduced-lead ECGs

Pan Xia[1,2], Zhengling He[1,2], Yusi Zhu[1,2], Zhongrui Bai[2,3], Xianya Yu[1,2], Yuqi Wang[2], Fanglin Geng[1,2], Lidong Du[2], Xianxiang Chen[2], Peng Wang[2], Zhen Fang[1,2]

[1]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China
[2]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
[3]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

Abstract

The PhysioNet/Computing in Cardiology Challenge 2021 focused on exploring the utility of reduced-lead ECGs for arrhythmia classification. We proposed a novel multi-scale convolutional neural network that can classify 30 scored arrhythmias from 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs. The proposed network was achieved by multiple branch networks to effectively extract the pathological information from the "detail" scale to the "approximation" scale via convolution kernels (filters). The backbone of each branch network was a carefully designed 2-D convolutional neural network (CNN) with residual connection and attention mechanism, and it can adapt to multi-lead ECGs as input. The first 10 seconds of records from corresponding leads were extracted and preprocessed as input for end-to-end training, and the prediction probabilities of 30 categories were output. The proposed algorithms were firstly evaluated on officially published datasets of 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs, and we achieved 5-fold cross-validation scores of 0.719, 0.687, 0.698, 0.695, and 0.678 by using the challenge metric. Finally, our team, AIRCAS_MEL1, achieved challenge validation scores of 0.63, 0.57, 0.58, 0.57, and 0.56 for 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs, respectively. Results showed that reduced-lead ECGs can indeed capture a wide range of arrhythmia diagnostic information, and some (4-lead, 3-lead) are even comparable to 12-lead ECGs in some limited contexts. Those reduced-lead ECGs (4-lead, 3-lead) are promising in developing smaller, lower-cost, and easier-to-use diagnostic devices that are comparable to the 12-lead standard diagnostic system.

1. Introduction

The goal of PhysioNet/Computing in Cardiology Challenge 2021 was to explore the utility of reduced-lead ECGs for arrhythmia classification. The participants were asked to build an algorithm that can classify 30 cardiac abnormalities from 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs. In this work, we proposed a novel multi-scale convolutional neural network with residual connection and attention mechanism. The proposed networks consisted of multiple branch networks, each branch network focused on pathological information of different time scales, and the range of receptive field was extended as far as possible within each branch, ultimately, the pathological information from the "detail" scale to the "approximation" scale was systematically perceived by series of 2-D convolution kernels (filters) with various sizes. The first 10 seconds of ECGs records from corresponding leads were extracted as input for end-to-end training, and the prediction probabilities of 30 categories were the outputs of the models. The proposed algorithms were firstly 5-fold cross-validated on officially published datasets of 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs. Ultimately, the performances of the proposed algorithms were evaluated on the officially hidden validation set.

2. Methods

2.1. Datasets and preprocessing

The publicly available datasets consisted of a total of 88243 subjects from 5 different sources. Only 30 categories will be considered in the final evaluation score. The sampling rate of the data set varied from 257 to 1000 Hz, the original ECG signals were firstly down-sampled to 100 Hz to reduce memory consumption and speed up

model training. In addition, a band pass filter with a cut-off frequency of 0.05 to 35 Hz was designed to eliminate baseline drift and high-frequency noise. The first 10 seconds of records were maintained, and then data will be truncated or expanded with 0 to a consistent length. Z-score normalization was applied to normalize signals in all leads.

2.2. Model architecture

A network consisting of three branch networks was built. Like our previous work [1], multi-lead ECG signals will still be regarded as 2-D input with dimensions of "lead" and "sample". The input shape of each branch network was $N*1000$, $N*500$, $N*250$, where N is the number of corresponding ECG lead. The dimensionality reduction of the "sample" dimension was achieved by directly down-sampling. Two specially designed blocks were stacked to form the backbone of each branch network, as shown in Figure 1. After the global average pooling (GAP) operation, the output feature maps of three branch networks were concatenated into the whole feature map. A fully connected layer was adopted to connect the whole feature map and the prediction probabilities of 30 categories.

•**Convolution kernel**: For each branch, a large convolution kernel size in the dimension of "sample" was adopted to the first 2 blocks to enhance the perception field, and the size of the last 3 blocks will become 3, 9, 27, which can ensure that the pathological information at different time scales was effectively extracted. Meanwhile, the convolution kernel sizes in the dimension of "lead" was set to 1 for the first 2 blocks, and the size for the last 3 blocks will be adjusted to [5, 5, 5], [3, 3, 3], [2, 2, 2], [1, 2, 2], [1, 1, 2] for 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead models, respectively, which makes the convolution filters just perceive the information among lead at first 2 blocks and adequately perceive cross-lead information at last 3 blocks. The above-mentioned convolution kernel configuration ensured the equivalence of the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead models in exploring the utility of reduced-lead ECGs.

•**Stride**: The step size of two adjacent convolutions was carefully controlled. The first 2 blocks were allocated a relatively large stride of 2 to significantly reduce the dimensionality of the feature map passed to the next block in each branch, and a stride of 1 was used in the last 3 blocks in each branch.

•**Batch normalization and spatial dropout**: Batch normalization (BN) [2] has been demonstrated to speed up the convergence of the network and alleviate the risk of over-fitting. Spatial dropout [3] is a widely used strategy to prevent over-fitting, which randomly drops the feature maps at a preset rate, and is more suitable for convolutional layers than the standard dropout strategy [4].

•**Attention module**: An attention module called Squeeze-and-Excitation (SE) [5] was added to those blocks, which is a channel-wise attention mechanism. Each feature map will be calculated with a channel-wise weight matrix, and
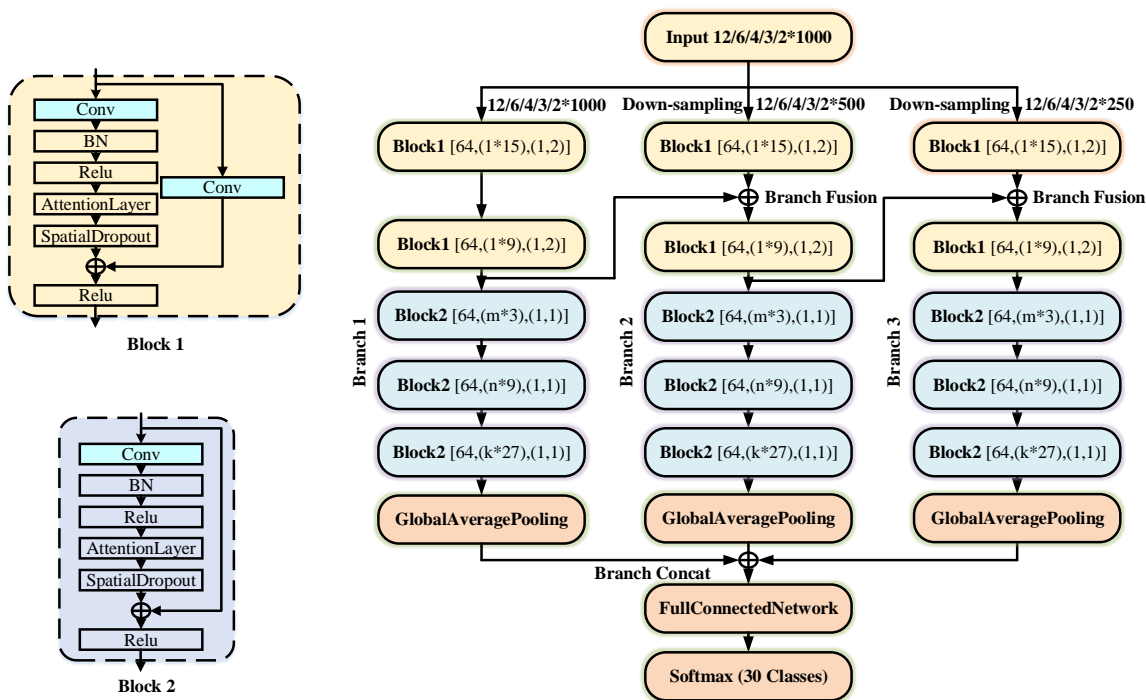

Figure 1. The proposed model architecture. The meaning of the parameters in brackets is [filter number, filter size, stride size]. The values of [m, n, k] are [5, 5, 5], [3, 3, 3], [2, 2, 2], [1, 2, 2], [1, 1, 2] for 12-lead, 6-lead, 4-lead, 3-lead, 2-lead models, respectively.

then the dot product of this matrix and the original map constitutes the reweighted feature map. This weight matrix reflects the importance of individual feature map, and these weights will eventually be automatically learned by the network in the process of gradient back-propagation.

•**Activation**: *Relu* is adopted to all layers except for the output layer, and the output layer uses *sigmoid* because ECG signal in the real world can contain more than one type of arrhythmia, so it's a multi-label classification problem. *sigmoid* can independently map the output to a probability value in the range of 0 to 1.

•**Loss function**: The 30 scored categories will be encoded in binary form as $label = [l_1, l_2,...,l_{30}]$, and for each record, if their category exists in the 30 scored categories, the corresponding binary bit is set to 1. A weighted binary cross-entropy loss was designed as the optimization target, which is defined as follows.

$$CE = -\frac{1}{m}\sum_{i=1}^{m} w_i * (y_i log(h_i) + (1 - y_i)log(1 - h_i)) \quad (1)$$

Where y_i and h_i are the true label and prediction score for the category i, respectively, and the total average is considered as the final loss. Weights are assigned to each class to alleviate the problem of class imbalance, which is defined as follows.

$$w_i = log_2(\frac{1}{n_i/N + e}) \quad (2)$$

Where n_i is the frequency of category i, N is the total number of samples, and e is set to 0.01 to prevent division by 0.

The above model was implemented using *Keras* with *Tensorflow* backend.

2.3. Model training

All publicly available data were randomly divided into training and testing at a ratio of 4:1, and the training set was further divided into training and validation datasets at a ratio of 9:1. *Adam* with an initial learning rate of 0.001 was applied for optimization. Hyper parameters of the network (dropout rate, convolution kernel size, etc.) were adjusted empirically according to the performance on the validation dataset to achieve optimal performance. Methods of early-stopping and reducing learning rate with a ratio of 0.5 during training were also adopted to alleviate over-fitting.

3. Results and discussion

We evaluated our proposed algorithms through 5-fold cross-validation on officially published datasets (including 88243 subjects). The average evaluation scores (challenge metric) on the publicly available dataset and the scores that our team (AIRCAS_MEL1) obtained on the official hidden

validation set were shown in Table 1.

Table 1. Challenge metric on the hidden validation set.

Dataset/Challenge metric	12-lead model	6-lead model	4-lead model	3-lead model	2-lead model
5-fold cross validation	0.719	0.687	0.698	0.695	0.678
Validation Set	**0.63**	**0.57**	**0.58**	**0.57**	**0.56**

A comparative experiment was conducted to analyze the effect of convolution kernel configuration (last 3 blocks of each branch) on the model performance elevation: (a) the convolution kernels with fixed sizes (3, 3, 3), the range of perception field were extended by multiplying the basic size (the value is 3 in this experiment) by 1; (b) the convolution kernels with variable sizes (3, 6, 12), the range of perception field was extended by multiplying the basic size by 2; (c) the convolution kernels with variable sizes (3, 9, 27), the range of perception field was extended by multiplying the basic size by 3. The cross-validated scores on publicly available datasets were shown in Table 2.

Table 2. Comparison of cross-validation results of the effect of convolution kernel configuration on model performance.

Convolution kernels / Challenge metric (5-fold)	12-lead model	6-lead model	4-lead model	3-lead model	2-lead model
Fixed size (3, 3, 3)	0.666	0.634	0.640	0.629	0.614
Variable sizes (3, 6, 12)	0.690	0.653	0.660	0.656	0.641
Variable sizes (3, 9, 27)	**0.700**	**0.663**	**0.672**	**0.672**	**0.653**

The results showed that the performance of models using convolution kernels with variable size was better than those with fixed size. The possible reason is that the convolution kernels with variable sizes can capture pathological information at different scales. More importantly, the model taking convolution kernels with variable sizes (3, 9, 27) has the best performance among all models. It's mainly because these convolution kernels can effectively perceive the pathological information from the "detail" scale to the "approximation" scale.

In addition, an ablation study was also adapted to individually evaluate the performance of branch models. The baseline models refer to that model taking convolution kernels with variable sizes (3, 9, 27) in the "sampling" dimension. The cross-validation scores of individual branch models on publicly available datasets were shown in Table 3.

Table 3. Comparison of cross-validation results of individual branch models using ablation study.

Strategy/ Challenge metric (5-fold)	12-lead model	6-lead model	4-lead model	3-lead model	2-lead model
First branch models	0.657	0.635	0.633	0.632	0.615
Second branch models	**0.664**	**0.637**	**0.636**	**0.631**	**0.621**
Third branch model	0.643	0.617	0.616	0.615	0.598
Baseline models	**0.700**	**0.663**	**0.672**	**0.672**	**0.653**

The results showed that the performance of the second branch network was best among all branch networks. It can be explained that this branch network can effectively capture more pathological features or most pathological features of arrhythmia concentrated on this scale.

In addition, an ablation study was also adapted to analyze the effect of the attention mechanism on the improvement of model performance. The baseline models refer to that model taking convolution kernels with variable sizes (3, 9, 27) in the "sampling" dimension. The cross-validation scores on publicly available datasets were shown in Table 4.

Table 4. Comparison of cross-validation results of the effect of attention mechanism on model performance using ablation study.

Strategy/ Challenge metric (5-fold)	12-lead model	6-lead model	4-lead model	3-lead model	2-lead model
Baseline	0.700	0.663	0.672	0.672	0.653
Baseline+Attention (proposed)	**0.719**	**0.687**	**0.698**	**0.695**	**0.678**

We can observe that using an SE attention layer can improve model performance. The general explanation is that the attention layer can help the network know where to emphasize or suppress by automatically learning the information flow from the above layer.

All in all, we have verified in this study that multi-scale or multi-branch networks can improve model performance. The relationship between pathological information and convolutional kernel sizes (scales) will be studied in future work.

4. Conclusion

We proposed a novel multi-scale convolutional neural network to identify 30 scored arrhythmias from 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs in an end-to-end training manner. A comparative experiment was conducted to analyze the effect of convolution kernel configuration (the last 3 blocks of each branch) on model performance. Results showed that models with variable convolution kernel size (3, 9, 27) have the best performance among all models. The effect of the attention mechanism on model performance elevation was also analyzed, and the results showed that the introduction of the attention mechanism has a positive effect on the improvement of model performance. In addition, the performances of branch models were individually evaluated on publicly available datasets, and the results showed that the second branch network has the best performance among all branch networks. The proposed models gained 5-fold cross-validation scores of 0.719, 0.687, 0.698, 0.695, and 0.678 on the publicly available 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs datasets, and they also achieved challenge validation scores of 0.63, 0.57, 0.58, 0.57, and 0.56, respectively. Results showed that reduced-lead ECGs can indeed capture a wide range of arrhythmia diagnostic information, and some (4-lead, 3-lead) are even comparable to 12-lead ECGs in some limited contexts. Those reduced-lead ECGs (4-lead, 3-lead) are promising in developing smaller, lower-cost, and easier-to-use diagnostic devices that are comparable to the 12-lead standard diagnostic system.

References

[1] He, Zhengling, et al. "A Novel Convolutional Neural Network for Arrhythmia Detection From 12-lead Electrocardiograms." *2020 Computing in Cardiology. IEEE*, 2020.

[2] Ioffe S and Szegedy C, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[3] Tompson J, Goroshin R, Jain A, et al., "Efficient object localization using convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648-656.

[4] Srivastava N, Hinton G, Krizhevsky A, et al., "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[5] Hu J, Shen L, and Sun G, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132-7141.

Address for correspondence:

Zhen Fang
Aerospace Information Research Institute, Chinese Academy of Sciences, No. 9, North Zhongguancun, Haidian District, Beijing.
zfang@mail.ie.ac.cn.