Heart Murmur Detection from Phonocardiogram Based on Residual Neural Network with Classes Distinguished Focal Loss

Pan Xia[1,2], Yicheng Yao[1,2], Changyu Liu[1,2], Hao Zhang[1,2], Lirui Xu[1,2], Yuqi Wang[1], Lidong Du[1,2], Yusi Zhu[3], Zhen Fang[1,2,4]

[1]State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
[2]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China
[3]School of Physics and Electronic Information, Yunnan Normal University, Kunming, China
[4]Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences, China

Abstract

The George B. Moody PhysioNet Challenge 2022 focused on detecting the presence or absence of murmurs from multiple auscultation locations heart sound recordings. Our team, MetaHeart, proposed a novel approach to detect heart murmurs by combing residual neural network and class distinguished focal loss. Firstly, the phonocardiogram (PCG) recordings were converted to the time-frequency Mel spectrograms to obtain a richer representation of cardiac mechanical activity. Secondly, a modified residual neural network with 30 layers was designed to extract the complex pathological patterns of heart murmurs. Thirdly, a joint loss combing class distinguished focal loss and center loss was designed in our approach. The class distinguished focal loss can give different degrees of attention to misclassified samples from different categories. The center loss learns a center for deep features of each class. The 15 seconds recordings from five corresponding auscultation locations were preprocessed and concatenated as model inputs for end-to-end training, and the prediction probabilities of 3 murmurs categories or 2 outcome categories were outputs. Finally, our murmur detection classifier received a weighted accuracy score of 0.72 (ranked 18th out of 40 teams) and Challenge cost score of 12536 (ranked 15th out of 39 teams) on the hidden test set.

1. Introduction

Congenital heart diseases affect about 1% of newborns, representing an important morbidity and mortality factor for several severe conditions, including advanced heart failure [1]. The phonocardiogram (PCG) analysis can unveil fundamental clinical information regarding heart malfunctioning caused by congenital and acquired heart disease in pediatric populations. The PhysioNet Challenge 2016 focused on classifying normal and abnormal heart sounds from a single short recording from a single precordial location [2-3]. The 2022 Challenge is devoted to detecting the presence or absence of murmurs from multiple heart sound recordings from multiple auscultation locations, as well as detecting the clinical outcomes [4-5]. The participants were asked to build an algorithm that can identify presence, absence, or unknown cases of murmurs from a patient's recordings and demographic data. Also, they need to identify the normal and abnormal clinical outcomes from such recordings and demographic data. In this work, we proposed a novel approach to achieve the target by combing residual neural network and class distinguished focal loss. The 15 seconds recordings from five corresponding auscultation locations were convert to a 2-dimentional Log-Mel spectrograms as inputs for end-to-end training, and the prediction probabilities of 3 murmurs categories or 2 outcome 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.

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.114

2. Methods

2.1. Datasets and Preprocessing

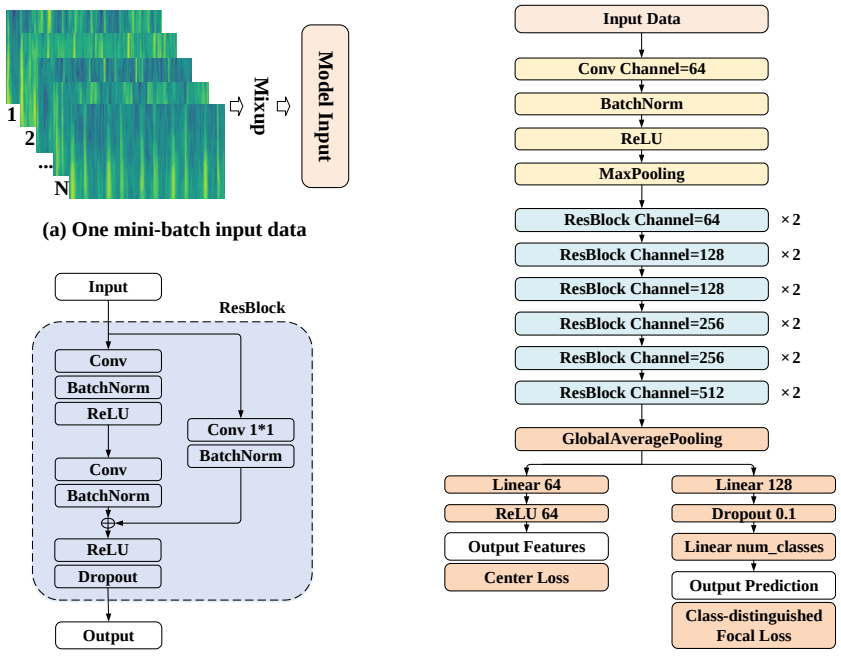
The Challenge data were collected from a pediatric population during two mass screening campaigns conducted in Northeast Brazil in July-August 2014 and June-July 2015 [6]. The Challenge data is composed of the public training set, the hidden validation set and the hidden test set. The public training set contains a total of 3,163 recordings from 942 patients [4]. Each patient in the Challenge data has one or more recordings from one or more prominent auscultation locations: pulmonary valve (PV), aortic valve (AV), mitral valve (MV), tricuspid valve (TV), and other (Phc).

We apply the following data pre-processing procedures. Firstly, the original PCG signals were down-sampled to 1000 Hz. Secondly, to eliminate artificial interference caused by manual operation at the beginning and end of signal acquisition, we discarded the first 5s and last 3s of each recording. The first 15s of truncated recordings were maintained, and then data will be truncated or expanded with 0 to a consistent length. Z-score normalization was applied to normalize each recording. Thirdly, all recordings within one sample were flattened in the order AV, MV, PV, TV, Phc. Finally, the flattened data was converted to a 2-dimensional Log-Mel spectrum with shape of (32, 25001). Where 32 is the size in the frequency dimension and 25001 is the size in the time dimension.

2.2. Model Architecture

Given that deep residual networks [7] allow more efficient feature extraction from longer signals, and the deep residual networks have been widely concerned and proved to be effective [8]. A 30-layer residual neural network was built in our approach. The overall structure of the model is shown in Figure 1. 2-dimensional Log-Mel spectrum maps were fed into the network after data augmentation. The input shape of main network was 32*25001, input data was first fed into a convolution layer with convolution kernel size (3, 11) and using stride 2. 12 residual blocks were stacked to form the backbone of network. After the global average pooling (GAP) operation, the output feature vector is diverted into two heads. The first head converts the feature vectors of length 512 to 64 by using linear operation. Then, the feature vectors are fed into the center loss function. The second head converts the feature vectors of length 512 to 3/2 (number of categories) by using two linear operations with dropout rate 0.1. Finally, the logit outputs are forwarded into the class-distinguished focal loss.

To improve the generalization performance and robustness of neural network when not enough training samples are provided, *Mixup* is adopted in our proposed approach to effective augment data. *Mixup* trains on virtual examples constructed as the linear interpolation of two random examples from the training set and their labels [9]. Assuming that (x_i, y_i) and (x_j, y_j) are two examples drawn at random from our training data in one batch, the mixup



(a) One mini-batch input data

(b) Residual block architecture

(c) Model architecture

Figure 1. The architecture diagram of our proposed approach. (a) One mini-batch input data for the modified residual neural network. (b) Residual block architecture. (c) The model architecture of modified residual neural network.

samples $(\tilde{x}, \tilde{y})$ are created as follows.

$$\tilde{x} = \omega x_i + (1-\omega)x_j$$
$$\tilde{y} = \omega y_i + (1-\omega)y_j$$

Where mixing coefficient $\omega \in [0, \varphi]$ is sampled from a Beta(φ, φ) distribution. The model is trained using the generated mixup samples $(\tilde{x}, \tilde{y})$.

ReLU is adopted to all layers except for the output layer, and the output layer uses *Softmax* because the murmur detection is a mutually exclusive multi-classification task in this year's Challenge. *Softmax* can map the model output to a vector whose probability values add up to 1.

2.3. Loss function

We adopted the joint supervision of classes distinguished focal loss and center loss to train the models for discriminative feature learning. The total loss formulation is given in Eq. 1.

$$L = l_1 + \lambda * l_2 \quad (1)$$

Where the λ denotes a hyper parameter for balancing the two loss functions. The conventional classes distinguished focal loss can be considered as a special case of this joint supervision, if λ is set to 0.

Classes distinguished focal loss

We counted the number of samples for present, unknown, and absent categories in the public training set, there was a significant class imbalance in the Challenge data, which could domains the optimization process, and easily classified samples comprise the majority of the loss and dominate the gradients while under-emphasizing gradients from hard classified samples during training. In addition, in this year's Challenge, owing to the evaluation tendency for each category is different, the hard samples from different categories should be paid attention to differently.

Inspired by Lin et al [10], we proposed a modified focal loss for our model training to alleviate the above problems. Firstly, we define a class-modulation factor θ_i for murmur detection task, which is defined as follows Eq. 2.

$$\theta_i = \begin{cases} 1 & class\ i\ is\ Present \\ 0 & class\ i\ is\ Unknown \\ -1 & class\ i\ is\ Absent \end{cases} \quad (2)$$

To make the algorithm pay more attention to the *Abnormal* category, the class-modulation factor θ_j for outcome classification task is defined as follows Eq. 3.

$$\theta_j = \begin{cases} 1 & class\ j\ is\ Abnormal \\ -1 & class\ j\ is\ Normal \end{cases} \quad (3)$$

Then, classes distinguished focal loss is defined as Eq. 4.

$$l_1 = -\frac{1}{m}\sum_{i=1}^{m} \alpha * (1-p_t)^{\gamma * e^{\mu\theta_i}} * log(p_t) \quad (4)$$

Where p_t denotes the model's corresponding prediction probability for the true label $y_t = 1$. The weighting factor α and the tunable focusing parameter γ are set to 0.5 and 2, respectively. The total average is considered the final loss.

Center loss

The convolutional neural network has been able to capture rich deep features. However, in order to enhance the discriminative power of the deeply learned features, the center loss was adopted in our approach, which is defined as Eq. 5 [11].

$$l_2 = \frac{1}{2}\sum_{i=1}^{n} \left\| \boldsymbol{x}_i - \boldsymbol{c}_{y_i} \right\|_2^2 \quad (5)$$

The $\boldsymbol{c}_{y_i} \in \mathbb{R}^d$ denotes the $y_i\ th$ class center of deep features. The $\boldsymbol{c}_{y_i}$ is updated as the deep features changed.

As above, the center loss simultaneously learns a center for deep features of each class and penalizes the distances between the deep features and their corresponding class centers.

2.4. Model Training

Each model is trained 40 epochs with a batch size of 16 using an NVIDIA GeForce RTX 3090. *Adam* with an initial learning rate of 0.0001 was applied for model optimization. *SGD* with an initial learning rate of 0.5 was applied for center loss parameters optimization. The multiple step learning rate scheduler was adopted to dynamically adjust the learning rate in the training process. The method of reducing the learning rate with a ratio of 0.5 during training was adopted to speed up model convergence. The category decision threshold is set to 0.5 and 0.55 for murmur detection task and outcome classification task, respectively. The main hyper-parameters of two models are showed in Table 1. Other hyper-parameter of the network (convolution kernel size, dropout rate, number of convolution layer, etc.) were adjusted according to the model 5-fold cross validation performance on the public training dataset to achieve optimal performance.

Table 1. The hyper-parameter of the proposed models.

Hyperparameter	Murmur model	Outcome model
epochs	40	40
batch size	16	16
φ	0.1	0.0001
λ	0.1	1
μ	2	0.5

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 obtained were shown in Table 2.

Models	Training	Validation	Test	Ranking
Murmur	0.73±0.05	0.71	0.72	18/40
Outcome	14521±371	9294	12536	15/39

Table 2. Official Challenge scores (weighted accuracy metric scores for murmur classifier, cost metric scores for outcome classifier) for our final selected entry (team MetaHeart), including the ranking of our team on the hidden test set. We used 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.

4. Discussion and Conclusions

In this paper, we proposed a novel approach to detect heart murmurs by combing residual neural network and classes distinguished focal loss and center loss. The PCG recordings were convert to 2-dimensional Log-Mel spectrum to obtain a richer representation of phonocardiograms. Our proposed models were firstly evaluated on the public training set, we achieved 5-fold cross-validation scores of 0.73 and 14521 for the murmur detection task and the outcome classification task with the Challenge evaluation metric. Finally, our murmur detection classifier received a weighted accuracy score of 0.72 (ranked 18th out of 40 teams) and Challenge cost score of 12536 (ranked 15th out of 39 teams) on the hidden test set.

Although our proposed models perform well on the murmur detection task on public training set, it performs poorly on the outcome identification task. An important reason for above problem may be that the presence or absence of murmurs is intuitive and easily distinguishable in PCG recordings. Many teams had also comparatively poor performance on the clinical outcome identification task. In fact, the diagnoses that determined the clinical outcome identification labels required echocardiograms. Therefore, identifying whether a patient is normal or abnormal based solely on the PCG recording is full of huge challenges.

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] D. S. Burstein, P. Shamszad, D. Dai, C. S. Almond, J. F. Price, K. Y. Lin, M. J. O'Connor, R. E. Shaddy, C. E. Mascio, and J. W. Rossano, "Significant mortality, morbidity and resource utilization associated with advanced heart failure in congenital heart disease in children and young adults,". American Heart Journal, vol. 209, pp. 9-19, 2019.

[2] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark. "Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016." In 2016 Computing in Cardiology Conference (CinC), 2016 Sep 11 (pp. 609-612).

[3] G. D. Clifford, C. Liu, B. Moody, J. Millet, S. Schmidt, Q. Li, I. Silva, R.G. Mark. "Recent advances in heart sound analysis," Physiol Meas., vol. 38, pp. E10-E25, 2017, doi: 10.1088/1361-6579/aa7ec8.

[4] Reyna, M. A., Kiarashi, Y., Elola, A., Oliveira, J., Renna, F., Gu, A., Perez-Alday, E. A., Sadr, N., Sharma, A., Mattos, S., Coimbra, M. T., Sameni, R., Rad, A. B., Clifford, G. D. (2022). Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv, doi: 10.1101/2022.08.11.22278688.

[5] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., … & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[6] Oliveira, J., Renna, F., Costa, P. D., Nogueira, M., Oliveira, C., Ferreira, C., … & Coimbra, M. T. (2022). The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification. IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2021.3137048.

[7] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[8] Andreotti F, Carr O, Pimentel M, et al. Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG[C]// 2017 Computing in Cardiology Conference. IEEE, 2017.

[9] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.

[10] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[11] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]//European conference on computer vision. Springer, Cham, 2016: 499-515.

Address for correspondence:

Yusi Zhu
School of Physics and Electronic Information, Yunnan Normal University, No. 1, Yuhua, Chenggong District, Kunming, China.
Zhuyusi16@mails.ucas.edu.cn.

Zhen Fang
State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences,
No. 19, North Fourth Ring West Road, Zhongguancun, Haidian District, Beijing, China.
zfang@mail.ie.ac.cn.