Transformer embedded with learnable filters for heart murmur detection

Pengfei Fan[1], Yucheng Shu[1], Yiming Han[1]

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

Abstract

Aim: Heart murmur detection can provide a preliminary diagnosis of heart disease, and has become increasingly important in assisting clinical diagnosis and treatment in recent years. The purpose of this study is to construct an automatic detection system for heart murmurs.

Methods: We build a learnable filter-based transformer architecture. The learnable filter is embedded between the embedding layer and the encoder layer of the transformer. The parameters of the filter are optimized by Adam to adaptively represent any filter in the frequency domain, thereby achieving the effect of adaptive noise reduction. Then, the transformer encoder module captures the long-term dependencies of the heart sound signal, allowing the network to learn more effective features from the input signal. Finally, the final classification result will be obtained according to the voting rules we set.

Results:Our(Bear_FH) method is trained and validated on a public dataset proposed by the challenge. In the formal phase of the challenge, the trained algorithm was tested using a hidden validation set, and we obtained challenge metric scores (weight accuracy and cost) of 0.367 and 19163, respectively.

1. Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide [1]. Currently, 17.7 million people worldwide die from cardiovascular disease every year. The study of heart sound signals has very important clinical value for the early diagnosis of cardiovascular disease. At present, the automatic analysis of heart sounds based on biological signal processing and artificial intelligence technology is becoming a popular research direction.

In the past few years, a large number of automatic heart sound classification algorithms based on machine learning [2–4] and deep learning [5, 6] have been proposed. Among them, deep learning methods usually use short-time Fourier transform and wavelet transform [7] to extract the effective features of heart sound signals, and subsequent classification models often use CNN and RNN [8]. At the same time, Transformer, as an attention-based method, has an excellent performance in time series forecasting, analysis of medical physiological signals [9], etc.

The goal of The 2022 PhysioNet/CinC challenge is to identify the presence of murmurs, as well as to detect the clinical outcomes from heart sound recordings collected from multiple auscultation locations on the body using a digital stethoscope [10]. In this paper, we adopt the transformer as the basic framework to classify clinical heart sound recordings according to their corresponding temporal features, and our method is described in detail below.

2. Methods

In this section, we first briefly describe the basic information of the dataset and then introduce the implementation details about our method. The process design of the entire framework is shown in Figure 1. The first is the preprocessing stage, which is divided into 3-second segments by sliding the heart sound recording between windows and normalized. Then, the deep neural network proposed in this paper is used to extract the features of heart sound segments. Finally, the final patient-level results are obtained through a decision-making strategy based on the majority voting rule.

2.1. Datasets

We will conduct experiments on the 2022 PhysioNet/CinC challenge dataset[11]. For clarity, we summarize the basic information of this dataset in Figure 2. The 2022 PhysioNet/CinC challenge data was primarily collected from the pediatric population (21 years or younger). The dataset contains a total of 1568 patients with one or more heart sound recordings (5192 heart sound recordings in total), of which 3162 records from 942 patients are publicly shared as training set, and 2030 records from 626 patients are used as validation set and test set set for private reservation.

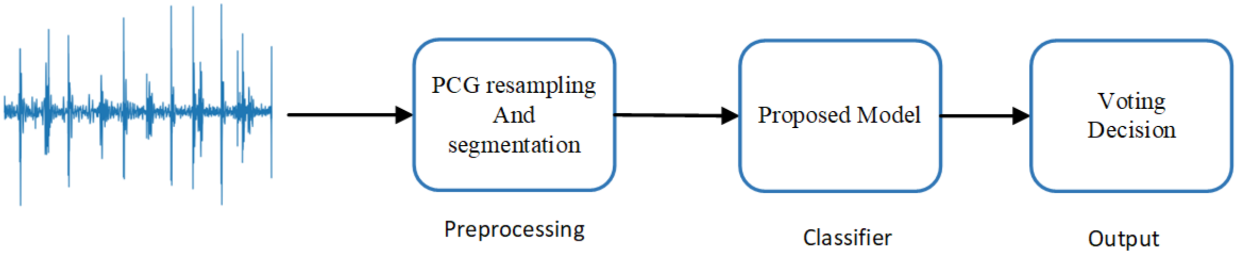

Figure 1. The flowchart of the proposed method.

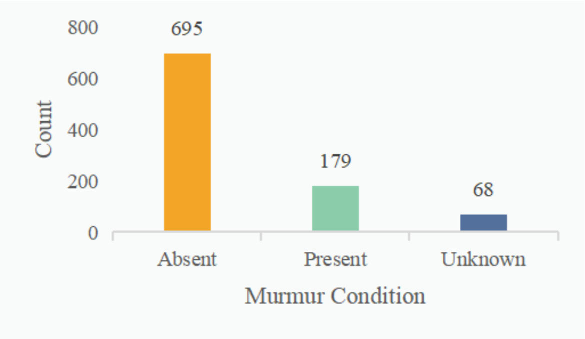

Figure 2. The PCG samples category distribution chart.

2.2. Data Pre-processing

When data is collected, each PCG record is naturally different due to differences in individuals and collection environments. In order to make the model training have better results, we adopted the following data preprocessing techniques. First, all heart sound recordings were resampled to a frequency of 2000 Hz. Second, normalize each heart sound recording to be between -1 and 1. Third, in order to ensure that the length of the data input to the deep learning model is fixed, each heart sound recording is cut into 3-second segments through a sliding window. Note that a 3-second heart sound recording is longer than a full cardiac cycle. Fourth, after data statistics, it is found that there is a serious data imbalance problem, so we use two data enhancement methods: random shift signal and random addition of Gaussian noise to the signal.

2.3. Model overview

Our learnable filter-based transformer model consists of three main components: 1) A basic CNN for learning shared low-level features. 2) The learnable filtering layer uses learnable filters to adaptively reduce noise information in the frequency domain and capture periodic features. 3) The transformer encoder layer captures the global dependencies of the heart sound signal. Figure 3 shows the overall framework of our proposed network.

2.3.1. Convolution layer

We mainly use multi-layer convolution to replace the embedding layer in the transformer. The original PCG waveform can obtain the shallow features of the heart sound signal through a series of convolution operations. At the same time, the original waveform is down-sampled by a factor of about 30, which reduces the number of model parameters. The obtained shallow features are then summed together with the positional encoding, and the obtained result is finally fed into the learnable filtering layer.

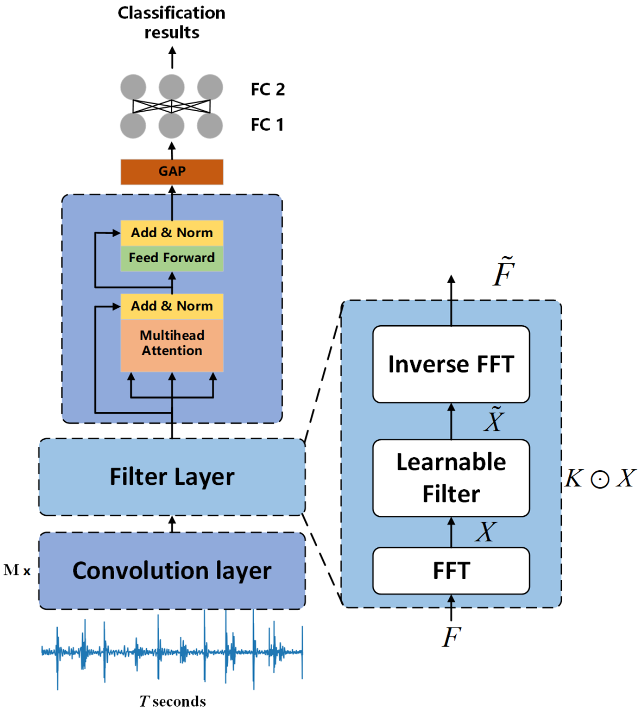

Figure 3. Illustration of our deep neural network.

2.3.2. The learnable filtering layer

In this module, we mainly realize the adaptive reduction of data frequency domain noise information and the capture of data periodic characteristics. First, we perform a Fast Fourier Transform (FFT) along the item dimension to transform the input information from the time domain to the frequency domain:

$$\text{X} = FFT(F) \in C^{n\times d}, \tag{1}$$

where $FFT(\cdot)$ denotes the one-dimensioal FFT. X is a complex tensor and represents the spectrum of F,and n,d denotes the length of feature sequence and the dimension of feature embedding,respectively. Then, we can multiply the obtained spectrum by the learnable filter K to achieve the effect of modulating the spectrum.

$$\tilde{X} = K \odot X, \tag{2}$$

$$\tilde{F} \leftarrow FFT^{-1}(\tilde{X}) \in R^{n\times d}, \tag{3}$$

where $\odot$ is the element-wise multiplication. Note that X is a complex tensor and represents the spectrum of F. Then, we can multiply the obtained spectrum by the learnable filter K to achieve the effect of modulating the spectrum. The learnable filter K can be optimized by the Adam optimizer to adaptively represent any filter in the frequency domain. At the same time, after the proof in [12], it can be known that the multiplication in the frequency domain is equivalent to the circular convolution in the time domain, and has a larger receptive field in the whole sequence, so it can better capture the periodic characteristics of the heart sound signal. Finally, we transform the modulation spectrum X back to the time domain by an Inverse Fast Fourier Transform (IFFT) and update the sequence representation. After this series of operations, the noise in the recorded data can be effectively reduced, resulting in a clearer feature representation.

2.3.3. Encoder

Since our current task is a classification task, not a time series prediction task, we only need to use the encoder module in the Transformer model. Each encoder layer consists of a multi-head self-attention mechanism sub-layer, followed by a fully connected feed-forward network. As described in [13], we use skip connections and layer normalization in each sublayer. The long-term dependencies of the heart sound signal are captured by the transformer encoder module, which enables the network to learn more effective features from the input signal.

2.4. Voting decision strategy

According to the scoring rules of the 2022 PhysioNet/CinC challenge, we know that the model's final prediction results are for patients. However, in the data preprocessing stage, this paper divides each heart sound recording of the patient into several segments. Therefore, how to aggregate the diagnostic results of the segmented fragments into the final diagnosis of the patient is crucial. Our aggregation strategy is based on the majority voting rule. First we aggregate the segment-level to record-level (summarize predictions from multiple segments of a heart sound recording into predictions for this heart sound recording). Then, the prediction results of one record or multiple records of the patient are aggregated into the prediction results of the patient.During the aggregation period, we may experience some tie-breakers. In response to these situations, we set priorities according to the official calculation cost flow chart. Priority for three categories: Present>Unkown>Absent.

2.5. Training Setup

Our learnable filter-based transformer model was trained for 60 epochs with a batch size of 64 on a machine with 64 GB RAM, 8-core CPU and one NVIDIA GeForce RTX 2080 Ti GPU.The model parameters were optimized with the Adam op-timizer [10]. The learning rate during training is set to 0.001 and rescheduled to 0.0001 at the 30th epoch and 0.00001 at the 50th epoch.

2.6. Reuslts

We train and evaluate the model using 10-fold cross-validation, where 10-fold is used for model training and the rest is used for model testing. Repeat ten times to generate ten trained models. Then select the best model from the 10 models and upload it for testing on the official test set. In the end, we obtained challenge metric scores (weight accuracy and cost) of 0.367 and 19163, respectively.

Weighted accuracy	Costs
0.367	19163

Table 1. 10-fold cross-validation results showing challenge scoring for our model.

3. Conclusions

In this paper, we propose a learnable filter-based transformer architecture. The model utilizes learnable filters to adaptively reduce noise information in the frequency domain and can better capture periodic features. At the same time, the long-term dependence of the heart sound signal is captured by the transformer encoder module. However, during the experiment, it was found that our method has potential over-fitting risks and does not take full advantage of the unique advantages of the transformer. We will find more detailed solutions in future work.

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] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. The Lancet 2018;392(10159):1789–1858.

[2] 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.

[3] 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.

[4] Uğuz H. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. Journal of medical systems 2012; 36(1):61–72. ISSN 1573-689X.

[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.

[6] Dominguez-Morales JP, Jimenez-Fernandez AF, Dominguez-Morales MJ, Jimenez-Moreno G. Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. IEEE transactions on biomedical circuits and systems 2017; 12(1):24–34. ISSN 1932-4545.

[7] Zeng W, Yuan J, Yuan C, Wang Q, Liu F, Wang Y. A new approach for the detection of abnormal heart sound signals using tqwt, vmd and neural networks. Artificial Intelligence Review 2021;54(3):1613–1647.

[8] Zhang W, Han J, Deng S. Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation. Biomedical Signal Processing and Control 2019;53:101560.

[9] Ahmedt-Aristizabal D, Armin MA, Denman S, Fookes C, Petersson L. Attention networks for multi-task signal analysis. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020; 184–187.

[10] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, et al. Heart murmur detection from phonocardiogram recordings: The george b. moody physionet challenge 2022. medRxiv 2022;.

[11] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C, et al. The circor digiscope dataset: from murmur detection to murmur classification. IEEE journal of biomedical and health informatics 2021;26(6):2524–2535.

[12] Zhou K, Yu H, Zhao WX, Wen JR. Filter-enhanced mlp is all you need for sequential recommendation. In Proceedings of the ACM Web Conference 2022. 2022; 2388–2399.

[13] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems 2017;30.

Address for correspondence:

Yucheng Shu
No.2, Chongwen Road, Nan'an district, Chongqing, China
shuyc@cqupt.edu.cn

Yiming Han
No.2, Chongwen Road, Nan'an district, Chongqing, China
s210231061@stu.cqupt.edu.cn