Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Research on heart sound detection and recognition based on convolutional neural network

Libiao Wang, Feng Li*

School of management science and Engineering, Anhui University of Finance and Economics,

Bengbu 233000, China

*Corresponding author Email: feli@ustc.edu.cn

Abstract: Cardiovascular diseases have become one of the most prevalent diseases in the world, and the detection and recognition of heart sounds plays an important role in early diagnosis. A heart sound recognition algorithm based on convolutional neural network is proposed to diagnose heart and blood diseases based on heart sound signals. Firstly, the structural characteristics of the heart and the generation of heart sounds are analyzed. Secondly, the analysis and extraction of heart sound features are realized by pre-processing the heart sound signal. Finally, the training was carried out by convolutional neural network, and the accuracy of the algorithm for heart sound recognition was effectively improved after experimental comparison.
Keywords: Heart sound detection, heart sound recognition, convolutional neural network, cardiovascular disease.

---

1. INTRODUCTION

At present, heart sound auscultation technology is one of the main clinical diagnostic methods for the treatment of cardiovascular diseases [1]. It is non-invasive, efficient and convenient. It can obtain physiological and pathological information about the heart. However, due to the complex clinical diagnostic conditions and a large amount of noise pollution, doctors who lack rich experience are often disturbed by environmental noise, Lead to inaccurate diagnosis of the disease. In 1929, German doctor Werner used a catheter to deliver drugs to the heart, opening the door to the study of cardiovascular diseases using physical models; In the 1970s, Dr. Marcus of the United States used angiography to observe the causes of cardiovascular diseases, overturning the long-standing misconception of heart disease; In the 1980s, the earliest cardiac defibrillator at Johns Hopkins University was put into clinical use, and the earliest telemetry system was developed, so that doctors can observe the vital signs of patients with heart disease from a long distance; In recent years, with the development of technology, devices such as integrated ECG and heart sound analyzer and intelligent electronic stethoscope have been put into clinical application. However, due to the inevitable factors in the use process, the collected heart sound signal will contain all kinds of noise to varying degrees,

affecting the final diagnosis results. At present, digital filter, wavelet decomposition and empirical mode decomposition are widely used in the digital denoising of heart sound signals. In recent years, with the rise of artificial intelligence, big data and other technologies, more accurate and effective heart sound detection methods are expected to be realized. This paper intends to use the heart sound recognition algorithm based on convolutional neural network to improve the recognition efficiency of cardiovascular diseases.

Convolutional neural network (CNN) is a feedforward neural network of heart sound. It is widely used in medical speech intelligent recognition and image recognition, and has obtained remarkable research results. CNN network can provide translation invariant convolution of heart sound at all levels at the same time. This deep thought algorithm of CNN is applied to the thinking model of medical heart sound intelligent recognition, Using the deep algorithm CNN widely used in speech and image to recognize the heart sound of its signal can not only effectively overcome the repeated and mixed heart sound noise in medical heart sound, but also provide a new idea of intelligent research for the heart sound intelligent recognition and analysis of non-stationary medical signals such as medical heart sound signals.

According to the idea of intelligent analysis of heart sound signal by using convolutional neural network, this paper introduces the structure and function of heart, the generation and characteristics of heart sound source components in turn, then introduces the design process of convolutional neural network and heart sound recognition algorithm, provides the research method of heart sound recognition by convolutional neural network, and finally shows the experimental background, experimental data Algorithm model and experimental results show the advantages of convolutional neural network in heart sound recognition.

2. MECHANISM OF HEART SOUND

2.1 Cardiac structure and function.

The heart is the most important organ of the human body, which can contract and relax rhythmically. It can be divided into two parts. The left contains the upper left ventricle and the lower left ventricle, and the right contains the upper right atrium and the lower right ventricle, with a total of four cavities [2]. The ventricles are separated from the atrium and cannot be connected, and the muscle and heart wall of the left ventricle are thicker than those of other atriums and ventricles, so as to provide strong power for promoting blood movement.

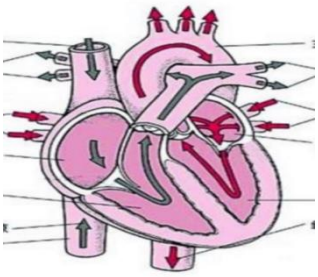

Fig.1 Cardiac structure diagram

As shown in Figure 1, we can clearly see that the heart wall of the left ventricle is the thickest, because the left ventricular capillaries contract rapidly and transport blood to the whole heart and

various organs of the whole body. The heart wall where the heart tissue membrane is located at the junction with cells is the heart capillaries, which are far away and need great power. According to the above structure, we can tank water into the pulmonary vein and drain the water out of the heart through the pulmonary vein - > left ventricle - > left ventricle - > median vessel. The main parts are atrioventricular slices and arterial valves. These common functions ensure that the flow of blood in our body can flow into the whole body and all major physiological parts of our body in the whole heart, rather than direct backflow.

The function of the heart is to send venous blood from the body to the lungs and arterial blood from the lungs to organs and tissues. There is more carbon dioxide produced by the respiration of body tissues and organs in the venous blood. The right ventricle sends the venous blood to the lungs. The arterial blood is oxygen rich blood just from the lungs, The left ventricle pumps arterial blood to various organs and tissues of the body, passing through the aorta (aorta). There are many valves on the myocardium, the mitral valve, which can prevent blood flow from the left ventricle to the left atrium; Large artery valve (aortic valve), which prevents blood flow from the median vessel to the left ventricle. When the heart works normally, the left atrium receives oxygen enriched arterial blood from the lung and then pumps it into the left ventricle. After the left ventricle is filled, the blood passes through the aortic valve and is pumped to the whole body.

2.2 Generation and characteristics of cardiac sound source componentsage Numbers.

Heart sound refers to the sound generated by the mechanical wave phenomenon formed when myocardial atrophy, cardiac valve atresia and blood flow impact the atrial wall and vascular wall. In some relatively small areas on the heart wall or chest wall, the heart sound can sometimes be heard in real time even with visceral ultrasound or stethoscope. Sometimes, some acoustic detection instruments such as visceral heart sound ultrasonic transducer can be used to record some mechanical waves that produce visceral heart sound in real time. The graph of Mechanical waves changing with time is called phonocardiogram.

A cardiac cycle consists of a systole and a diastole. According to the relative position of each heart sound, the motion mechanism and the formation time of heart sound are roughly composed of the first heart sound, the second heart sound, the third heart sound and the fourth heart sound. Most people can hear the first and second heart sounds during auscultation. The third heart sound is relatively weak, while the fourth heart sound is even weaker than the third heart sound, but pregnant women and children can sometimes hear the third heart sound.These four heart sound signals are generated by different principles [3], as follows:

First heart sound: it is generated after the closure of mitral and tricuspid valves. The tone is low and lasts for a long time. The apex of the heart is the loudest.

Second heart sound: it is produced by the sudden closure of the main and auxiliary pulmonary valves and the sudden decline of blood from the main and auxiliary pulmonary arteries, resulting in valve vibration. The tone is high and crisp, but the wavelength is short, weaker than the first heart sound, and the duration is short, and the bottom of the heart is the loudest.

Third heart sound: it is the sound formed by the rapid filling of blood in the atrium and the vibration generated by the impact of the atrium on the ventricular wall. The pitch is light and low, and the time is not long. It is only at the apex of the heart or above it.

Fourth heart sound: the atrioventricular slices and corresponding structures are suddenly tense and vibrated due to ventricular contraction. It is relatively weak and pathological.

Under normal circumstances, the normal periodic movement of various organs in the heart received by the heart sound collector makes the heart sound periodic signal have quasi periodic characteristics. When analyzing heart sound signals, a cycle is also called a cardiac cycle. In each cardiac cycle, the heart relaxes, the atrioventricular valve closes, the indoor pressure rises the fastest, the arterial pressure is the lowest, and the vena cava blood flows back into the heart; The heart contracts, the atrioventricular valve opens and the internal pressure rises, pumping blood to the arteries. It can be seen that heart sound is mainly composed of two motion stages: systolic stage and diastolic stage, and there is a static period in both stages.

3. HEART SOUND RECOGNITION ALGORITHM BASED ON CONVOLUTIONAL NEURAL NETWORK

3.1 Convolutional neural network.

Convolutional neural network is developed from traditional artificial neural network. It not only has the characteristics of traditional fully connected neural network, but also has many differences and improvements on this basis. The basic principle of convolutional neural network is to convert the original data into two-dimensional matrix format, which is superior to the traditional artificial neural network in the extraction performance of image eigenvalues. In CNN, the function of initial convolution layer is similar to that of edge detector, which can be used to identify low-level features, that is, low-dimensional features. Although the network near the convolution layer is more sensitive, more complex or abstract high-dimensional, because of the weight and sparse connection sharing characteristics of CNN, and its network requires less parameter training than fully connected to the Internet, it shows that when the network layer and each layer output, the dimension required by CNN to process the same data in the stage is far lower than that of the whole connected to the Internet, that is, CNN has higher operation efficiency. Compared with other feature extraction methods, CNN has the characteristics of simple structure, no manual setting of parameters, strong fitting ability and trainability. Therefore, this paper selects CNN network to train heart sound algorithm.

3.2 Flow design of heart sound recognition algorithm.

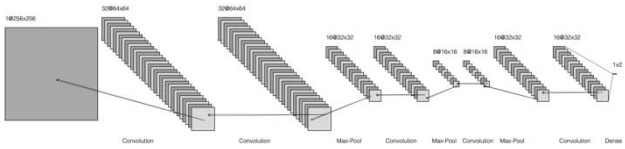

Fig.2 Network structure diagram

After the processing shown in Fig. 2, the required classification result is obtained.

The activation function of each layer is: relu, and the last full connection classification activation function is: softmax.

Convolution neural network can be understood as a kind of down sampling. For an image, the convolution kernel is scanned from the upper left corner to the lower right corner at one time according to the size of convolution kernel and step size, as shown in Figure 3.

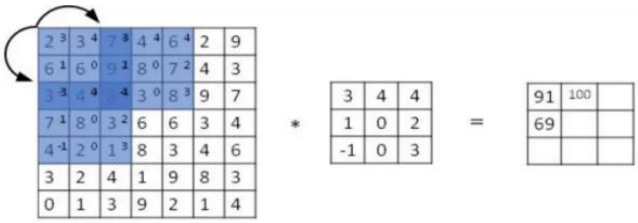

Fig.3 Convolution diagram

For signal, convolution neural network is the same as signal convolution, in essence, that is:In essence, the signal of each dimension is convoluted [4] to reduce signal redundancy. However, such feature extraction is prone to local convergence after the bias of the activation function. Therefore, it is usually necessary to add a pooling layer after one or more convolution layers. The basic idea of the pooling layer is to mark (replace) the new matrix obtained by convolution with some features (average, maximum and minimum). Because the information difference between the first two convolution layers is very large, the convolution neural network is directly connected with the convolution neural network, and then through the pooling layer [5,6,7]. Twice compression compresses the signal to a size that is easy to deal with. At this time, the idea of cotconv convolution is adopted, and the down sampled signal is up sampled once to further reduce the spatial redundancy and obtain the signal matrix with low information entropy. Naturally, through such processing, the spatial frequency is divided. Each layer of signal matrix at different depths is equivalent to a group of signal matrices with the same frequency. The classification results are obtained through the subsequent fully connected network and activation function.

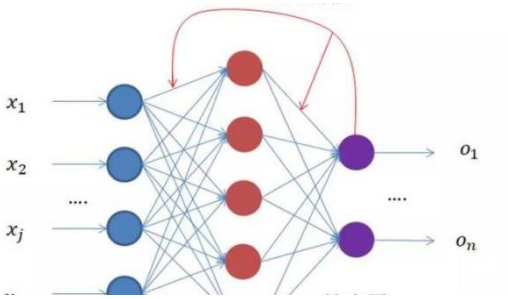

Fig.4 Schematic diagram of the whole connection layer

The activation function of each layer of the network is: relu () so that the matrix of neurons to be output each time becomes. This activation function can better extract the effective information in the matrix, avoid gradient explosion or gradient disappearance, and better cooperate with the softmax activation function of the full connection layer.

4. HEART SOUND RECOGNITION ALGORITHM BASED ON CONVOLUTIONAL NEURAL NETWORK

4.1 Experimental background.

In the past decades, heart sound signals (i.e. echocardiography or PCGs) have been widely studied. Automatic heart sound segmentation and classification technology has the potential to screen pathology in various clinical applications. However, due to the lack of a large and open heart sound recording database, the comparative analysis of algorithms in the literature has been hindered. In 2016, PhysioNet / computing # in # Cardiology (CINC) challenge opened the largest heart sound

database to solve this problem. The database comes from 8 sources of 7 independent research groups around the world. The database included 4430 recordings from 1072 subjects, and a total of 233512 heart sounds from various groups were collected. These records were collected using heterogeneous devices in clinical and non clinical (such as home visits). The recording time ranges from a few seconds to a few minutes. Other data provided include subject demographics (age and gender), recorded information (number of patients per patient, body position and length of recording), synchronously recorded signals (such as ECG), sampling frequency and type of sensor used. Participants were asked to rate the recordings as normal, abnormal or impossible (noisy / uncertain).

4.2 Data sets.

For a total of 8 independent heart sound databases, 4 of them are used as training and test sets for 30-70 training tests. The remaining four databases are specifically assigned to the training set (or test set). Training and test sets are two mutually exclusive groups (i.e., there are no records from the same subject / patient in both training and test sets). The challenge training set (a to f) included a total of 3153 cardiac recordings from 764 subjects / patients, and the test set (B to e, plus g and I) included a total of 1277 cardiac recordings from 308 subjects / patients. After the manual correction process of segmentation annotation, there are 84425 times in the training set and 32440 times in the test set. The recording duration ranges from a few seconds to more than 100 seconds. All data were resampled to 2000 Hz and used in uncompressed wav format. Table 1 briefly summarizes the data of the "challenge". The number distribution of training sets is shown in Table 1.

Tab.1 Data set structure

file name	quantity
training-a	409
training-b	490
training-c	31
training-d	55
training-e	2054
training-f	114

4.3 Algorithm model.

In this experiment, res net in CNN model is used for speech recognition training. The network structure of res net is shown in Figure 5.

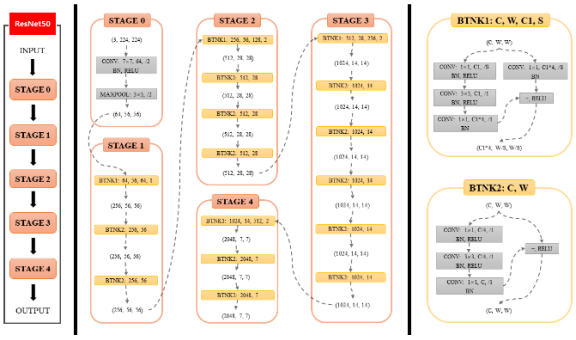

Fig.5 Network structure diagra

The difficulty of CNN model training has always been a problem. For example, VGg has only 19 layers. Once the number of layers is too deep, the gradient of the model will disappear, which makes

it more difficult to train the deep network model. Res net alleviates this problem by proposing the residual module. The residual module is shown in Figure 6.

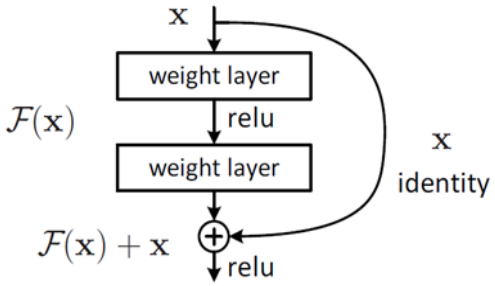

Fig.6 residual structure

Intuitively, the residual block directly bypasses the input information to the output to protect the integrity of the information. The whole network only needs the part of the difference between input and output, which simplifies the experimental goal and difficulty.

4.4 Modeling.

First, generate a data list. Because before processing the data, we write the voice data name and its path under each category in one Txt folder, so that subsequent data can be read.

Then, use librosa Load reads voice data, and its original data waveform is shown in Figure 7.

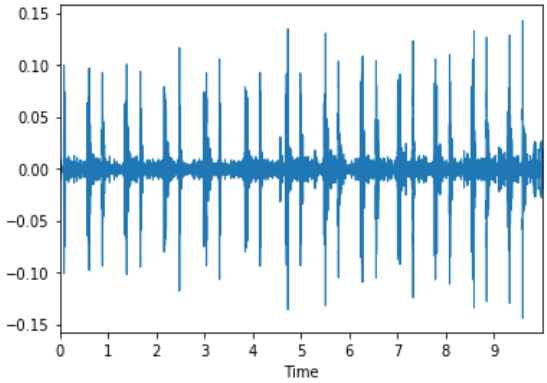

Fig.7 raw data signal

Use librosa feature. Melspectrogram calculates Mel spectrum, and its visualization is shown in Figure 8.

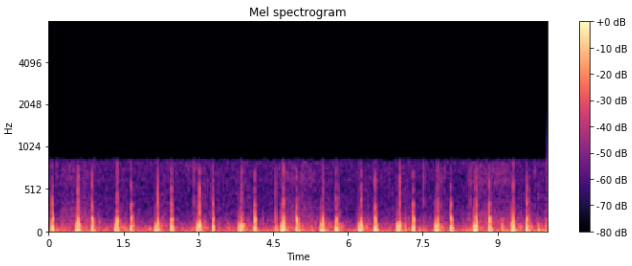

Fig.8 Mel spectrum

In this experiment, res net in CNN model is used as the training model, and the model parameters are shown in Table 2.

Tab2. network parameters

Parameter name	Parameter interpretation	parameter
batch_size	Batch size of training	32
num_workers	Number of threads reading data	10

num_epoch	Number of rounds of training	50
num_classes	Number of categories classified	6
learning_rate	Initial learning rate	1e-3
train_list_path	Data list path of training data	'data/train_list.txt'
test_list_path	Data list path of test data	'data/test_list.txt'
save_model	Path to model save	'models/'

The loss function in the model uses cross entropy loss, which is optimized by Adam and attenuated by steplr.

The loss function decline process and accuracy change process during training are visualized as shown in Figures 9 and 10.

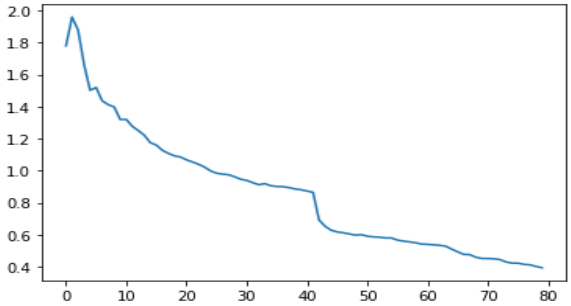

Fig.9 Loss function descent process

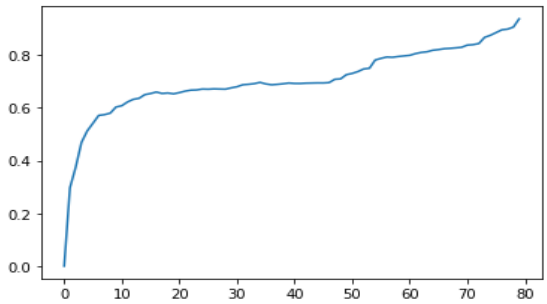

Fig.10 Accuracy change process

5. EXPERIMENTAL RESULT

In order to compare the experimental results, the methods of the top nine contestants in the 2016 PhysioNet cardiology calculation challenge were selected as the comparison. The comparison results are shown in Table 3.

Tab.3 Comparative experiment

Refference	ACC	Method
Potes et al.[8]	0.8602	AdaBoost & CNN
Zabihi et al[9]	0.8590	Ensemble of SVMs
Kay & Agarwal[10]	0.8520	Regularized Neural Network
Bobillo[11]	0.8454	MFCCs, Wavelets, Tensors & KNN
Homsi et al[12]	0.8448	Random Forest + LogitBoost
Maknickas[13]	0.8415	Unofficial entry - no publication
Plesinger et al[14]	0.8411	Probability-distribution based
Rubin et al.[15]	0.8399	Convolutional NN with MFCs

Abdollahpur et al.[16]	0.8263	HMM
Ours	0.9075	Res-Net

Among the top nine methods, the ten fold cross validation method is basically used to increase the performance of the model, and the weighted voting algorithm is used in the final result to further improve the result. They mostly use some scoring skills and do not improve the performance of the model fundamentally, In addition, we can also find that the top ten methods mostly use machine learning methods, and only a few combine or use deep learning methods, and the players who use deep learning methods rank very high. It can be seen that in terms of feature extraction, deep learning methods are more powerful and the extracted features are more robust, Our method can achieve such good results, which depends on the deeper network structure layer of our model. Through the residual module, we can effectively alleviate the gradient disappearance or gradient explosion, so that the model can extract deeper useful features, so that the model can learn more useful information, and improve the prediction effect of the model from the method level.

In this paper, a heart sound recognition algorithm based on convolutional neural network is proposed, which is combined with deep learning method to deal with the recognition of cardiovascular diseases. By analyzing the causes and characteristics of heart sound signal and the environment diagnosed by clinicians, a convolution neural algorithm which can be applied to various listening devices and can collect, separate, classify and recognize heart sound is designed. On this basis, the experiment is carried out. Firstly, the heart sound is preprocessed, and then the res net in CNN model is used as the training model for training. The training results show that convolutional neural network is a good method to recognize heart sound in complex clinical environment. This is undoubtedly a new way of heart sound recognition algorithm.

ACKNOWLEDGMENTS

This work was financially supported by Preferential subsidy program for overseas students' innovation project (Project No.: 2021lcx032); Innovation and entrepreneurship training program for college students (Project No.: 202110378072).

REFERENCES

[1] Yu Yunzhi，Nie Bangji.Clinical significance of heart sounds and current status of research[J]. Modern medical instruments and applications, 1997(03):9-12.
[2] Chow CL,Chen S S M,Gaal W V.An unusual finding in the right atrium[J].Heart(British Cardac Society) ,2016,103(2):heartjn-2016-311056.
[3] Liu C, Springer D, Li Q. et al. An open access database for the evaluation of heat sound algorithms[J],P hysiological Measurement, 2016,37(12):2181-2213.
[4] Ioffe S,Szegedy C.Batch Normalization:Accelerating Deep Network Training by Reducing Internet Cov ariate Shift[J].2015：448-456.
[5] Hung S L,Adeli H.Parallel backpropagation algorithms on CRAY Y-MAP8/864 supercomputer[J].Neurc omputing,1993,5(6):287-302.
[6] Hinton G E,Srivatava N,Krizhevsky A,et al.Improving neural networks by preventing co-adaptation of f eature decetors[J].Computer Science，2012，3(4):212-223.
[7] Mousavi Z,Rezail T Y,Sheykhivand S,et al.Deep convolutional neural network for classification of sleep stages from single-channel EEG signals[J]. Journal of Neuroscience Methods, 324:108312-108312.
[8] Potes C , Parvaneh S , Rahman A , et al. Ensemble of Feature-based and Deep learning-based Classifie rs for Detection of Abnormal Heart Sounds[C]// 2016 Computing in Cardiology
[9] Conference (CinC). Piscataway, NJ;IEEE, 2016.43:621-624.

[10] Zabihi M , Rad A B , Kiranyaz S , et al. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation[C]// 2016 Computing in Cardiology Conference (CinC). Piscata way, NJ;IEEE, 2017:180-213.
[11] Edmund, Kay, Anurag, et al. DropConnected neural networks trained on time-frequency and inter-beat f eatures for classifying heart sounds[J]. Physiological Measurement, 2017：1645 - 1657.
[12] Bobillo I D . A Tensor Approach to Heart Sound Classification[C]// Computing in Cardiology Conferenc e.Piscataway, NJ; IEEE, 2017.184-315.
[13] Homsi M N , Medina N , Hernandez M , et al. Automatic Heart Sound Recording Classification using a Nested Set of Ensemble Algorithms[C]// Computing in Cardiology 2016.Piscataway, NJ; IEEE,2016.2 37-325.
[14] Maknickas V , Maknickas A . Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients[J]. Physiological Measurement, 2 017：1671-1684.
[15] Plesinger F , Jurco J , Jurak P , et al. Robust Multichannel QRS Detection[C]// Computing in Cardiolo gy. Piscataway, NJ;IEEE, 2014.:557-560.
[16] Rubin J , Rui A , Ganguli A , et al. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients[C]// 2016 Computing in Cardiology Conf
[17] Ab Dollahpur M , Ghiasi S, Mollakazemi M J , et al. Cycle selection and neuro-voting system classifying heart sound recordings[C]//Computing in Cardiology Conference. Piscataway, NJ;IE EE, 2017.176-238.erence (CinC). Piscataway, NJ;IEEE, 2017.236-175.