Learning Time-Frequency Representations of Phonocardiogram for Murmur Detection

Jae-Man Shin[1], Hyun-Seok Kim[2], Woo-Young Seo[2], Sung-Hoon Kim[1]

[1]Department of Anesthesiology and Pain Medicine, Asan Medical Center, Seoul, Korea
[2]Biomedical Engineering Research Center, Asan Medical Center, Seoul, Korea

Abstract

In this Physionet2022 challenge, our goal is to diagnose whether a patient has heart murmur or cardiac diseases. In this study, we propose new approaches to diagnose cardiac abnormality or murmur. Proposed deep learning models for detecting heart murmur were based on EEGNet and temporal convolutional networks to employ learning frequency-temporal-specific representation from phonocardiogram. To learn patient-specific representation of phonocardiogram, we also utilized demographic information: age, sex, BMI, pregnancy status. Demographic features concatenated to convolutional feature vector then predict murmur presence. Our convolutional networks predict per segmentation for recorded heart sounds. Then we utilized periodical property of heart murmur. We utilized how often murmur characteristics in systole or diastole intervals appeared from phonocardiogram recording. From view of frequentist inference, we extracted statistical features and employed to train machine learning models, RandomForest, to diagnose a patient's condition of heart sounds. In official phase, our team(amc-sh) recorded 0.689 weighted accuracy in murmur detection, 9203 challenge cost for outcome detection based on the highest scores.

1. Introduction

Phonocardiogram (PCG) is sounds recorded activity of the heart when it is beating, it is an important sign when patients are diagnosed for heart diseases. In heart sounds, there exist two sounds, first sound(S1) and second sound(S2), generated from valves activity of the heart. Systole interval is defined by interval between S1 and S2, and diastole interval is between S2 and S1. In general, heart sounds waveform is a collection of cycles of S1-Systole-S2-Diastole without considering noises.

Heart murmurs are noises generated from valves diseases including aortic stenosis, mitral regurgitation, etc. Heart murmurs usually appear in systole or diastole. Because heart murmurs can be clue for heart diseases, detecting murmurs is essential problem aspect clinical applications of deep learning.

In this study, our goal is to detect heart murmurs on pediatric dataset using 1-D convolutional neural networks. The main approach in this paper is to determine the overall tendency of the heart sound signals after short-time detection for systole and diastole. Heart murmur has the property of periodical appearance in systole or diastole interval. we noted how often the murmur sounds appear in the overall sounds. The final goal in this study is to diagnose the patient's cardiac diseases by using the frequency of occurrence of murmur signal using 1-D convolutional networks and machine learning models.

2. Methods

In this paper, we present the architecture and processes to detect murmur existence. The Proposed method consists of two phases, murmur detection model and diagnosis model. Our workflow was described at Figure 1.

On the first phase, we trained deep learning models to detect murmur presence of pre-processed and segmented signal. If a recoding is murmur sounds, then we labelled all segmented signals as murmur presence. Our model is based on 1D convolutional neural networks. In the first phase, CNN models only predict whether each segmented signal is murmur present or absent except unknown class.

At the next phase, our models determined whether an arbitrary patient has heart murmur or abnormality. Because heart murmur appears periodically during systole or diastole interval, most segmentations from a murmur recording should have murmur characteristics. Once CNN models predict most segmentations as murmur presence, then we classified the recording and patient as murmur presence. As a criterion of classification, we extracted statistic from outputs of CNN models for segmentations. We worked above processes per auscultation locations. Then we employed to learn decision rules using machine learning models.

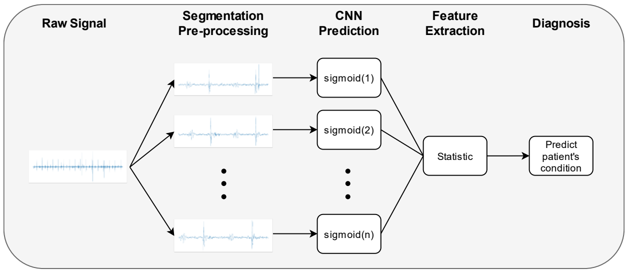

Figure 1. Workflow

We worked each phase after splitting training dataset. Frist split training data is used for training deep learning models, not including unknown class. Second split dataset is used for training diagnosis models, including unknown class.

For abnormality of heart sound, we attempted to predict abnormal signal using murmur detection models.

2.1. Dataset

Public dataset in this challenge is pediatrics dataset recording heart sounds and collected from four cardiac auscultation: Aortic valve, Pulmonary valve, Tricuspid valve, Mitral valve [1]. Training dataset includes recordings of heart sounds, segmentation labels, demographic information and murmur characteristics like shape, pitch, etc. There are two goals in this challenge, prediction for murmur presence and abnormality. Each recording or patient was labelled by expert. Every heart sound was sampled by 4000hz. When deep learning model was trained, we used only parts that the segmentation labels are non-zero. When diagnosis model was trained, entire wave was used.

2.2. Pre-processing

Each heart sound signal was pre-processed as followings.

• Segmentation: Each raw signal was segmented by 1 second and overlapped by 0.25 seconds.

• Spike removal: Schmidt spike removal is used, introduced by [2].

• Resampling: After spike removal, every signal was resampled to 2000hz.

• Normalization: Every segmented signal is normalized by $hs(i) = \frac{hs(i)}{\max(|hs(0)|,\ \cdots,\ |hs(n)|)}$, where $hs(n)$ is a segmented heart sound with length n.

For demographic features, each feature was normalized by maximum value. We replaced missing value to mean value for weight and height, and mode for age. Then we calculated body mass index (BMI) from height and weight, defined by $\frac{weight(kg)}{height^2(m)}$. Upper bound of BMI was restricted by 30. We used four demographic features: age, sex, BMI, pregnancy status.

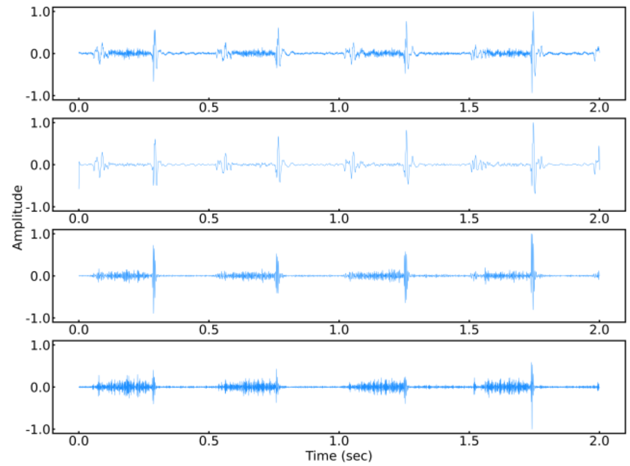

Figure 2. Holosystolic murmur with grade 2. First: raw signal. Second: band-pass filtered signal (20~200hz). Third: band-pass filtered signal (200~400hz). Fourth: band-pass filtered signal (400~600hz).

2.3. Models

Proposed models in this study were based on learning characteristics of heart murmur aspect frequency range. Typically, heart sounds, including first and second heart sound, are audible in frequency range from 20 to 200hz [3]. But some abnormal signal due to cardiac diseases can appear on other frequency range [4]. Figure 2 described the change of murmur signal according to the frequency range.

Because capturing frequency range for abnormal signal can help to learn murmur representation, proposed model was constructed to learn frequency-temporal features. Figure 3 described architecture of our models.

From above motivations, proposed models were based on two model, EEGNet and Temporal Convolutional Networks (TCN).

EEGNet is introduced by [5] to learn frequency-specific spatial filters for EEG signal. EEGNet consist of three parts: 2D convolution networks, Depthwise Convolution, Separable convolution. EEGNet learns specific frequency range through 2D convolutional layers and Depthwise Convolutional layers. Then separable convolution learns relationship and combines from different feature maps. In this study, EEGNet employ to learn murmur-specific band-pass filtering.

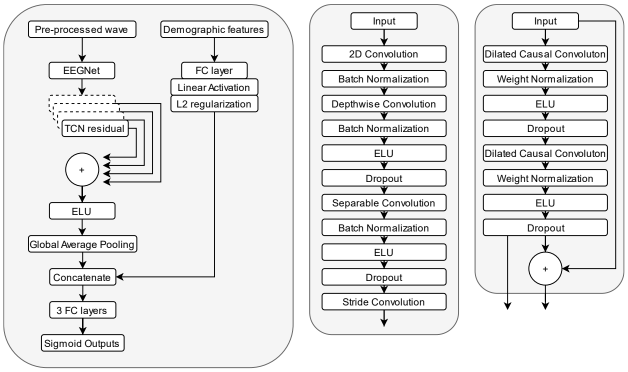

Figure 3. Model architecture. Left: Overall architecture. Middle: EEGNet. Right: TCN residual block

After feature extraction through EEGNet, we expanded our model using temporal convolutional networks (TCN) [6, 7]. TCN architecture was built by stacking residual networks with dilated causal convolution. Dilated causal convolution learn temporal representation of input data with large receptive field.

Our combined EEG-TCN models referred to [8]. We added stride convolution between EEGNet and TCN blocks. Also, every activation function was exponential linear units [9]. We used batch normalization in EEGNet block and weight normalization [10] in TCN residual block. Then we used skip connections throughout each residual block [7]. Classifiers consist of 3 Fully connected layers after 1D global average pooling layers. Our loss function was focal loss, proposed by [11]. It is expanded binary cross entropy loss and its formula is as followings.

$$Focal\ Loss = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

In this study, we set the parameters of loss function as $\alpha = 0.25,\ \gamma = 2$.

We additionally utilized demographic information to employ learning patient-specific heart sound signals. Pre-processed demographic features were trained by 1 fully connected layer with linear activation and L2 regularization. Then demographic features were concatenated with feature vectors after global average pooling.

Each TCN residual block had 128 filters, 3 kernel size. We stacked 8 residual block and set the dilation rate as (1, 2, 4, 8, 16, 32, 64, 128), exponentially. In EEGNet block, first 2D convolution had 32 filters and (1, 64) kernel size. We set 64 filters and (1, 16) kernel size in separable convolution. And stride convolution had 64 filters and 7 kernel size and 2 strides size. Every dropout rate was set as zero.

As our diagnosis models, Random Forest was used with grid search using stratified 5-fold cross-validation. We searched only estimators and max depth parameters. Outcome diagnosis model and murmur diagnosis model were made, respectively.

For training diagnosis model, features were extracted from sigmoid outputs of EEG-TCN models. We utilized statistic from distribution of EEG-TCN classifier's outputs for all segmentations.

- mean: $\frac{\sum_{k=1}^{n} S(k)}{n}$, where $S(k)$ is sigmoid outputs of EEG-TCN models and n is the number of segmentations get from an arbitrary PCG recording.
- ratio: percentage of segmentation detected to be murmur presence by EEG-TCN models for all segmentation, defined by $\frac{\left|\{S(k) \middle| S(K) < threshold\}\right|}{n}$, where n is the number of segmentations get from an arbitrary PCG recording and threshold value was optimized by step size 0.05 in interval (0, 1).
- kurtosis: kurtosis of sigmoid outputs of EEG-TCN models, defined by $\frac{E(S-\mu)^4}{\sigma^4} - 3$ [12] where S is the sigmoid outputs and μ is the mean of S and σ is the standard deviation of S.
- skewness: skewness of sigmoid outputs of EEG-TCN models, defined by $\frac{E(S-\mu)^3}{\sigma^3}$ [12].

Our assumption for above features extraction is that models will yield skewed prediction for murmur presence or absence from an overall signal. We also expected that unknown class has different distribution than present case or absent case (for example, normal distribution).

Our feature extraction was processed for each auscultation locations (Aortic valve, Pulmonary valve, Tricuspid valve, Mitral valve), respectively. We totally extracted 16 features ($4\ locations \times 4\ features$) from one patient. If there were no auscultation locations, we filled zero values.

For outcome diagnosis model, we utilized same features.

3. Results

In official phase, our state-of-the-art scores on public validation set are 0.689(9th submission) in murmur score and 9203(3rd submission) in outcome score. The 5-fold cross-validation results on training set were described in Table 1~3. Each table was filled averaged score on 5-fold results. We work experiment using 9th submission code. In 9th submission code, our methods achieved 11991 challenge scores for outcome detection.

Table 1. 5-fold cross-validation results on training set.

Metrics	Murmur	Outcome
Weighted Accuracy	0.727	0.537
Challenge Cost	16956	15122
AUROC	0.863	0.666
AUPRC	0.678	0.668
F-measure	0.663	0.631

Table 2. 5-fold cross-validation results on training set. Results per class for murmur detection.

Metrics	Present	Unknown	Absent
AUROC	0.890	0.825	0.874
AUPRC	0.812	0.291	0.931
F-measure	0.767	0.330	0.891
Accuracy	0.662	0.408	0.908

Table 3. 5-fold cross-validation results on training set. Results per class for outcome detection.

Metrics	Abnormal	Normal
AUROC	0.666	0.666
AUPRC	0.701	0.635
F-measure	0.564	0.699
Accuracy	0.481	0.806

For murmur challenge scores, results on public validation set are 0.038 less than results on training set.

4. Discussion

There were differences between 3rd submission and 9th submission. The followings are applied in 3rd submission.

- Resampling: Resampling was excluded in pre-processing steps.
- Model architecture: Two residual block was added with (256, 512) dilation rate. And EEG-TCN models was trained without demographic information.
- Diagnosis models: Models was trained without grid search.

Our further study is to work ablation study. The followings are our further study.

- Robustness: We work our study for others public dataset. Using others dataset, our models will have robustness or will be extended to pre-trained models.
- Ablation study: We study how patient's demographic information affects learning the PCG signal. We also study how the heart murmur is detected by our models per four auscultation locations. We also study the performance of our models depending on murmur characteristic such as shape, pitch, and grade.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), which is funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C2383), and was supported by Institute of Information & communications Technology Planning & Evaluation (IITP). This study was also supported by a grant (2022IP0053) from the Asan Institute for Life Sciences, Asan Medical Center, Seoul, Korea.

References

[1] J. Oliveira *et al.*, "The CirCor DigiScope dataset: from murmur detection to murmur classification," *IEEE journal of biomedical and health informatics,* vol. 26, no. 6, pp. 2524-2535, 2021.

[2] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiological measurement,* vol. 31, no. 4, p. 513, 2010.

[3] S. Choi and Z. Jiang, "Comparison of envelope extraction algorithms for cardiac sound signal segmentation," *Expert Systems with Applications,* vol. 34, no. 2, pp. 1056-1069, 2008.

[4] S. Choi, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine," *Expert Systems with Applications,* vol. 35, no. 4, pp. 1679-1687, 2008.

[5] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 15, no. 5, p. 056013, 2018.

[6] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271,* 2018.

[7] A. v. d. Oord *et al.*, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499,* 2016.

[8] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020: IEEE, pp. 2958-2965.

[9] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289,* 2015.

[10] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," *Advances in neural information processing systems,* vol. 29, 2016.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision,* 2017, pp. 2980-2988.

[12] J. Behar, J. Oster, Q. Li, and G. D. Clifford, "ECG signal quality during arrhythmia and its application to false alarm reduction," *IEEE transactions on biomedical engineering,* vol. 60, no. 6, pp. 1660-1666, 2013.

Address for correspondence:

Sung Hoon Kim.
88, Olympic-ro 43-gil, Songpa-gu, Seoul, Republic of Korea.
shkimans@gmail.com