A Real-time ECG Classification Scheme Using Anti-aliased Blocks With Low Sampling Rate

Yunkai Yu[1†], Zhihong Yang[2†], Peiyao Li[3], Zhicheng Yang[4], Yuyang You[1]

[1] Beijing Institute of Technology, Beijing, China

[2] Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing, China

[3] Tsinghua University, Beijing, China

[4] PAII Inc, Palo Alto, USA

[†] First co-authors, [*] Corresponding author

Abstract

Feasible real-time ECG classification algorithms contribute to an early and correct diagnosis of cardiac abnormalities. In this paper, we (team Triology) leverage 80 Hz ECG signals to develop a lightweight end-to-end neural network. A soft voting scheme is applied to determining the prediction in a long record from multiple segments. The model has a ResNet-18 backbone. It integrates standard and dilated convolutions to extract multi-scale information. Anti-aliased blocks are used for shift invariance. Age and sex are included. To encourage the inter-class competition in the multi-label classification task, lovász softmax and weighted cross entropy loss are randomly selected in the training process, which facilitates model convergence. In order to derive a robust model, data augmentation approaches like Gaussian noise, random erasing and shifting are implemented. Our offline validation is carried out using databases from four sources. We score 0.328 using the challenge metric. False negatives are main errors.

1.   Introduction

Cardiovascular abnormalities are common health problems. They are potential risk factors for cardiovascular diseases such as stroke [1]. High quality diagnosis at an early stage can bring increased successful treatments and life quality to potential patients. Moreover, a good automatic detection and classification system of cardiac abnormalities emancipates the clinicians from the onerous manual interpretation workload. The standard 12-lead Electrocardiogram provides sufficient physiological information, and it is an widely used as an invasive way to monitor heart activity.

Deep neural networks (DNNs) generalize well to large-scale database. An end-to-end scheme enables DNNs to learn effective representations automatically from raw data. Correspondingly, classical classifiers like support vector machine and random forest use hand-crafted features, which are designed according to expert knowledge. However, hardware resource limits the applicability of DNNs. In this manuscript, we explored a lightweight DNN method based on ResNet18 model for 1-D input, integrating anti-aliased design and multi-scale design into model structure. Two relevant factors, age and sex are considered when making predictions. The PhysioNet/Computing in Cardiology Challenge 2020 [2] provides large scale databases from multi-centers, enabling us to validate the credibility of our algorithm. Normal class accounts for a large proportion. In regard of the long-tailed distribution of categories and the label distribution, we utilize undersampling method. Weighted cross entropy loss and lovász softmax loss are combined to optimize the model. Furthermore, data augmentation approach is applied to improve model robustness. Our model needs to reduce false positives in the offline validation.

2.   Related work

This section illustrates DNNs for classification tasks from ECG signals. Previous work shows the feasibility of deep neural networks in heart rhythm classification. ResNet appears to be a good baseline. In Physionet 2017 challenge [3], Xiong et al design a ResNet block, achieving good performance in AF classification [4]. Hannun et al reported a DNN to capture morphological characteristics of ECG. It shows good performance in a wide range of heart rhythm categories [5]. Chen et al stack CNN blocks to compute useful representations. The subsequent bio-directional RNN layer and an attention layer are used to learn temporal information. They suggest that using single-lead information can have good performance from 12-leads in predicting multiple cardiac arrhythmia (CA) [6]. Our previous work illustrates that the sampling rate of ECG data can be as low as 60 Hz when records comprise of atrial fibrillation and normal rhythm. It validates

the feasibility of light-weight neural networks [7]. However, this technique should be tested in a larger database with more CA.
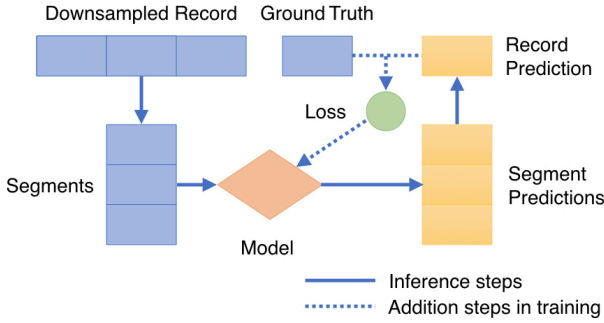
3. Methods

Figure 1: System overview.

Fig. 1 gives an overview of our system. Firstly, we split downsampled records into small segments. Therefore, the system can process records with different duration. Secondly, our end-to-end deep neural network outputs predictions of each segments. We take averaged probability of each segments as record prediction. Finally, in the training process, we use record prediction, together with ground truth to compute loss.

Following sections introduce our notations and the challenge metric. Afterwards, we illustrate each procedure of the system.

3.1. Notations

Throughout the paper we denote a 12-lead ECG record data as $X \in R^{12 \times N}$, where N is the number of samples. The corresponding label is defined as an one-hot vector z. The record is divided into small segments $x_i \in R^{12 \times n}, i \in [1, floor(\frac{N}{n})]$ after removing $mod_n(N)$ samples, where n is the number of samples of a segment, and $floor(\cdot)$ takes the nearest integer that is smaller than the input. Each segment is processed by the DNN $T(\cdot)$, where the output $y \in R^{1 \times m}$ satisfies $y = T(x_i)$, where m is the size of the heart rhythm set $C = \{c_1, c_2, \cdots, c_m\}$. Both $softmax$ and $sigmoid$ methods are used to transform y into a probability estimation for each class. The elements of $p_{softmax}$ and $p_{sigmoid}$ are computed as shown in Eq (1) and (2).

$$softmax: p^{c_i}_{softmax} = \frac{e^{y_i}}{\sum_{k \in C} e^{y_k}}, \quad (1)$$

$$sigmoid: p^{c_i}_{sigmoid} = \frac{1}{1 + e^{-y_i}}. \quad (2)$$

We also denote the categories that are scored with the challenge metric using sorted class index as shown in Table 1.

Table 1: Amount of each heart rhythm category.

SNOMED CT CODE	Sorted Class Index	Amount
10370003	0	299
111975006	1	277
164889003	2	3473
164890007	3	314
164909002	4	1041
164917005	5	1013
164934002	6	4673
164947007	7	340
251146004	8	556
270492004	9	2394
284470004	10	2534
39732003	11	6086
426177001	12	2359
426627000	13	277
426783006	14	20486
427084000	15	2391
427172004	16	1283
427393009	17	1238
445118002	18	1806
47665007	19	427
59931005	20	1111
698252002	21	996
713426002	22	1611
713427006	23	3083

3.2. Challenge metric

The evaluation metric generalizes the traditional accuracy by assigning different weights to the multi-class confusion matrix A, where the weight matrix W reflects the outcome or treatment that is on the basis of true diagnosis as well as partial misdiagnosis. It demonstrates that confusing some classes is much less harmful. The score s is given by adding up the elements of the weighted multi-class confusion matrix $W \circ A$, where $\circ$ is the Hadamard product. Then the score is mapped to $[0, 1]$ using the min-max scaling scheme, where the maximum value is the score of a perfect classifier with 100% accuracy and the minimum value is that of an inactive classifier whose output is always normal class [2]. According, true negatives are neither rewarded nor penalized.

3.3. Data and preprocessing

Training data for the challenge are from multiple sources. The first source is the training and unused data from the China Physiological Signal Challenge in 2018 (CPSC2018). The second source is from the publicly available St Petersburg INCART 12-lead Arrhythmia Database.

The third source set from the Physikalisch Technische Bundesanstalt (PTB) consists of two public databases: the PTB Diagnostic ECG Database which contains 549 records and the PTB-XL, a large dataset containing 21,837 ECG records. The fourth source is a Georgia database provided by Emory University, representing a unique demographic of the Southeastern United States. We pool the four databases together. The basic information of the pooled database is summarized in Table 2. Using

Table 2: Basic information of the pooled dataset

Female		22988 (0.534)
	1	22551
Label quantity	2	10138
	≥ 3	5954

downsampled signal is the core design in our light-weight scheme. Our previous work demonstrated that sampling frequency for distinguishing atrial fibrillation and normal rhythms can be as low as 60 Hz. Considered to the fact that the heart rhythm categories are largely expanded, we adopt 80 Hz signal to preserve sufficient information. Mean values are subtracted from each ECG lead.

3.4. Model structure

Our model structure is presented in Fig 2. The model backbone is modified from ResNet18 for 1-D signals, where 12 leads are regarded as different channels. We integrated two useful model architecture designs namely multi-scale design and anti-aliased design to embed important priors. Multi-scale design means using kernels of different sizes to capture information in multi-scale receptive fields. To further reduce the computation load, we implemented dilation convolution blocks where kernel size and stride is 3 and 2, respectively. It's an alternative of 1-D convolution blocks where kernel size is 5. The anti-aliased design accounts for the prior that models should possess shift invariance. It improves downscaling blocks by blur filters [8]. Batch normalization are replaced by instance normalization (IN) [9]. Age and sex are introduced into the feature set due that many ECG parameters are age- and sex-dependent [10, 11].

3.5. Loss

The loss function is defined as shown in Eq (3).

$$Loss(X, z) = \begin{cases} Loss_1\left(\frac{1}{n}\sum_{i=1}^{n} p_{softmax}, z\right), v \leq 0.5 \\ Loss_2\left(\frac{1}{n}\sum_{i=1}^{n} p_{sigmoid}, z\right), v > 0.5 \end{cases},$$
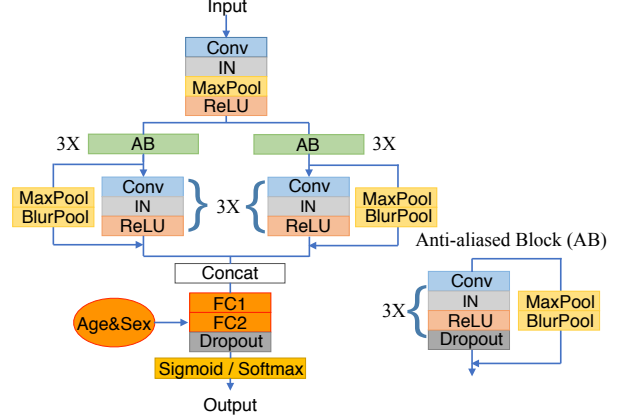
$$(3)$$

Figure 2: Model structure. The model backbone is a ResNet18 for 1-D input. The convolutional layers in two branches are of same kernel size (3) but different dilation (1 and 2). IN means instance normalization. The anti-aliased design here means using a max-pooling layer ($stride = 1$) and a blur-pooling layer ($stride = 2$) instead of a max-pooling layer ($stride = 2$).

where $v \sim U(0, 1)$ is a random variable, $Loss_1$ and $Loss_2$ are weighted cross entropy loss and lovász-softmax loss [12], respectively. Random shifts between the two loss function are designed to encourage the inter-class competition. Table 2 shows that many records have very few labels. The overwhelming size of output space is a key challenge of developing models for multi-label classification tasks [13], which is the motivation to design the loss.

To obtain final prediction needs category-wise threshold $\{th_i, i \in C\}$. The estimated probability possesses following property:

$$sgn(p_{softmax}^{c_i} - th_i) = sgn(p_{sigmoid}^{c_i} - th_i). \quad (4)$$

Therefore, we obtain

$$(p_{softmax}^{c_i} - th_i)(p_{sigmoid}^{c_i} - th_i) \geq 0.$$

Denote $k_i = \frac{p_{sigmoid}^{c_i}}{p_{softmax}^{c_i}}$, we have

$$(p_{sigmoid}^{c_i} - th_i)(p_{sigmoid}^{c_i} - k_i th_i) \geq 0. \quad (5)$$

Therefore, using two probability estimation increases the chance to penalize low-confidence predictions.

3.6. Training process

Considered to the fact that different categories may arise at different period, only record prediction and ground truth were used to compute loss. Using thresholds from the training set, we transformed the predictions from probabilities to their binary form. The model was trained using an Adam optimizer, with a 0.0003 learning rate. We accumulated the gradients and make back propagation every

64 records. During the training process, the dropout probability was set to 0.3. Last but not the least, we used data augmentation methods including adding Gaussian noise, random erasing and shifting.

4. Results and discussion

To evaluate the performance of our algorithm, the model for diagnosing the hidden test set is determined by the training and validation set. In total 80% of the data are selected as the training set, and the rest are used as validation set. Both of them include all 24 categories. We also randomly excluded records with signal absence and noisy records based on standard deviation threshold. Undersampling strategy was adopted to remove 30% normal records from training set. We measured the challenge metrics in the validation set every 3 epochs and stored the best parameters.

Currently, false positives are major errors in our offline validation, where the model scores 0.328. From our view, it indicates that a more competitive inter-class competition is needed. To find the best hyper-parameters, we compared using 60, 80 and 100 Hz signal, and ablation experiments are necessary.

5. Conclusion

This paper validates a light-weight end-to-end DNN for downsampled 12-lead ECG diagnosis. The model integrates multi-scale and anti-aliased design and basic characteristics like age and sex. It achieves feasible results when the amount of heart rhythm classes is small. Switching between softmax and sigmoid activation is a strategy to get multi-class output and encourage inter-class competition. Future work is to perform ablation experiments and improve model performance when the amount of heart rhythm classes is large.

Acknowledgements

References

[1] Go A, et al. Association of burden of atrial fibrillation with risk of ischemic stroke in adults with paroxysmal atrial fibrillation: The kp-rhythm study. JAMA Cardiology 05 2018; 3.

[2] Alday P, et al. Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. Physiol Meas 2020 (Under review);.

[3] Clifford GD, et al. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In 2017 Computing in Cardiology (CinC). 2017; 1–4.

[4] Xiong Z, Stiles MK, Zhao J. Robust ecg signal classification for detection of atrial fibrillation using a novel neural network. In 2017 Computing in Cardiology (CinC). Sep. 2017; 1–4.

[5] Hannun AY, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine 2019;25(1):65–69.

[6] Chen TM, et al. Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model, 09 2019.

[7] Yu Y, et al. Work-in-progress: On the feasibility of lightweight scheme of real-time atrial fibrillation detection using deep learning. In 2019 IEEE Real-Time Systems Symposium (RTSS). 2019; 552–555.

[8] Zhang R. Making convolutional networks shift-invariant again. In ICML. 2019; .

[9] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization. ArXiv 2016;abs/1607.08022.

[10] Rijnbeek PR, et al. Normal values of the electrocardiogram for ages 16–90years. Journal of Electrocardiology 2014; 47(6):914 – 921. ISSN 0022-0736.

[11] Laureanti R, et al. Sex-related electrocardiographic differences in patients with different types of atrial fibrillation: Results from the swiss-af study. International Journal of Cardiology 2020;307:63 – 70. ISSN 0167-5273.

[12] Berman, et al. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018; 4413–4421.

[13] Zhang M, Zhou Z. A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering 2014;26(8):1819–1837.

Address for correspondence:

Yuang You
Beijing Institute of Technology, Haidian District, Beijing, China
arthurwy@163.com