Classification of Cardiac Abnormalities From ECG Signals Using SE-ResNet

Zhaowei Zhu[1*], Han Wang[2*], Tingting Zhao[1], Yangming Guo[1], Zhuoyang Xu[1], Zhuo Liu[1], Siqi Liu[3], Xiang Lan[2], Xingzhi Sun[1#], Mengling Feng[2#]

[1]Ping An Technology, Beijing, China
[2]National University of Singapore, National University Health System, Singapore
[3]NUS Graduate School for Integrative Sciences & Engineering, Singapore

Abstract

In PhysioNet/Computing in Cardiology Challenge 2020, we developed an ensembled model based on SE-ResNet to classify cardiac abnormalities from 12-lead electrocardiogram (ECG) signals. We employed two residual neural network modules with squeeze-and-excitation blocks to learn from the first 10-second and 30-second segments of the signals. We used external open-source data for validation and fine-tuning during the model development phase. We designed a multi-label loss to emphasize the impact of wrong predictions during training. We built a rule-based bradycardia model based on clinical knowledge to correct the output. All these efforts helped us to achieve a robust classification performance. Our final model achieved a challenge validation score of 0.682 *and a full test score of* 0.514, *placing our team HeartBeats 3rd out of 41 in the official ranking. We believed that our model has a great potential to be applied in the actual clinical practice, and planned to further extend the research after the challenge.*

1. Introduction

Electrocardiogram (ECG) examines the physiological activities of the heart and helps the diagnosis of many cardiovascular abnormalities [1]. Diagnosis with the ECG signal requires the clinician to careful examine and identify the inter-beat and intra-beat patterns. The process can be both time-consuming and error-prone [2]. Thus, a quick and accurate algorithm for automatic ECG pattern classification is always desired. In the PhysioNet/Computing in Cardiology Challenge 2020, participants were to develop a model to automatically identify the cardiac abnormalities present in 12-lead ECG recordings [3]. In this paper, we will describe our approaches to tackle the challenge.

---

* These authors contributed equally and are co-first authors.
\# These authors contributed equally and are co-corresponding authors.

2. Methods

The overall system design is shown in Figure 1 and will be elaborated below.

2.1. Datasets & Labelling

The public challenge data consist of 43,101 12-lead ECG signals from 6 different datasets, namely CPSC, CPSC2, PTB, PTB-XL, Georgia and INCART. The sampling frequency of the signals varies from 257 Hz to 1000 Hz, and the length of the signals varies from 6 seconds to 30 minutes. There are 111 labeled abnormalities in total, of which 27 are included in the final scoring metrics. From these data, we created our offline training set and test set via the following approaches.

Processing original data. INCART dataset was excluded from our training data since it has only 74 30-minutes records with a sampling frequency of 257 Hz and is significantly different from other datasets. All data without a label in the 27 scored classes were excluded as well. PTB dataset was downsampled from 1000 Hz to 500 Hz to make the sampling frequency of all training data unified. Since lead III, aVR, aVL and aVF are linearly dependent on other leads and can be calculated based on Einthoven's Law [4] and Goldberger's equations [5] , these 4 leads were also excluded. In the rest of the data, we randomly split 80% as the training set and 20% as the offline test set. The final sizes of the training set and test set are 30,172 and 7,544 respectively.

Relabelling CPSC data. CPSC dataset was relabeled due to the fact that the labels cover only 9 classes and the class distribution is significantly different from other datasets. A baseline model was first trained on the original training set, and used for inference on CPSC dataset. For each signal, among all the classes with inference output probability higher than 0.8, the classes that were not in the original 9 classes but in the 27 scored classes were added as a new label.

To check the validity of our relabelling strategy, out of all
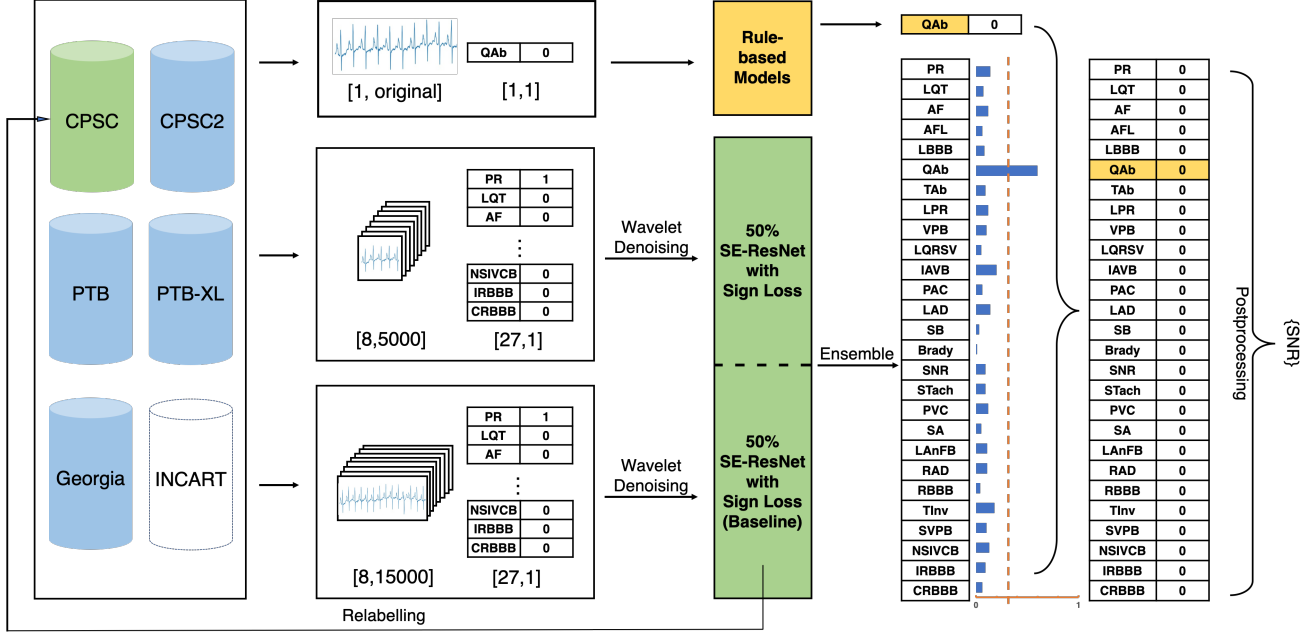
Figure 1. Illustration of the system design.

the relabelled data with inference output probability higher than 0.95, 11 records were reviewed by a clinician. The feedback that most of the new labels were valid testified that CPSC dataset has missing labels.

Including external dataset. In order to help validate the generalizability of our model, an external dataset from Hefei Hi-tech Cup ECG Intelligent Competition [6] (Hefei dataset in short) was introduced. Hefei dataset consists of 40,000 records of 8-lead ECG signals with a sampling frequency of 500 Hz and length of 10 seconds. Out of all the records, 6,500 records with labels in the 27 scored classes were randomly selected and formed an external test set.

2.2. Data Preprocessing

To better prepare the data for model training, we adopted the following data preprocessing techniques.

Truncating & padding. For the baseline model, all input signals were fixed at 30 seconds in length. This was done by truncating the part exceeding the first 30 seconds for longer signals and padding the shorter signals with zero. For the other ensembled model, the input length was fixed at 10 seconds with the same preprocessing method.

Wavelet denoising. To reduce the noise in ECG signals, biorthogonal wavelet transformation was applied. The numbers of vanishing moments for the decomposition and reconstruction filters were 2 and 6 respectively. The level of refinement was set to be 8.

2.3. System Architecture

SE-ResNet. SE-ResNet [7] was used as our main model for the task. Introducing Squeeze-and-Excitation (SE) block into the ResNet [8] structure allows the model to adaptively adjust the weights of each channel in the feature maps. In this way, more important channels could be given higher weights, leading to a better prediction performance. Our baseline model was an SE-ResNet model with the input length of 30 seconds. To minimize the effect of padding on the shorter signals, another SE-ResNet model was trained with the input length of 10 seconds and ensembled with the baseline model. The structure of our SE-ResNet model is shown in Figure 2.

Sign Loss. A significant issue observed in our data was class imbalance, which resulted in predictions biased towards the majority class. To mitigate this issue, we designed a multi-label sign loss for our model training. The loss is defined as follows:

$$\text{sign}(p) = \left\{ \begin{array}{ll} y - 2py + p^2 & , |y - p| < 0.5 \\ 1 & , |y - p| \geq 0.5 \end{array} \right.$$

$$\text{Loss}(p, y) = \text{sign}(p) \times BinaryCrossEntropyLoss(p, y)$$

For the correctly classified labels, a coefficient smaller than 1 was multiplied to the default binary cross entropy loss. By doing so, the accumulated loss from the large number of true negative labels became smaller, and the loss from the mis-classified labels became more prominent.
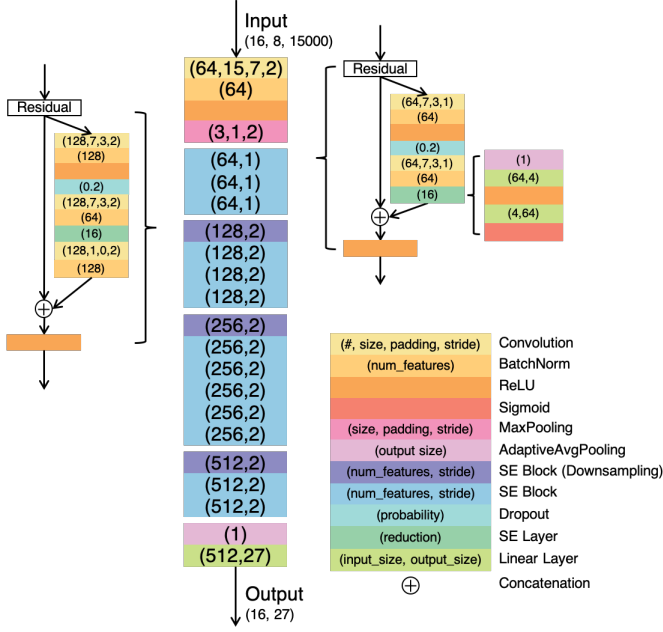
Figure 2. Architecture of the SE-ResNet model.

Rule-based model. The baseline model did not perform well enough for certain classes while there were relatively clear clinical rules to follow. One of such classes is bradycardia, which indicates the heart rate is slower than 1 beats per second, or the R-R interval between two heartbeats is longer than 1 second. To detect the R-R intervals, Pan & Tompkins algorithm [9] was used to detect the R-peaks on lead I, and R-R intervals could be easily calculated. The pseudocode of the rule-based model for bradycardia is shown in Algorithm 1.

Algorithm 1: Rule-based bradycardia classifier

Input: List of R-R intervals
Output: Classification of bradycardia
brady_beats = 0;
foreach *R-R interval*
    if *1s ≤ length of interval ≤ 1.6s*
        brady_beats += 1
if *brady_beats / # of R-R intervals ≥ 0.5*
    return True
else
    return False

However, the final bradycardia prediction of the system was not purely decided by the rule-based model. A very high recall and low precision were observed when doing so, and this could be attributed to the sub-optimal label quality of the datasets. Therefore, the prediction of rule-based model was only taken when its output is negative. The pseudocode for the final bradycardia prediction is shown in Algorithm 2.

Algorithm 2: Final bradycardia prediction

Input: Prediction from ensembled model
        Prediction from rule-based model
Output: Final classification of bradycardia
if *Prediction from rule-based model is False*
    return False
else
    return Prediction from ensembled model

SNR Postprocessing. Signals that were predicted to be negative for all classes were revised to be positive for the default normal class, sinus rhythm (SNR).

2.4. Training Setup

The ensembled model was trained for 19 epochs with a batch size of 16 on a machine with 117 GB RAM, 4-core CPU and one NVIDIA V100 GPU. The model parameters were optimized with the Adam optimizer [10]. The learning rate during training was set as 0.001, and rescheduled to 0.0001 at the 13th epoch. The optimal binarization threshold was found to be 0.36 on the offline test set.

3. Results

Table 1 shows the locally evaluated challenge scores on a) the test set divided from the official offline data, b) the test set from Hefei dataset and c) the full training set consisting of both offline training and test set.

	20% test	Hefei test	Full training
1) $Model_{15000}$	0.674	0.300	0.805
2) $Model_{5000}$	0.674	0.236	0.782
3) $Model_{1) + 2)}$	0.680	0.305	0.883
4) $Model_{3) + Rule}$	**0.683**	**0.319**	**0.885**

Table 1. Model performance on offline datasets. SNR postprocessing were done on all four models.

The 4th model scored highest in all three sets, hence was chosen as our best entry, and obtained challenge scores of 0.682 on the online validation data and 0.514 on the online test data.

4. Discussion

From the results above, our ensembled approach demonstrated its ability to classify the cardiac abnormalities despite of the challenges presented, e.g. noise in the signals and labels. The offline score on the 20% test set is 0.683, only differed the online score by 0.001, suggesting good generalizability and little overfitting.

During the challenge, we also experimented with several other ideas. One idea that we would like to share and hopefully inspire further explorations is the segmentation of abnormal heartbeats.

Abnormal heartbeats segmentation via 1D U-net. Some of the cardiac abnormalities are associated with individual heartbeats. If the model can learn to recognize these individual heartbeats that directly lead to the abnormality label, it may perform and generalize better, especially when the abnormal heartbeats are rare and sparse over the signal span.

Based on the preliminary analysis of the prediction results and advice from a clinician, we selected data with premature ventricular complex (PVC) labels to annotate. The raw lead-II signals were transformed into images by plotting on a grid background. The images were then imported into Colabeler, and we manually annotated all suspected PVC heartbeats by specifying their x-axis spans. Lastly, we translated the x-axis spans back to the actual locations on the signals. Due to time constraint, only 160 PVC data were annotated and used for subsequent training. We adapted U-net [11], a popular segmentation model in medical imaging to segment the PVC heartbeats. The 160 annotated PVC records were used as positive samples and 500 randomly selected records without PVC label were used as negative samples to train the model. The PVC classification is considered as positive if there is any positive PVC signal output. When we incorporated U-net into our system, the PVC predictions were solely determined by U-net.

Our experiments showed that incorporating U-net increased the F_{beta} and G_{beta} measures in the evaluation metrics, but no challenge score improvement was observed. Due to the additional training time required, we did not incorporate U-net in our final system.

5.   Conclusions

In this paper, we have described our approach used in the PhysioNet/Computing in Cardiology Challenge 2020. Our ensembled SE-ResNet model is able to classify 27 cardiac abnormalities on 12-lead ECG signals with a challenge validation score of 0.682 and a full test score of 0.514. Since the system is trained on real-life datasets, we believe that it has a great potential in the actual clinical practice.

Acknowledgments

References

[1] Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, Van Herpen G, Kors JA, Macfarlane P, Mirvis DM, et al. Recommendations for the standardization and interpretation of the electrocardiogram: part i: the electrocardiogram and its technology a scientific statement from the American Heart Association electrocardiography and arrhythmias committee, council on clinical cardiology; the American college of cardiology foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. Journal of the American College of Cardiology 2007;49(10):1109–1127.

[2] Bickerton M, Pooler A. Misplaced ecg electrodes and the need for continuing training. British Journal of Cardiac Nursing 2019;14(3):123–132.

[3] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad BA, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurements 2020;.

[4] Kligfield P. The centennial of the Einthoven electrocardiogram. Journal of Electrocardiology 2002;35(4):123–129.

[5] Goldberger AL, Goldberger ZD, Shvilkin A. Goldbergers clinical electrocardiography: a simplified approach. Elsevier, 2018.

[6] URL https://tianchi.aliyun.com/competition/entrance/231754/introduction.

[7] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEECVF Conference on Computer Vision and Pattern Recognition 2018;.

[8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition CVPR 2016;.

[9] Pan J, Tompkins W. A real-time QRS detection algorithm. IEEE Transaction on Biomedical Engineering ;32.

[10] Kingma DP, Ba J. Adam: A method for stochastic optimization. In Bengio Y, LeCun Y (eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. 2015; URL http://arxiv.org/abs/1412.6980.

[11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science Medical Image Computing and Computer Assisted Intervention – MICCAI 2015 2015; 234–241.

Address for correspondence:

Xingzhi Sun
3 Xinyuan Road, Chaoyang District, Beijing, China
sunxingzhi820@pingan.com.cn

Mengling Feng
12 Science Drive 2, Singapore
ephfm@nus.edu.sg