Detection of Murmurs from Heart Sound Recordings with Deep Residual Networks

Lei Hu, Wenjie Cai, Xinyue Li, Jia Li

University of Shanghai for Science and Technology, Shanghai, China

Abstract

Cardiac auscultation is an effective method to screen hemodynamic abnormalities. As part of the George B. Moody PhysioNet Challenge 2022, this paper aims to propose an automated algorithm to identify the presence of murmurs in heart sounds from multiple auscultation locations and to determine whether the heart sounds signal is normal. Two methods are explored. In method one, we perform a series of pre-processing such as denoising and segmentation of the heart sounds signal, extract Log Mel-Spectrogram as features, and use fastai's built-in xResNet 18 pre-trained model for classification. In method two, we extract Mel-frequency cepstral coefficients (MFCCs) as features without any pre-processing and build a customized model based on deep residual networks using one-dimensional convolutional neural layers. Our team, USST_Med, received a challenging score of weighted accuracy of 0.642 (ranked 26th out of 40 teams) and cost of 14529 (ranked 30th out of 39 teams) on the final hidden test set.

1. Introduction

According to the World Health Organization, it is estimated that 17.9 million people died from cardiovascular diseases (CVD). Most populations in low and middle-income countries do not have access to comprehensive primary health care systems, which may delay the diagnosis and treatment of CVD, and lead to early death. It becomes important to explore early prevention methods for heart diseases. Auscultation of the heart plays an important role in the early detection for CVD. However, traditional auscultation methods are highly dependent on physicians' experience, and large-scale auscultation screening of a population is a big challenge for doctors. Therefore, the automatic classification of heart sounds signal is of great practical importance for the screening and diagnosis of heart disease [1]. The George B. Moody PhysioNet Challenge 2022 [2, 3] focused on automated approaches for detecting heart murmurs and predicting clinical outcomes using a new open-source dataset [4].

The fundamental heart sounds (FHS) signal usually includes the first (S1) and second (S2) heart sounds. S1 occurs at the beginning of ventricular systole and S2 occurs at the beginning of diastole. The diastole interval is usually longer than the systole interval. Although the FHS is the most recognizable sound in the cardiac cycle, the mechanical activity of the heart may also cause other audible sounds such as the third heart sounds (S3), the fourth heart sounds (S4), systolic ejection clicks (EC), mid-systolic clicks (MC), diastolic sounds or open pops (OS), and cardiac murmurs caused by turbulent, and high-speed blood flow [5].

Heart sounds segmentation is a crucial step for feature extraction and classification of heart sounds. The primary purpose of segmentation is to divide the heart sounds into four parts: first heart sounds (S1), systole, second heart sounds (S2), and diastole [6]. Many heart sound segmentation methods have been proposed, including ECG signal-based segmentation methods, envelope-based segmentation methods, feature-based segmentation methods, machine learning-based segmentation methods, Hidden Markov Model (HMM)-based segmentation methods, and deep learning-based segmentation methods [7].

After the segmentation, feature extraction is used to convert the original high-dimensional heart sounds signal into low-dimensional features to facilitate the analysis of the heart sounds [6]. In general, the extracted features can be classified into three main types: time domain-based, frequency domain-based, and time-frequency domain-based features.

In the final step, the extracted features are fed into a classifier for feature learning and classification. Recently, with the rapid development of artificial intelligence, deep neural networks (DNN) have been explored for human heart sounds classification. The significant advantage of deep learning algorithms over traditional machine learning algorithms is the feature extraction function for complex heart sounds [8]. Most of the existing works use 1D/2D convolutional neural networks (CNN) [9], recurrent neural networks (RNN) [10], or deep convolutional and recurrent neural networks (CRNN) [8] as their classifiers.

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.047

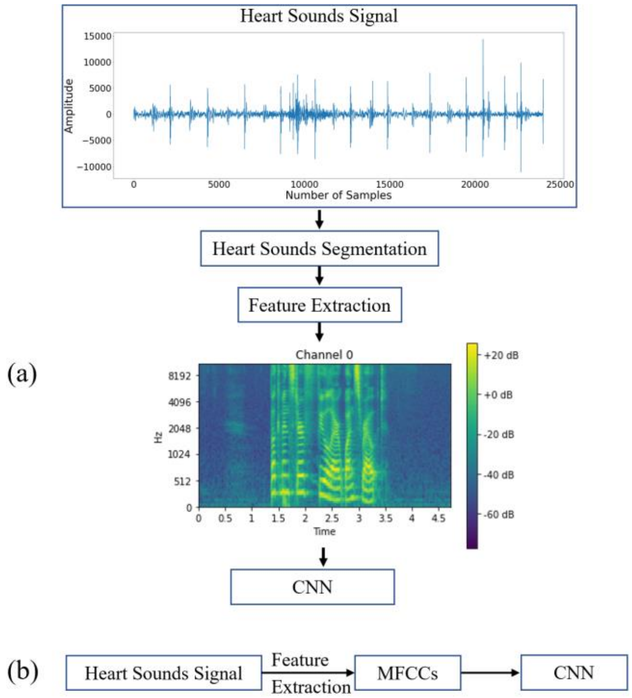

Figure 1. The system-level flow diagram of the proposed methods, (a) denotes method one, (b) denotes method two.

2. Methods

Two methods are applied in this challenge, Method one uses the discrete wavelet transform (DWT) to denoise the heart sounds signal, and a proposed segmentation method to segment the signal. Then the Log Mel-Spectrogram is extracted as input features. Finally, a transfer learning method is used to classify the heart sounds signal. The deep learning model is implemented with fastai [11]. Method two requires feature extraction of Mel-frequency cepstral coefficients (MFCCs) and uses customized deep residual networks to classify the heart sounds. The system-level flow diagram of the proposed methods is shown in Figure 1.

2.1. Heart Sounds Segmentation

Heart sounds signal is interfered by various factors and noises, such as electromagnetic interference (EMI) from the surrounding environment, industrial frequency interference, electrical signal interference from the human body, and breath sound [12]. The presence of noise makes it difficult to localize S1 and S2. Thus, it is necessary to denoise the signal before segmentation.

In this paper, a fifth-level discrete wavelet decomposition using order six Daubechies filter of the original signal is performed to obtain the approximate coefficients, the third-level, fourth-level, and fifth-level

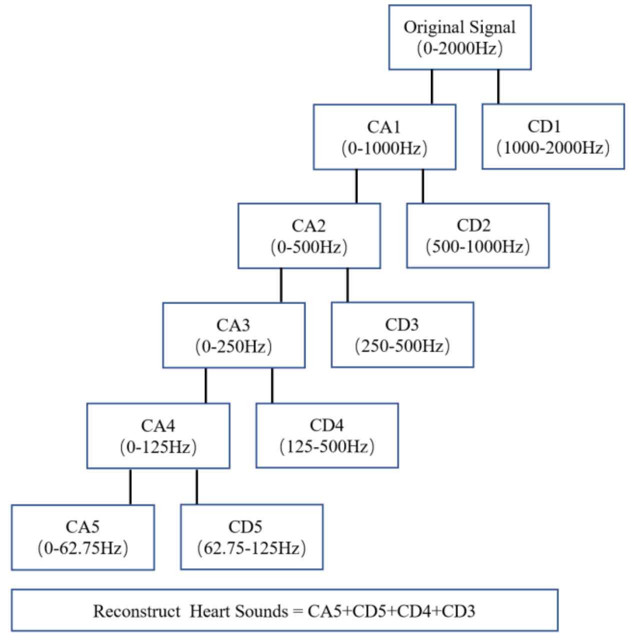

Figure 2. The framework of the corresponding heart sounds signal wavelet denoising algorithm.

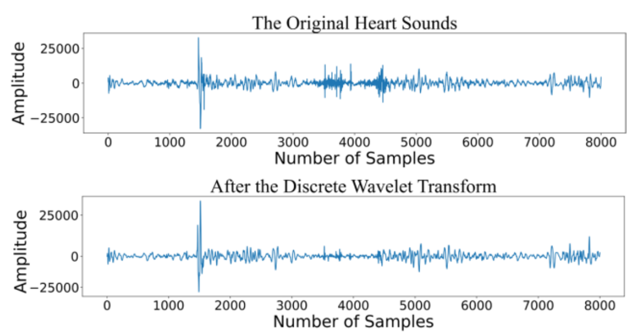

Figure 3. The original heart sounds signal and the signal after discrete wavelet transform.

detail coefficients. Then they are used to reconstruct the signal. The framework of the corresponding heart sounds signal wavelet denoising algorithm is shown in Figure 2. The denoising effect based on wavelet decomposition is shown in Figure 3.

After denoising, the heart sounds signal was normalized by equation 1 and equation 2:

$$X(t) = \frac{x(t) - \mu}{\sigma} \tag{1}$$

where $x(t)$ is the filtered heart sounds signal, μ is the mean of the filtered signal, and σ is the standard deviation of the filtered signal.

$$s(t) = \frac{X(t)}{\max(X(t))} \tag{2}$$

In this work, instead of choosing the envelope-based segmentation method, we directly localize the peak of the filtered signal.

First, the filtered signal s(t) is differentiated to provide information about the slope by equation 3:

$$d(t) = s(t+1) - s(t) \tag{3}$$

Then, points with a positive slope are marked as 1, and others are marked as 0. Only when the current point is marked as 1 and the next point is marked as 0, it is considered a potential S1 peak. We use the Python package PeakUtils to detect the final peak. The threshold is set as 0.3 and the minimum distance between each detected peak is set as 1800 after several experiments. At last, all S1 peaks are identified for segmentation.

After the S1 peak detection of the heart sounds signal, we segment the signal to every five cardiac cycles (S1-S1) until the last S1. Due to the inconsistent length of the segmented heart sounds signal, we fixed the signal length to 8800 and used zero padding for insufficient length. The distributions of two types of labels of the segments are shown in Table 1 and Table 2.

Labels	Number	Percentage
Present	1963	17.5
Absent	8962	80
Unknown	290	2.5
Total	11215	100

Table 1. The sample distribution of “Present” “Absent” and “Unknown”.

Labels	Number	Percentage
Abnormal	5482	49
Normal	5733	51
Total	11215	100

Table 2. The sample distribution of “Abnormal” and “Normal”.

2.2. Feature extraction

The Log Mel-spectrogram and MFCCs have been widely used in audio signal processing. In this stage, due to the excellent performance of CNN in image processing, we convert the audio signal processing problem into image processing by extracting frequency features. In method one, we extract the Log Mel-Spectrogram out of each segment heart sounds signal. Since the imbalance of our dataset, spectrogram augmentation techniques are applied [13]. A random time and frequency domain mask on the Log Mel-spectrogram is used in this work.

In method two, we consider the MFCCs features as time-series data instead of images. For the classification of the presence, absence, or unclear cases of murmurs, we randomly segment MFCCs in a shape of (13, 2500). For the classification of the normal and abnormal clinical outcomes, we randomly segment MFCCs in a shape of (13, 1500). On both classification tasks, zero padding is used for insufficient length.

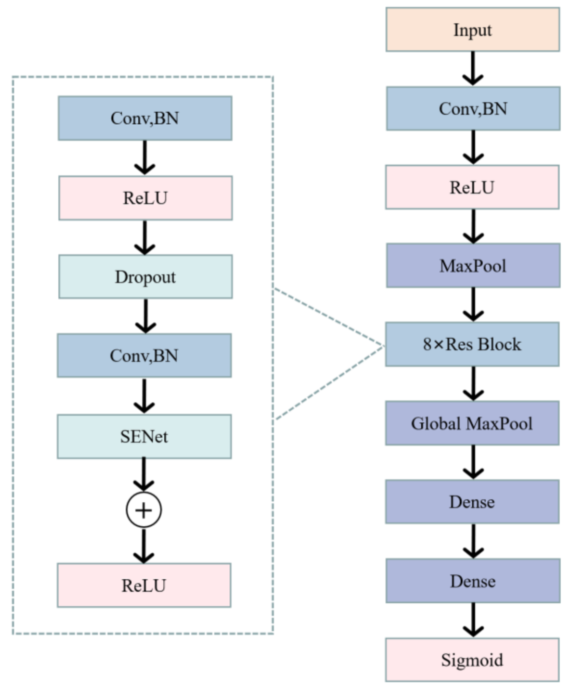

Figure 4. The CNN network architecture used in method two

2.3. Classification

In method one, instead of building a model from scratch, we select the xResNet 18 pre-trained model built into fastai to classify the heart sounds signal. Fastai has a very easy-to-use workflow, which makes the process of debugging much simpler. We use the ‘lr_find’ method to find the optimal learning rate, and in this work, we choose 3e-3. Due to the ‘fit_one_cycle’ method, the learning rate is uniformly varied over 20 training epochs, Callbacks and EarlyStopping are used to efficiently track training and validation losses at the specified learning rate.

The CNN network architecture in method two is shown in Figure 4. One-dimensional convolution layers are used to extract class-related features, and the main architecture is based on eight residual blocks [14]. For both tasks, the heart sounds signal from multiple auscultation locations is input to the network, after convolutional and global max-pooling layers, all the outputs are concatenated. The last fully connected layer with the Sigmoid function returns the result of classification.

3. Results

After model training, we evaluated the performance of the classifiers on the online validation dataset. We got a score of weighted accuracy of 0.518 and a cost of 11752 with method one. The method two achieved a score of weighted accuracy of 0.56 and a cost of 11114 on the online validation dataset. It got a score of weighted accuracy of 0.68 and a cost of 10852 on the hold-out dataset which accounted for one-fifth of the public training set. The official test score was based on method two. Its weighted accuracy was 0.64 and its cost was 14529, which ranked 26th and 30th in all challenge participants, respectively (Table 3).

Metrics	Training	Validation	Test	Ranking
Weighted accuracy	0.68	0.56	0.64	26/40
Clinical outcome	10852	11114	14592	30/39

Table 3. Cost metric scores (official Challenge score) for our final selected entry (team USST_Med) for the clinical outcome identification task on method two, including the ranking of our team on the hidden test set. Our algorithm was tested on part of the public training set, validated on the hidden validation set, and evaluated on the hidden test set.

4. Discussion and Conclusions

In this paper, we proposed two methods for heart sounds signal classification. Our results are not satisfactory. The pre-trained model used in method one is designed for natural images which are totally different from the extracted Log Mel-spectrogram. Due to relatively small amounts of data, fine-tuning does not achieve optimal results. The segmentation method is not robust enough to get an accurate result in the validation set, which limits the model's performance. In method two, we randomly segment fixed-length MFCCs features as model inputs. These inputs could be just some noisy signals, which may prevent the model from achieving good results.

In future work, we can try different denoising methods to obtain better quality heart sounds signal. We can conduct more in-depth studies for different hyperparameters (window size, hop length, number of MFCCs features, FFT size, etc.) to obtain the optimal features when performing feature extraction. In addition, we can compare the classification effects of different types or sizes of pre-trained models. Finally, the CNN network architecture used in method two achieves a good result on electrocardiogram (ECG) classification, we can compare the differences between heart sounds and ECG to optimize the network architecture structure for better heart sounds signal classification.

In conclusion, although we did not achieve satisfactory scores, our approach at least demonstrates the good potential of deep learning for heart sounds classification.

References

[1] Chen JF, Dang X. Heart Sound Analysis Based on Extended Features and Related Factors. IEEE Symposium Series on Computational Intelligence (SSCI) 2019;2189–2194; URL https://doi.org/10.1109/SSCI44817.2019.9003008

[2] Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[3] Reyna MA, Kiarashi Y, Elola A, et al. Heart Murmur Detection from Phonocardiogram Recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv 2022; URL https://doi.org/10.1101/2022.08.11.22278688.

[4] Oliveira J, Renna F, Costa PD, et al. The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification. IEEE Journal of Biomedical and Health Informatics 2021;26(6):2524–2535.

[5] Liu C, Springer D, Li Q, et al. An Open Access Database for the Evaluation of Heart Sound Algorithms. Physiological Measurement 2016;37(12):2181-2213.

[6] Chen W, Sun Q, Chen X, et al. Deep Learning Methods for Heart Sounds Classification: A Systematic Review. Entropy 2021;23(6):667.

[7] He Y, Li W, Zhang W, et al. Research on Segmentation and Classification of Heart Sound Signals Based on Deep Learning. Applied Sciences 2021;11(2):651.

[8] Deng M, Meng T, Cao J, et al. Heart Sound Classification Based on Improved MFCC Features and Convolutional Recurrent Neural Networks. Neural Networks 2020;130:22-32.

[9] Abduh Z, Nehary EA, Wahed MA, et al. Classification of Heart Sounds Using Fractional Fourier Transform Based Mel-frequency Spectral Coefficients and Traditional Classifiers. Biomedical Signal Processing and Control 2020;57:1746-8094.

[10] Khan FA, Abid A, Khan MS. Automatic Heart Sound Classification from Segmented/Unsegmented Phonocardiogram Signals Using Time and Frequency Features. Physiological Measurement 2020;41(5):055006.

[11] Howard J, Gugger S. fastai: A Layered API for Deep Learning. Information 2020;11(2):108.

[12] Ali MN, El-Dahshan EA, Yahia AH. Denoising of Heart Sound Signals Using Discrete Wavelet Transform. Circuits Syst Signal Process 2017;36,4482–4497.

[13] Park DS, Chan W, Zhang Y, et al. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. Interspeech 2019;2613–2617. URL https://dx.doi.org/10.21437/Interspeech.2019-2680

[14] Cai W, Liu F, Xu B, et al. Classification of Multi-lead ECG with Deep Residual Convolutional Neural Networks. Physiological Measurement 2022;43(7):074003.

Address for correspondence:
Wenjie Cai
School of Health Science and Engineering,
University of Shanghai for Science and Technology,
516 Jungong Road, Yangpu Distric, Shanghai, China