# Classification of Phonocardiograms Using Residual Convolutional Neural Network and MLP

Guohui Peng[1], Haitao Zou[1], Jin Wang[1]

[1]School of Computer, Jiangsu University of Science and Technology, Jiangsu, China

## Abstract

*Our method does not require signal segmentation or segment identification, it is simpler than identifying and segmenting S1, systole, S2 and diastole.*

*Our team is JUST_IT_Academy1, in this competition, we design our strategy aiming at avoiding signal segmentation and segment identification. During our data preprocessing, each audio data is converted to a 128-dimensional vector by computing its Mel-scaled spectrogram. After fixed-length processing, we input such data into a residual convolutional neural network (ResNet), and input age, height, weight, and other characteristics to Multilayer Perceptron (MLP). Then we connect the output of ResNet and MLP, and assign the result to the fully connected layer for classification. The loss is calculated by binary cross entropy.*

*Our method was applied to the 2022 George B. Moody PhysioNet Challenge. In the test set, the Accuracy score of murmur classification was 0.757, AUROC score was 0.797, AUPRC score was 0.610, and Weighted Accuracy score was 0.671. The outcome Accuracy score was 0.562, AUROC score was 0.624, AUPRC score was 0.631, Weighted Accuracy score was 0.612, and the Cost was 13,394. We were ranked 23th in the murmur classification and 24th in the clinical outcome classification.*

## 1. Introduction

Using artificial intelligence (AI) technology to assist diagnosis and treatment is a very good job, it can save our medical costs, and can screen out people who are not sick before experts make a diagnosis, so that medical resources can be used in the best way. The AI recognition of heart disease diagnosis is such a work, in this year's challenge is to determine whether people have heart disease mainly through heart sound data.

The traditional way of judging a patient based on heart sounds requires identifying whether the heart sounds of S1, systole, S2 and diastole are abnormal, and combining other judgments to determine whether the patient has heart disease, and then determining the treatment plan. However, using this method requires professional medical personnel and needs to accurately segment the heart sound data into four stages of heart sounds, and the segmentation of the segments may also cause errors due to the influence of noise in the data.

Our team, JUST_IT_Academy1, investigated a novel method that we process the audio data using Residual Convolutional Neural Networks (ResNet) [1] and Multilayer Perceptron (MLP) [2, 3] for other features such as age, gender, height, weight, etc., which are then concatenated for classification. Our method is applied to the 2022 George B. Moody PhysioNet Challenge [4].

Our approach does not need medical expert necessarily involved, and does not require the segmentation of heart sounds, allowing only experts in the field of machine learning to contribute to the algorithm.

## 2. Data Preprocessing

We use 60% heart sound data provided by George B. Moody PhysioNet Challenge 2022 dataset[5-7], which has 942 patients and 3,163 recordings for training and cross-validation. The distribution about the data set is shown in Table 1. One may notice that both the murmur and outcome classification are unbalanced. Heart disease is diagnosed in 83.8% of Present Murmur patients, 63.2% of Unknown Murmur patients and 37.8% of Normal Murmur patients, respectively.

Table 1. Classification distribution of the dataset.

| Classification | Abnormal | Normal | Total Murmur |
|---|---:|---:|---:|
| Present | 150 | 29 | 179 |
| Unknown | 43 | 25 | 68 |
| Absent | 263 | 432 | 695 |
| Total Outcome | 456 | 486 | 942 |

For the WAV audio signal of the heart sound, we convert the heart sound data to a Mel-scaled spectrogram using Mel-spectrographic transformation, resulting in 128-dimensional data.

Since our network input requires fixed-length data, we

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.001

truncate the converted data to a length of 128 bits, and zero-pad data less than 128 bits to obtain 128×128 two-dimensional data. Each patient in this dataset has the feature values collected from AV, MV, PV, TV, and Phc. The number of the patients who have all feature values takes 0.3% of the total, which is a very less common case. Therefore, we assign the value of 0 for patients' empty features, and finally construct 128×128×5 three-dimensional data.

Other features include age, gender, height, weight, pregnancy, and the mean, variance, and skew values of AV, MV, PV, TV, and Phc. One-dimensional data of length 26 are obtained using the standard method provided by the official.

## 3. Methods

We design a method that uses Residual Convolutional Neural Network (ResNet) [1] to learn the characteristics of WAV data, and design a 2-layer Multilayer Perceptron (MLP) [2, 3] to learn the characteristics of patients from all their other features, and finally connect the outputs of the two networks for classification. We compute its loss using binary cross-entropy.

Since the datasets is very small for the needs of deep neural networks [8], instead of using the standard ResNet, we simplify the model and tune some hyperparameters for better performance. In each block of ResNet, we perform 3 convolution operations and perform residual computation on the input data. The detailed descriptions are shown in the following subsections.

### 3.1. Murmur Classifier

The murmur classifier is composed by ResNet and MLP networks. After connecting their outputs, they are input to the full-link layer for murmur classification. This classifier will output the probabilities of *Absent*, *Unknown*, *Present*, and take the max one as its prediction result.

As shown in Figure 1, we use 3 ResNet blocks, each ResNet Block has the same structure as the standard version. Its output is connected with the data processed by the 2-layer MLP through global pooling, and finally classified as *Present*, *Unknown* and *Absent* through the fully connected layer. MLP consists of 1 Dense plus ReLU layer and 1 Dense plus sigmoid layer, and we use dropout to suppress its overfitting.

The hyperparameter settings of 3 blocks for ResNet network and MLP network are shown in Table 2.

Table 2. Hyperparameters for Murmur Classifier and Outcome Classifier.

| ResNet Block | Murmur Classifier | | Outcome Classifier | |
|---|---|---|---|---|
| | Filters | Kernels | Filters | Kernels |
| 1 | 32*32 | 8*8 | 32*32 | 8*8 |
| 2 | 64*64 | 5*5 | 64*64 | 5*5 |
| 3 | 64*64 | 3*3 | 64*64 | 3*3 |
| MLP Dense | Units | | Units | |
| 1 | 54 | | 54 | |
| 2 | 4 | | 4 | |
| FC | 3 | | 2 | |

### 3.2. Outcome Classifier

As shown in Figure 2, the outcome classifier also uses simplified ResNet to process the heart sound audio features, and use MLP network to process other features. The structures of ResNet and MLP are the same as Murmur Classifier. We finally concatenate the outputs of ResNet, MLP, and the probabilities which predicted by the previously trained murmur classifier, and input them into the fully connected layer to classify as *Abnormal* and *Normal*. The hyperparameters of outcome classifier are shown in Table 2.

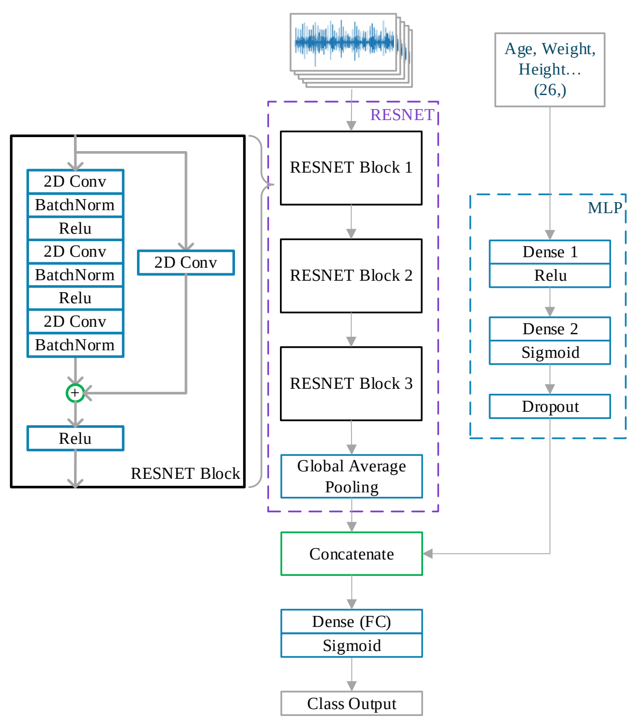


Figure 1. The composition about murmur classifier.

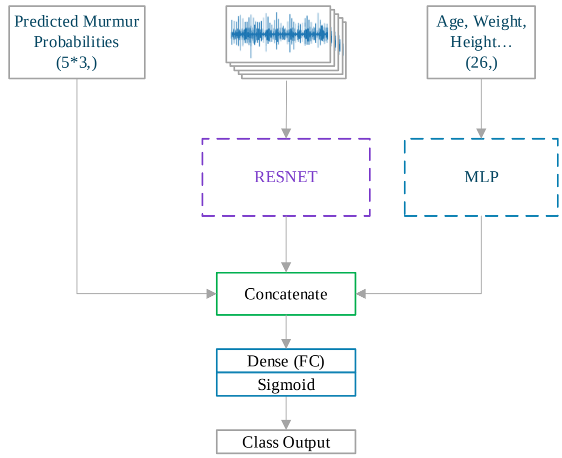


Figure 2. The outcome classifier is concatenate by ResNet blocks and MLP and the predicted values of 5 trained murmur classifiers, then do 2 classifications.

## 3.3. Final Decision Rule

We use the *k*-fold method [9] to split the dataset into 5 folders, and then we train the model 5 times to get 5 sets of parameters. At each training time, we take 4 of these folders as the training set and the remaining folder as the validation set, choosing the parameters that maximize AUPRC in validation. The final classification is chosen by voting.

We design two different voting selectors for final murmur and outcome classifications to reduce the probability of genuine patients missing treatment.

The process about murmur voting selector is shown in Figure 3. If all five predictions of the model are *Absent*, then its classification is *Absent*. Otherwise, if one is predicted to be *Present*, then the result is *Present*. And the rest are treated as *Unknown*.

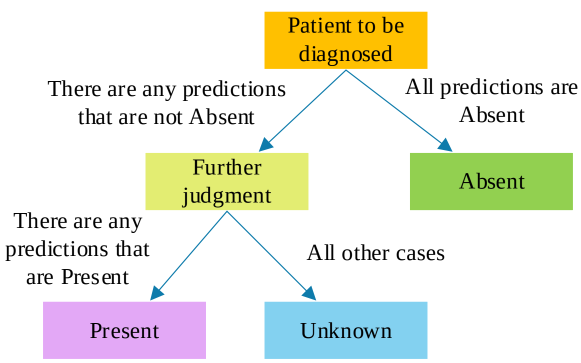


Figure 3. Murmur voting selector.

The outcome voting selector is based on the predictions of the five different sets of parameters of the outcome model. If two of the five outcome classifiers predict the result as *Abnormal*, then the result is *Abnormal*, and the others are *Normal* (see Figure 4).

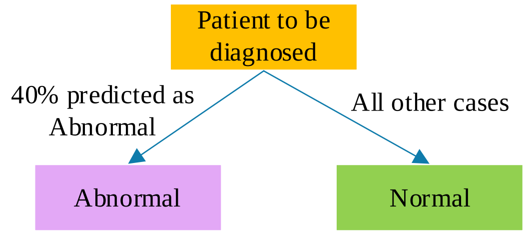


Figure 4. Outcome voting selector.

## 4. Results

The scores obtained by our murmur classifier in the validation set are shown as: Accuracy score was 0.799, AUROC score was 0.839, AUPRC score was 0.667, and Weighted Accuracy score was 0.723. The outcome classification for Accuracy score was 0.537, AUROC score was 0.604, AUPRC score was 0.638, and Weighted Accuracy score was 0.573 (higher is better), Cost was 10,692 (lower is better).

In the test set, murmur's scores were shown as: Accuracy score was 0.757, AUROC score was 0.797, AUPRC score was 0.610, and Weighted Accuracy score was 0.671. The outcome score for Accuracy was 0.562, AUROC score was 0.624, AUPRC score was 0.631, Weighted Accuracy score was 0.612, and Cost was 13,394. We achieved the 23th in the murmur classification ranking and the 24th in the clinical outcome classification ranking.

The Weighted Accuracy of our method in the murmur classification is shown in Table 3, and the Cost in the clinical outcome classification is shown in Table 4.

Table 3 Weighted accuracy metric scores (official Challenge score) for our final selected entry (team JUST_IT_Academy1) for the murmur detection task, 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.

| Training | Validation | Test | Ranking |
|---|---|---|---|
| 0.941±0.032 | 0.723 | 0.671 | 23/40 |

Table 4 Cost metric scores (official Challenge score) for our final selected entry (team JUST_IT_Academy1) for the clinical outcome identification task, 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.

| Training | Validation | Test | Ranking |
|---|---|---|---|
| 10,034±933 | 10,692 | 13,394 | 24/39 |

## 5. Discussion and Conclusion

The *k*-fold method has a great contribution to the stability of the result prediction. As shown in Table 5, when we utilize the *k*-fold method, the murmur classification in the official competition improves by 14.26%, and the outcome classification improves by 19.16%. And when we optimize our voting selector based on the principle of sending people who may be sick as much as possible to experts for confirmation, the murmur classification and the outcome classification in the official competition is improved by 14.22% and 21.15%, respectively.

Table 5. After adding *k*-fold and the optimized voting selector, the prototype network gets about 14% improvement in murmur (higher is better) and about 20% improvement in outcome (lower is better).

| Model | Murmur weighted accuracy | Outcome cost |
|---|---|---|
| Add *k*-fold with optimized voting selector | **0.723** | **10,692** |
| Add *k*-fold only | 0.633 | 13,559 |
| Original | 0.554 | 16,772 |

ResNet is a deep neural network, the more data, the better the accuracy. In this dataset, we design a simplified version of ResNet combined with MLP to obtain better results comparing with the standard ResNet (Table 6). The number of our network parameters is 0.73 million, and the number for the smallest ResNet18 parameters exceed to 11.5 million. Although our proposed model uses fewer parameters, it outperforms the standard version of ResNet in all metrics on the competition dataset.

Table 6. Comparison of our method with standard ResNet in murmur classification (higher is better, cross-validation on training data).

| Model | Accuracy | AUROC | AUPRC |
|---|---|---|---|
| Our Method | **0.872** | **0.923** | **0.874** |
| ResNet18 | 0.844 | 0.894 | 0.811 |
| ResNet34 | 0.859 | 0.898 | 0.821 |
| ResNet50 | 0.853 | 0.897 | 0.819 |
| ResNet101 | 0.831 | 0.868 | 0.770 |

Our method does not require data segmentation, we still need to perform preprocessing about audio information, such as Mel-spectral transformation, energy density transformation, etc. It would be an even bigger advance if we could have a straightforward algorithm for audio feature recognition like CNN [10] does for image feature recognition without requiring image processing experts.

## References


[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[2] C. M. Bishop, “Neural Networks for Pattern Recognition. Repr,” Agricultural Engineering International the Cigr Journal of Scientific Research & Development Manuscript Pm, vol. 12, no. 5, pp. 1235 - 1242, 1995.

[3] F. Rosenblatt, “Perceptrons and the Theory of Brain Mechanisms,” Cornell Aeronautical Laboratory Report No, 1960.

[4] M. A. Reyna, Y. Kiarashi, A. Elola, J. Oliveira, F. Renna, A. Gu, E. A. Perez-Alday, N. Sadr, A. Sharma, S. Mattos, M. T. Coimbra, R. Sameni, A. B. Rad, and G. D. Clifford, “Heart Murmur Detection from Phonocardiogram Recordings: The George B. Moody PhysioNet Challenge 2022,” medRxiv, 2022, doi:10.1101/2022.08.11.22278688.

[5] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. J. c. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” Circulation [Online], vol. 101, no. 23, pp. e215-e220, 2000.

[6] J. Oliveira, F. Renna, P. D. Costa, M. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares, A. Elola, A. B. Rad, R. Sameni, G. D. Clifford, and M. T. Coimbra, “The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification,” IEEE J Biomed Health Inform, vol. 26, no. 6, pp. 2524-2535, Jun, 2022, doi:10.1109/JBHI.2021.3137048.

[7] J. Oliveira, F. Renna, P. D. Costa, M. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares, A. Elola, A. B. Rad, R. Sameni, G. D. Clifford, and M. T. Coimbra, “The CirCor DigiScope Phonocardiogram Dataset (version 1.0.3),” PhysioNet, 2022, doi:10.13026/tshs-mw03.

[8] Y. Lecun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[9] T. Fushiki, “Estimation of Prediction Error by Using K-fold Cross-validation,” Statistics and Computing, vol. 21, no. 2, pp. 137-146, 2009, doi:10.1007/s11222-009-9153-8.

[10] A. Ajit, K. Acharya, and A. Samanta, “A Review of Convolutional Neural Networks,” in 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-5.



Address for correspondence:

Guohui Peng
School of Computer
Jiangsu University of Science and Technology
China
pguohui2004@163.com