A Novel Convolutional Neural Network for Arrhythmia Detection From 12-lead Electrocardiograms

Zhengling He[1], Pengfei Zhang[1], Lirui Xu[1], Zhongrui Bai[1], Hao Zhang[1], Weisong Li[1], Pan Xia[1], Xianxiang Chen[1,2]

[1]State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

[2]Research Unit of Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences

Abstract

Electrocardiogram (ECG) is a widely medical tool used in the clinical diagnosis of arrhythmia, numerous algorithms based on deep learning have been proposed to achieve automatic arrhythmia detection. In PhysioNet/Computing in Cardiology Challenge 2020, inspired by the deep residual learning and attention mechanism, we proposed a novel neural network to accomplish this classification task. The backbone of the network is a carefully designed 2-D convolutional neural network (CNN) with residual connection and attention mechanism, and it can adapt to multi-lead ECG signals as input. The first 10 seconds of records from all leads are extracted and preprocessed as input for end-to-end training, and the prediction probabilities of 27 categories are output. The proposed algorithm was firstly verified and adjusted via 5-fold cross-validation on officially published datasets from 4 multiple sources. Finally, our team (MetaHeart) achieved a challenge validation score of 0.616 and full test score of 0.370, but were not ranked due to omissions in the submission.

1. Introduction

Electrocardiogram (ECG) is a non-invasive medical tool to record the rhythm of the heart, the diagnosis of abnormal ECG rhythms requires a professionally trained cardiologist to carefully study the records, which is extremely time-consuming. Automatic diagnosis by computer-aided system can help improve doctors' work efficiency, and these algorithms can even be transplanted to wearable devices to enable real-time arrhythmia detection in the home environment.

A large number of algorithms based on machine learning and deep learning have been established for this purpose in the past few decades. General machine learning approaches for arrhythmia classification task includes signal pre-processing, feature extraction, feature selection and final classification. The key to the machine learning algorithm is to obtain numerous carefully designed features based on expert knowledge, such as statistical features [1, 2], RR-interval and morphological features [3-5], wavelet-based features [6, 7]. However, performances of these models largely depend on the hand-crafted features. In deep learning, feature engineering has almost being replaced due to the automatic feature extraction capabilities of neural networks, that is to say, an end-to-end deep learning framework allows a machine to automatically discover the features that are necessary and suited to the classification task. 1-D convolutional neural network (CNN) applied to ECG signals [8, 9], 2-D convolutional neural network applied to time-frequency spectrogram [10, 11], combination of CNN and Long Short-Term Memory (LSTM) network [12, 13], and numerous other end-to-end networks have been proposed and verified to be effective in the field of arrhythmia detection.

The goal of PhysioNet/Computing in Cardiology Challenge 2020 is to identify 27 types of arrhythmia from 12-lead ECG recordings, and performance of all submitted models will be evaluated by a defined evaluation score called challenge metric. We proposed a novel 2-D convolutional neural network with residual connection and attention mechanism to achieve the target in this study. The first 10 seconds of records from all leads are extracted as input for end-to-end training, and the prediction probabilities of 27 categories are the output of the model. Performance of the proposed model was cross-validated on officially published datasets from 4 multiple sources and ultimately evaluated on hidden test set.

2. Methods

2.1. Datasets and preprocessing

The publicly available datasets consist of a total of 43,101 subjects from 4 different sources. 12-lead ECG signals and corresponding labels of all individuals in dataset were recorded in separate files, only 27 categories will be considered in the final evaluation score [14]. The sampling rate of the data set varies from 257 to 1000 Hz, the original ECG signals are 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 20 Hz is designed to eliminate baseline drift and high frequency noise. The first 10 seconds of records are maintained due to the median length of all records is 10 seconds, and then data will be truncated or expanded with 0 to a consistent length. Z-score normalization is applied to normalize signals in all leads.

2.2. Model architecture

Three specially designed blocks are stacked to form the backbone of the proposed model, as shown in Figure 1. 2-D convolutional network is mainly applied to the field of image recognition, so Huang et al. [10] and Salem et al. [11] transformed ECG signals to time-frequency images at first in their frameworks. Different from their approaches, however, multi-lead ECG signals can also be regarded as 2-D input with two dimensions, we call them "lead" and "sample", and we can enable the network perceive information from the single lead as well as cross-lead by controlling the shape of the 2-D convolutional filter size.

●**Filter:** A large convolution filter (or kernel) size in the dimension of "sample" is adopted to the first block to enhance the perception field, and decay from the second block, and the size will become 3 from the 9th block on. For the first 6 blocks, the filter size in the direction of "lead" is set to 1, which can ensure that the information across lead is not exchanged at lower layers but just share the same filter, and the filter size will be adjust to 3 at higher blocks in the direction of "lead", which makes each individual convolution filter not only perceives the information of the current lead, but also perceives cross-lead information, thus making cross-lead information fusion possible.

●**Stride**: It controls the step of two adjacent convolution operations, the first several blocks are allocated the relatively lager stride as 2 to significantly reduce the dimension of feature map passed to the next block, and blocks use a stride of 1 from the 6th block on.

●**Batch normalization and spatial dropout**: Batch normalization (BN) [15] has been demonstrated to speed up the convergence of network and alleviate the risk of over-fitting. Spatial dropout [16] is a widely used strategy to prevent over-fitting, which drop the feature maps randomly with a preset rate, and is more suitable for convolutional layers than standard dropout strategy [17].

●**Attention module**: We added an attention module called Squeeze-and-Excitation (SE) [18] to blocks. It 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, and this weight matrix reflects the importance of individual feature map. These weights will be finally 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 use *sigmoid* due to the fact that ECG signal in the real-world can contain more than one type of arrhythmia, so it's actually a multi-label classification problem. *sigmoid* can independently map the output to a probability value in the range of 0 to 1.

●**Global average pooling layer**: A global average pooling

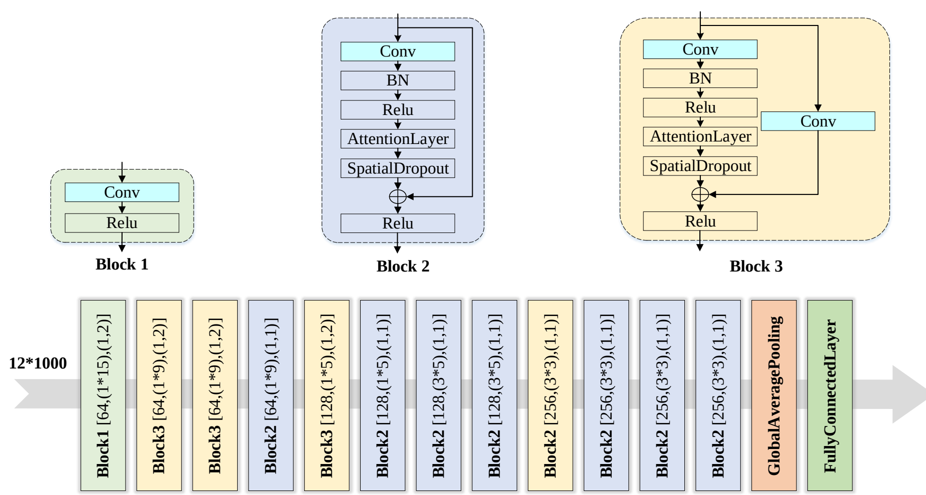

Figure 1. The proposed model architecture. The meaning of the parameters in brackets is [filter number, filter size, stride size].

(GAP) layer [19] is stacked between the last convolutional layer and the output layer.

•**Loss function**: The 27 categories considered in evaluation score will be encoded in binary form as $label = [l_1, l_2, ..., l_{27}]$, and for each record, if their category exists in 27 categories, set the corresponding binary bit to 1. A weighted binary cross-entropy loss is designed as the optimization target, its definition is 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 category i, separately, 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 ε is set to 0.01 to prevent division by 0.

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

2.3. Model training

Data are randomly divided in a ratio of 8:2 for training and testing, and the training set is further divided in a ratio of 9:1 as training and validation dataset. *Adam* with an initial learning rate of 0.001 is applied for optimization. Hyper parameters of the network (dropout rate, number of blocks, etc.) are adjusted empirically according to the performance on validation dataset to achieve optimal performance. Early-stopping and reduce learning rate with a ratio of 0.5 during training are also adopted to alleviate over-fitting.

3. Results and discussion

We verified and adjusted our algorithm on officially published datasets from 4 multiple sources (including 43,101 subjects) by 5-fold cross-validation, the average evaluation score (challenge metric) on publicly available dataset was 0.621, and the score on hidden test set our team (MetaHeart) obtained is shown in Table 1.

Table 1. Challenge metric on hidden test set.

Dataset	Challenge metric
Validation Set	0.616
Hidden CPSC Set	0.758
Hidden G12EC Set	0.590
Hidden Undisclosed Set	0.194
Full Test Set	0.370

An ablation study was adapted to analyze the effect of 3 strategies on model performance elevation: (a) Remove the attention layer; (b) Replace the spatial dropout by standard dropout; (c) Replace GAP by fully connected (FC) layer. The cross-validated averaged evaluation score on publicly available datasets is shown in Table 2.

Dropout is a strategy we used to alleviate over-fitting and improve generalization performance, the performance of both standard dropout and spatial dropout have been compared in this study, the results show that the performance of using standard dropout is worse than using spatial dropout. It is mainly because the model is easy to over-fit to training set according to our experiment. In fact, dropout works by preventing activations from becoming strongly correlated in network, it assumes independence between the activations. However, the activations in a same feature map are also strongly correlated in CNN, so standard dropout may not work effectively [16]. Spatial dropout between convolutional layers drops the entire feature maps rather than individual activations to alleviate this problem, so it achieved better performance.

The GAP layer between the last convolutional layer and the output layer also helps to improve the performance of the model. The FC layer will increase the amount of model parameters, and the loss in validation set in our experiment indicates that the model is easier to over-fit. In fact, the training data set is relatively smaller compared to other fields such as images recognition, and our network is consist of a larger number of filters and deep layers, which makes the capacity of the network pretty large. The use of the GAP layer after the last convolutional layer is more native to the convolution structure by enforcing correspondences between feature maps and categories [19], and it greatly reduces the amount of model parameters and reduces the risk of over-fitting.

It can be also observed that using an SE attention layer provides improvement in performance. Generally explanation is that attention layer can help the network to know where to emphasize or suppress by automatically learning the information flow from the above layer.

Table 2. An ablation study performed on officially published datasets to analyze the effect of 3 strategies on performance elevation.

Strategy	Challenge metric (5-fold)
Spatial dropout+GAP	0.569
Standard dropout+Attention+GAP	0.607
Spatial dropout+Attention+FC	0.571
Spatial dropout+Attention+GAP (proposed)	**0.621**

Limitations of this study includes that we only used ECG signals with a length of 10 seconds, which may ignore information especially some intermittent abnormal rhythms, a network that can accommodate variable length input will be designed in the future. In addition, we only used a single model to predict the output, which can be replaced by an ensemble model in the future to further improve performance.

4. Conclusion

Inspired by the deep residual learning and attention mechanism, we proposed a novel 2-D convolutional neural network to identify 27 types of arrhythmia from 12-lead ECG recordings in an end-to-end training manner. An ablation study was performed to analyze the effect of 3 strategies on performance elevation, it was found that spatial dropout instead of standard dropout, GAP instead of FC layer after the convolutional layer and the introduction of the attention mechanism had a positive effect on model performance improvement. The average challenge metric of 5-fold cross-validation on publicly available datasets was 0.621, and we ultimately achieved a challenge validation score of 0.616 and full test score of 0.370. In future work, we will design a network that can accommodate variable length input and use an ensemble framework to fuse multiple models to further enhance the model performance.

Acknowledgments

This work is supported by the National Key Research and Development Project 2020YFC2003700, 2016YFC1304302, 2018YFC2001100, 2018YFC2001800 and CAMS Innovation Fund for Medical Sciences (CIFMS No. 2019-I2M-5-019).

References

[1] Martis RJ, Acharya UR, Lim CM *et al.*, "Application of higher order cumulant features for cardiac health diagnosis using ECG signals," *International Journal of Neural Systems,* vol. 23, no. 04, p. 1350014, 2013.

[2] Martis RJ, Acharya UR, Prasad H *et al.*, "Application of higher order statistics for atrial arrhythmia classification," *Biomedical Signal Processing and Control,* vol. 8, no. 6, pp. 888-900, 2013.

[3] Afkhami RG, Azarnia G, and Tinati MA, "Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals," *Pattern Recognition Letters,* vol. 70, pp. 45-51, 2016.

[4] Xiang Y, Luo J, Zhu T *et al.*, "ECG-based heartbeat classification using two-level convolutional neural network and RR interval difference," *IEICE Transactions on Information and Systems,* vol. 101, no. 4, pp. 1189-1198, 2018.

[5] Llamedo M and Martínez JP, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Transactions on Biomedical Engineering,* vol. 58, no. 3, pp. 616-625, 2010.

[6] Sarkaleh MK and Shahbahrami A, "Classification of ECG arrhythmias using discrete wavelet transform and neural networks," *International Journal of Computer Science, Engineering and Applications,* vol. 2, no. 1, p. 1, 2012.

[7] Arif M, "Robust electrocardiogram (ECG) beat classification using discrete wavelet transform," *Physiological Measurement,* vol. 29, no. 5, p. 555, 2008.

[8] Yıldırım Ö, Pławiak P, Tan R-S *et al.*, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Computers in Biology and Medicine,* vol. 102, pp. 411-420, 2018.

[9] Kiranyaz S, Ince T, and Gabbouj M, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering,* vol. 63, no. 3, pp. 664-675, 2015.

[10] Huang J, Chen B, Yao B *et al.*, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access,* vol. 7, pp. 92871-92880, 2019.

[11] Salem M, Taheri S, and Yuan JS, "ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2018, pp. 1-4: IEEE.

[12] 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,* vol. 94, pp. 19-26, 2018.

[13] Oh SL, Ng EY, San Tan R *et al.*, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Computers in Biology and Medicine*, vol. 102, pp. 278-287, 2018.

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

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

[16] 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.

[17] 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.

[18] 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.

[19] Lin M, Chen Q, and Yan S, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

Address for correspondence:

Xianxiang Chen
State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, No. 9, North Zhongguancun, Haidian District, Beijing.
chenxx@aircas.ac.cn.