Heart Murmur Detection of PCG Using ResNet with Selective Kernel Convolution

Yonghao Gao[1], Lihong Qiao[1], Zhixiang Li[1]

[1]Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing, China

Abstract

Aims: Heart murmur detection plays a crucial role in the early diagnosis of congenital and acquired heart diseases in children. This study aimed to construct a deep neural network architecture for detecting heart murmurs from PCG recordings. The model was created by the team "fly_h" for the PhysioNet/Computing in Cardiology 2022.

Methods: The PCG signals collected from different auscultation positions were downsampled to 2000Hz, and then a sliding window method was used to clip the signal to 6000 samples. Next, the MFCC features of the PCG signals are extracted. To learn effective features, we constructed a ResNet with selective kernel convolution (SK-Conv). The SK-Conv was embedded into each ResBlock, which adaptively captures multi-scale features using convolution filters of different kernel sizes and applies a channel attention module (similar to Squeeze-and-Excitation) to emphasize the representation of important features. Our model was trained on record-level data and validated on patient-level data. For each patient-level data, it may contain 1-5 PCG recordings, i.e. AV, MV, PV, TV, Phc, and the final prediction result was selected from the corresponding record-level prediction results in the priority order of presence, absence and unknown.

Results: Using the scoring metric based on the costs for algorithmic prescreening for human experts for heart murmur identification, we scored 11331 in the official stage. In addition, the weighted accuracy of the proposed method reached 0.557.

Conclusion: The proposed heart murmur detection model performed well on the validation set. Such models may be used to assist physicians in diagnosing

1. Introduction

Heart murmur detection plays a crucial role in the early diagnosis of congenital and acquired heart diseases in children. Traditional cardiac auscultation is a non-invasive and cost-effective method. However, its accuracy depends on the physician's clinical technique and subjective experience. Therefore, it is very important to develop a computer-aided PCG diagnostic technique.

Over the past few decades, Methods based on phonocardiogram (PCG) anomaly detection have been extensively studied. Overall, these methods divide the process of PCG classification into three main steps: preprocessing, feature extraction, and classification. Among them, the method of feature extraction and the construction of the classifier are the focus of the research. However, traditional PCG anomaly detection methods [1, 2] have some common defects: 1) They require manual feature engineering with a lot of expert knowledge. 2) It is difficult for handcrafted feature-based classifiers to capture the latent connections between features, which affects the classification effect. Recently, deep neural networks have been applied to PCG tasks due to their powerful feature representation capabilities. For time-domain PCG signals, a DNN [3] based on Dense and clique structures has a good performance. In particular, the DNN method proposed by Deng *et al.* using MFCC features has achieved surprising performance.

The 2022 PhysioNet/CinC Challenge focuses on automated open-source methods for the identification of cardiac murmurs from AV, PV, TV, and MV heart sound recordings that may be present in each patient [7]. In this paper, as part of the Physiological Networks/Computation in the Cardiology Challenge 2022, we develop a convolutional neural network (CNN) embedded with selective kernels that integrates feature representations convolved with different receptive fields to improve PCG heart murmur recognition efficiency. We will describe our approach to the challenge.

2. Methods

According to previous research, we designed a detection system with multiple core steps, as shown in Figure 1. The PCG signal was divided into 3s segments by sliding windows. After preprocessing, the MFCC features of heart sounds are fed into the proposed model. Finally, the patient-level classification results are obtained based on a

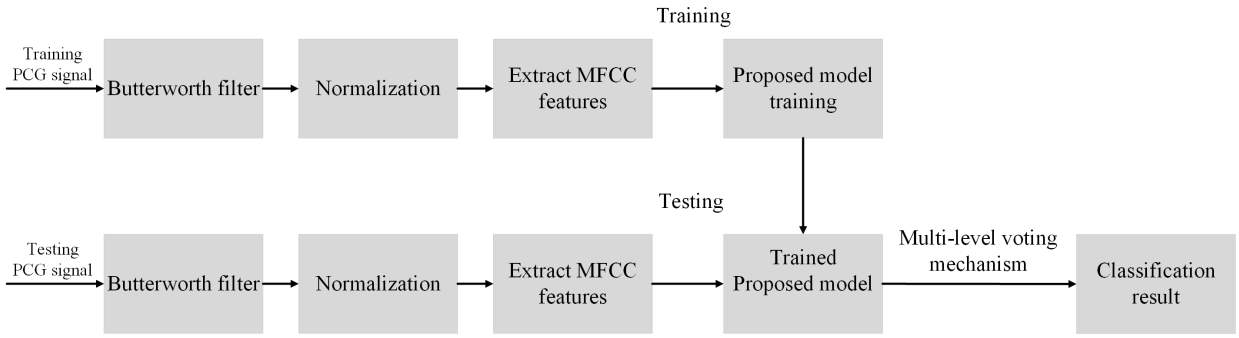

Figure 1. The flowchart of the proposed method.

multi-level voting strategy.

2.1. Data Pre-processing

Filtering and normalization are important preprocessing operations. In our method, we downsample the PCG signal to 2000Hz and use a second-order Butterworth filter for denoising. Since the PCG data is of non-fixed length, we adopt a sliding window algorithm to intercept the heart sound signal into blocks of 3s length with a step size of 1s. It should be noted that we are not using the official segmentation annotation files (with the .tsv extension). Then, the 81-dimensional MFCC from the 256ms window and 128ms frame offset is fed into our model. MFCC is a widely used time-frequency feature extraction method [4, 5]. The MFCC feature representation of PCG is shown in Figure 2. In this paper, we concatenate static MFCC, first-order MFCC, and second-order MFCC in the channel dimension.

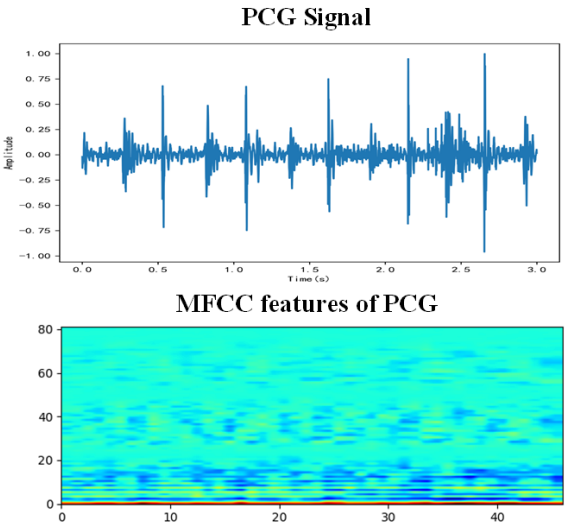

Figure 2. Original signal and MFCC representation of heart sound.

2.2. Model overview

To learn effective features, we constructed a ResNet with selective kernel convolution (SK-Conv). The SK-Conv was embedded into each ResBlock, which adaptively captures multi-scale features using convolution filters of different kernel sizes and applies a channel attention module (similar to Squeeze-and-Excitation) to emphasize the representation of important features.Figure 3 shows the details of the model.We take 81-dimensional MFCCs as feature channels and use 3-layer Resblock with selective convolution kernels to learn high-level knowledge of PCG. Finally, a combination of fully connected layers and softmax is used for heart sound classification.

2.2.1. Selective Kernel Convolution

Selective kernel convolution(SK-Conv) [6] has recently achieved striking success in image processing. In this paper, we embed SK-Conv into Resnet for the task of heart murmur detection.

Suppose the input feature of SK-Conv is $X \in R^{C \times T}$, where C,T denote the dimension of the channel and time axes, respectively. We first construct a 1D convolution $\hat{F}$ with a kernel size of 3 and a dilated convolution $\tilde{F}$ with a kernel size of 3:

$$BN(\mathrm{ReLU}(\hat{F}(X))) : X \to \hat{U} \in R^{C' \times T'}, \quad (1)$$

$$BN(\mathrm{ReLU}(\tilde{F}(X))) : X \to \tilde{U} \in R^{C' \times T'}, \quad (2)$$

where BN and ReLu are batch normalization and ReLU function. $\hat{F}$ and $\tilde{F}$ have the same kernel size. However, $\tilde{F}$ has a larger receptive field. We employ an element-wise summation to integrate information from two branches:

$$U = \hat{U} \oplus \tilde{U}, \quad (3)$$

where $\oplus$ denotes the element-wise summation. U is the global representation that integrates different receptive field features.

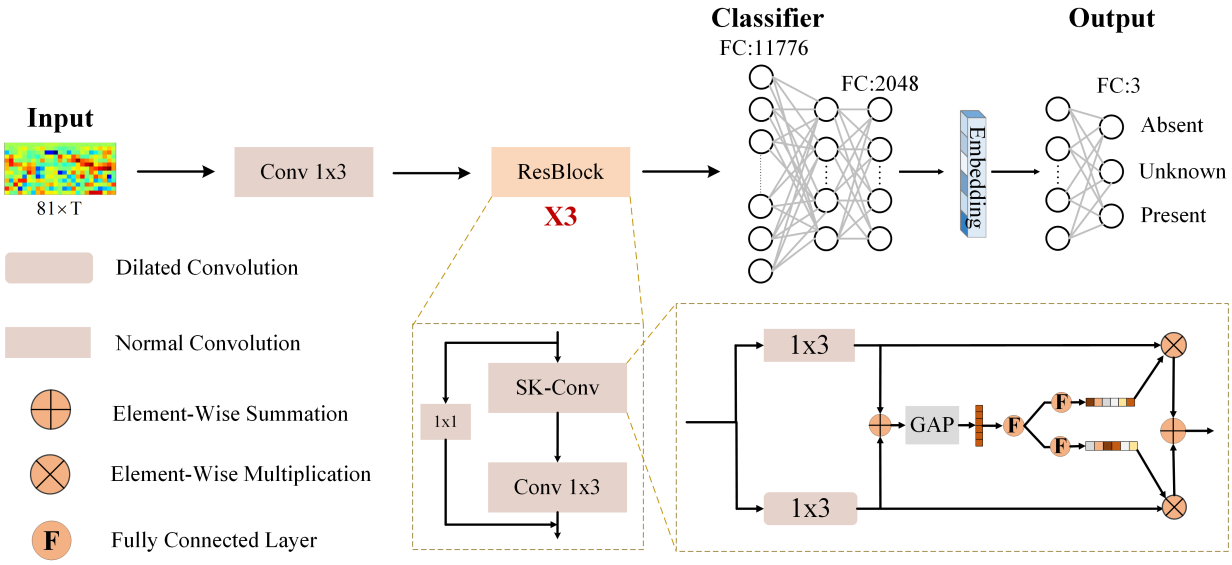

Figure 3. Illustration of the network structure in our proposed method.

Next, we perform a process similar to Squeeze-and-Excitation (SE). Global average pooling (GAP) is used to generate channel-wise statistics $s \in R^C$. Then, two independent fully-connected with softmax modules are used to generate channel excitations $\hat{W}$ and $\tilde{W}$ corresponding to $\hat{U}$, $\tilde{U}$. The final output O of SK-Conv is obtained as follows:

$$O = (\hat{W} \otimes \hat{U}) \oplus (\tilde{W} \otimes \tilde{U}), \tag{4}$$

where $\otimes$ denotes the element-wise multiplication and $O \in R^{C' \times T'}$.

2.2.2. Loss Function

To improve learning efficiency, Center Loss [7] and Focal Loss [8] are used to constrain the model. They are represented as:

$$L_C = \frac{1}{2} \sum_{i=1}^{m} \|x_i - c_{y_i}\|_2^2, \tag{5}$$

$$L_F = -\frac{1}{m} \sum_{i=0}^{m} (1 - \hat{y}_i)^{\gamma} \log(\hat{y}_i), \tag{6}$$

In Equation 5 and 6, $\hat{y}_i$ denote the probability of the i-th sample. The γ is an adjustable hyperparameter that can be adjusted to control the classification of hard-to-classify and easy-to-classify samples. The size of mini-batch is m. $x_i \in R^{2048}$ denote the embedding representation of the i-th sample.The embedding representation is the intermediate layer output of the fully connected layer. $c_{y_i} \in R^{2048}$ represents the class center vector corresponding to label y_i of the i-th sample.

2.3. Voting decision strategy

In the data processing stage, we divide the PCG into several segments of equal length. The model can only output slice-level predictions. Therefore, how to convert the classification results of PCG fragments into patient-level records is crucial. Given the hierarchical relationship of slice-level records, record-level records, and patient-level records. The predicted value of patient-level records can only be aggregated after the record-level record prediction results are obtained first. Therefore, we design a multi-level voting strategy. First, we use slice-level records to aggregate record-level record results based on the majority voting mechanism. Then, we select the prediction result of a PCG record-level record as the final patient-level record prediction value according to the priority order of presence, absence and unknown.

2.4. Training Setup

The proposed model is trained from scratch on a NVIDIA GeForce RTX 3090 GPU. Adaptive moment (Adam) estimation algorithm is used as the optimizer. The hyperparameter γ of the proposed loss function is set to 2. Our proposed model is trained for 40 epochs and a batch size of 512. The learning rate is initially set to 0.002 and the cosine annealing algorithm is used to decay the learning rate to 0.00002 during the training. Finally, 5-fold cross validation is adopted to evaluate the performance of our algorithm.

3. Results

We train and evaluate the models using 5-fold cross-validation, where 4 folds are used for model training and the rest is used as a test set. This was repeated 5 times to produce 5 training models and the scores of the challenge for each model is averaged to obtain an estimate of the model performance. The averaged results are shown in Table 1. We report the accuracy and weighted accuracy as well as the cost. Compared with Resnet, our Sk-Resnet has higher accuracy. The performance of the proposed method in the official phase also proves its effectiveness.

Model	Accuracy	Weighted accuracy	Costs
Resnet	0.8672	-	-
SK-Resnet	0.8862	0.557	11331

Table 1. 5-fold cross-validation results showing challenge scoring for different models.

4. Conclusions

In this paper, we proposed a resnet based on selective kernel convolution, which improves the efficiency of heart murmur recognition by integrating feature representations of different sensory fields. In the training phase, we exploit the clustering properties of Center Loss and Focal Loss to accelerate the learning progress of the model. Compared with Resnet, the method effectively compensates the limitation of a single convolutional kernel and provides multi-sensing of potential pathological information in PCG signals. The experimental results demonstrate the effectiveness of the method. In future work, we will apply the selectable convolutional kernel network to other physiological signal analysis and processing needs.

Acknowledgments

This work was supported in part by the National Key Research and Development Project under Grant 2016YFC1000307-3 and Grant 2019YFE0110800, in part by the National Natural Science Foundation of China under Grant 61806032 and Grant 61976031, in part by the National Major Scientific Research Instrument Development Project of China under Grant 62027827, in part by the Scientific and Technological Key Research Program of Chongqing Municipal Education Commission under Grant KJZD-K201800601.

References

[1] Markaki M, Germanakis I, Stylianou Y. Automatic classification of systolic heart murmurs. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE. ISBN 1479903566, 2013; 1301–1305.

[2] Quiceno-Manrique A, Godino-Llorente J, Blanco-Velasco M, Castellanos-Dominguez G. Selection of dynamic features based on time–frequency representations for heart murmur detection from phonocardiographic signals. Annals of biomedical engineering 2010;38(1):118–137. ISSN 1573-9686.

[3] Xiao B, Xu Y, Bi X, Li W, Ma Z, Zhang J, et al. Follow the sound of children's heart: a deep-learning-based computer-aided pediatric chds diagnosis system. IEEE Internet of Things Journal 2019;7(3):1994–2004. ISSN 2327-4662.

[4] Deng M, Meng T, Cao J, Wang S, Zhang J, Fan H. Heart sound classification based on improved mfcc features and convolutional recurrent neural networks. Neural Networks 2020;130:22–32. ISSN 0893-6080.

[5] Dissanayake T, Fernando T, Denman S, Sridharan S, Ghaemmaghami H, Fookes C. A robust interpretable deep learning classifier for heart anomaly detection without segmentation. IEEE Journal of Biomedical and Health Informatics 2020; 25(6):2162–2171. ISSN 2168-2194.

[6] Li X, Wang W, Hu X, Yang J. Selective kernel networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019; 510–519.

[7] Wen Y, Zhang K, Li Z, Qiao Y. A discriminative feature learning approach for deep face recognition. In European conference on computer vision. Springer, 2016; 499–515.

[8] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. 2017; 2980–2988.

Address for correspondence:

Lihong Qiao
No.2, Chongwen Road, Nan'an district, Chongqing, China
qiaolh@cqupt.edu.cn

Zhixiang Li
No.2, Chongwen Road, Nan'an district, Chongqing, China
1176590542@qq.com