

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 Wang², Fanglin Geng^{1,2}, Lidong Du², Xianxiang Chen², Peng Wang², Zhen Fang^{1,2}

¹School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

²Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

³School 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. Our team, AIRCAS_MELI, 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 a series of 2-D 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 inputs for end-to-end training, and the prediction probabilities of 30 categories were outputs. Finally, the proposed algorithms were evaluated on the hidden test set, and our classifiers received scores of 0.38, 0.33, 0.37, 0.43, and 0.38 (ranked 21th, 23th, 20th, 16th, and 20th out of 39 teams) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions of the hidden test set with the Challenge evaluation metric.

1. Introduction

The PhysioNet/Computing in Cardiology Challenge 2020 focused on identifying 27 arrhythmias from 12-lead ECG recordings [1]. Building on last year’s Challenge, the goal of PhysioNet/Computing in Cardiology Challenge 2021 was to explore the utility of reduced-lead ECGs for arrhythmia classification [2]. 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 2-D convolutional neural network with residual connection and attention mechanism to achieve the target. The first 10 seconds of ECGs records from corresponding leads were extracted as inputs for end-to-end training, and the prediction probabilities of 30 categories were the outputs of the models. Our final selected entry was firstly 5-fold cross-validated on the public training set by using the Challenge evaluation metric and then was scored on the hidden validation set by the Challenge organizers with the Challenge evaluation metric. Ultimately, our final selected entry was scored as well as ranked on the hidden test set with the Challenge evaluation metric.

2. Methods

2.1. Datasets and Preprocessing

The public training set consisted of a total of 88253 subjects from 6 different sources [3-8]. 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 [9], 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 in 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 made 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.

•**Attention Module:** An attention module called Squeeze-and-Excitation (SE) [10] 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 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 $label = [l_1, l_2, \dots, l_{30}]$, 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

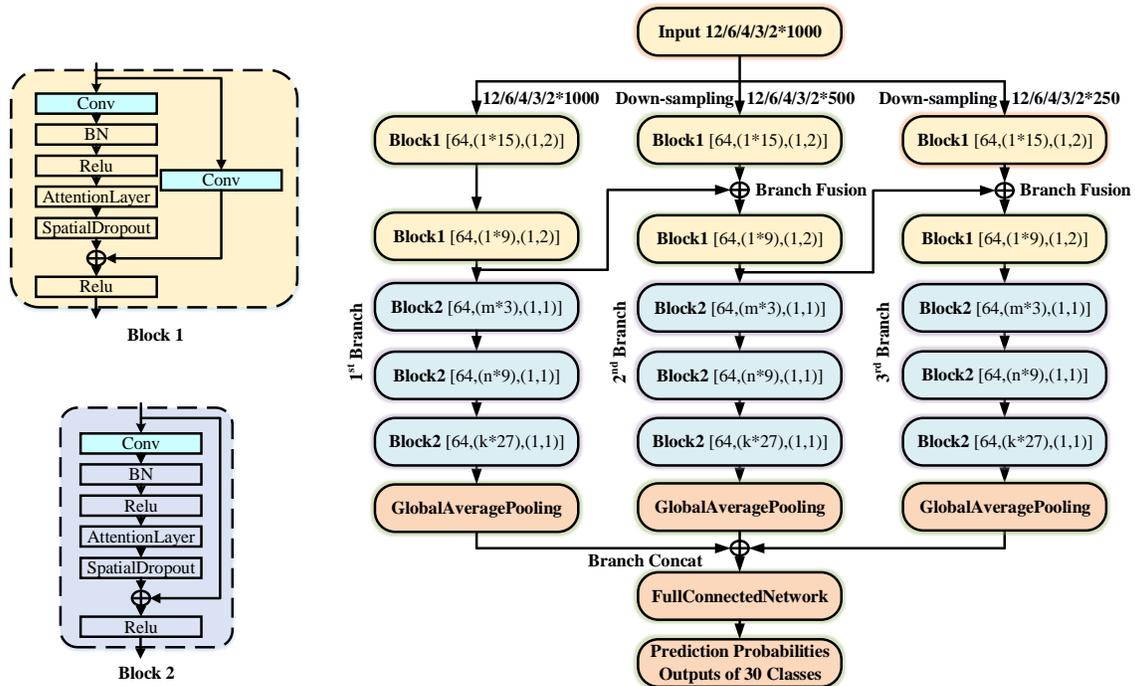


Figure 1. The proposed model architecture. The meaning of the parameters in brackets is [filter number, filter size, stride size]. The values of (m, n, k) are (5, 5, 5), (3, 3, 3), (2, 2, 2), (1, 2, 2), (1, 1, 2) for 12-lead, 6-lead, 4-lead, 3-lead, 2-lead models, respectively.

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.

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

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

The public training set was randomly divided into training and testing at a ratio of 4:1, and the training dataset 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. Hyperparameters of the network (convolution kernel size, number of blocks, etc.) were adjusted according to the performance on the validation dataset to achieve optimal performance. The method of reducing the learning rate with a ratio of 0.5 during training was adopted to alleviate over-fitting.

3. Results

We evaluated our proposed algorithms through 5-fold cross-validation on the public training set with the Challenge evaluation metric. The Challenge scores on both the public training set, hidden validation set, and hidden test set that our final selected entry (team AIRCAS_MEL1) obtained were shown in Table 1.

Leads	Training	Validation	Test	Ranking
12	0.721±0.002	0.63	0.38	21
6	0.693±0.004	0.57	0.33	23
4	0.702±0.002	0.58	0.37	20
3	0.699±0.004	0.57	0.43	16
2	0.683±0.004	0.56	0.38	20

Table 1. Challenge scores for our final selected entry (team AIRCAS_MEL1) using 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set as well as the ranking on the hidden test set.

4. Discussion and Conclusions

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 5-fold cross-validated scores with the Challenge evaluation metric on the public training set were shown in Table 2.

Leads	Fixed sizes (3, 3, 3)	Variable sizes (3, 6, 12)	Variable sizes (3, 9, 27)
12	0.666	0.690	0.700
6	0.634	0.653	0.663
4	0.640	0.660	0.672
3	0.629	0.656	0.672
2	0.614	0.641	0.653

Table 2. Comparison of 5-fold cross-validation results of the effect of convolution kernel configuration on model performance.

The results showed that the performances of models using convolution kernels with variable sizes were better than those with fixed sizes. The possible reason is that the convolution kernels with variable sizes can capture pathological information at different scales. More importantly, the models taking convolution kernels with variable sizes (3, 9, 27) have 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 performances of three branch models. The baseline models refer to that model taking convolution kernels with variable sizes (3, 9, 27) in the “sampling” dimension. The 5-fold cross-validation scores with the Challenge evaluation metric of individual branch models on the public training set were shown in Table 3.

Leads	1 st Branch Models	2 nd Branch Models	3 rd Branch Models	Baseline Models
12	0.657	0.664	0.643	0.700
6	0.635	0.637	0.617	0.663
4	0.633	0.636	0.616	0.672
3	0.632	0.631	0.615	0.672
2	0.615	0.621	0.598	0.653

Table 3. Comparison of 5-fold cross-validation results of individual branch models using ablation study.

The results showed that the performances of the second branch networks were best among all branch networks. It can be explained that the second branch networks can effectively capture more pathological features, or in other words, the majority of pathological features of arrhythmias concentrated on this scale range.

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 5-fold cross-validation scores with the Challenge evaluation metric on publicly available datasets were shown in Table 4.

Leads	Baseline Models	Baseline + Attention Models (proposed)
12	0.700	0.721
6	0.663	0.693
4	0.672	0.702
3	0.672	0.699
2	0.653	0.683

Table 4. Comparison of 5-fold cross-validation results of the effect of attention mechanism on model performance using ablation study.

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.

In this paper, we proposed a novel multi-scale convolutional neural network for 30 arrhythmias classification on reduced-lead ECGs. 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 sizes (3, 9, 27) have the best performance among all models. The correspondence between the pathological information of different scales and convolutional kernel sizes (scales) will be studied in future work. The performance of each branch model was also individually evaluated on the public training set, and the results showed that the second branch network has the best performance among all branch networks. In addition, 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.

Finally, our proposed models were evaluated on the public training set, and we achieved 5-fold cross-validation scores of 0.721, 0.693, 0.702, 0.699, and 0.683 for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions of the public training set with the Challenge evaluation metric, and these models received scores of 0.38, 0.33, 0.37, 0.43, and 0.38 (ranked 21th, 23th, 20th, 16th, and 20th out of 39 teams) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions of the hidden test set with the Challenge evaluation metric. The proposed models showed poor performance on the hidden test set. Some necessary strategies will be adopted to improve the generalization performances of the models in future work.

Acknowledgments

This work is supported by the National Key Research and Development Project 2020YFC1512304, 2020YFC2003703, 2018YFC2001101, 2018YFC2001802, National Natural Science Foundation of China (Grant 62071451), and CAMS Innovation Fund for Medical Sciences (2019-I2M-5-019).

References

- [1] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiological Measurement* 2020;41.
- [2] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. *Computing in Cardiology* 2021;48:1–4.
- [3] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. *Journal of Medical Imaging and Health Informatics* 2018;8(7):1368–1373.
- [4] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead Arrhythmia Database. *PhysioBank PhysioToolkit and PhysioNet* 2008; Doi:10.13026/C2V88N.
- [5] Boussejot R, Kreisler D, Schnabel A. Nutzung der EKG Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik* 1995;40(S1):317–318.
- [6] Wagner P, Strodthoff N, Boussejot RD, Kreisler D, LunzeFI, Samek W, et al. PTB-XL, a Large Publicly Available Electrocardiography Dataset. *Scientific Data* 2020;7(1):1–15.
- [7] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients. *Scientific Data* 2020;7(48):1–8.
- [8] Zheng J, Cui H, Struppa D, Zhang J, Yacoub SM, et al. Optimal Multi-Stage Arrhythmia Classification Approach. *Scientific Data* 2020;10(2898):1–17.
- [9] He Z, Zhang P, Xu L, Bai Z, Zhang H, et al. A Novel Convolutional Neural Network for Arrhythmia Detection From 12-lead Electrocardiograms. *Computing in Cardiology* 2020;1-4.
- [10] Hu J, Shen L, Albanie S, Sun G, and Wu E. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2020;42(8):2011–2023.

Address for correspondence:

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