Learning Time-Frequency Representations of Phonocardiogram for Murmur Detection

Jae-Man Shin[1], Seong-Yong Park[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

As part of the George B. Moody PhysioNet Challenge 2022, our team (amc-sh) developed new computational approaches to diagnose cardiac abnormality or heart murmur. The 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. From view of frequentist inference, we extracted statistical features and trained two separate Random Forest models to classify heart murmur and patient outcome. The weighted accuracy of murmur detection in 5-fold CV in public training set was 0.727 and challenge cost for outcome detection was 11,886. Our team (amc-sh) recorded 0.688 (rank: 22/40) weighted accuracy in murmur detection, 13,002 (rank: 20/39) challenge cost for outcome detection as final score.

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.

The George B. Moody PhysioNet Challenge 2022 provide patient heart sound recordings [1] to enable participants designing novel algorithms to detect heart murmur and cardiac abnormality. As part of the Challenge [2], our goal is to detect heart murmurs on pediatric dataset using 1-D convolutional neural networks.

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 is 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.

ISSN: 2325-887X DOI: 10.22489/CinC.2022.126

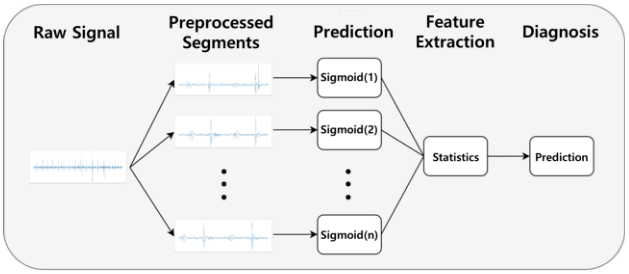

Figure 1. Workflow for heart murmur and outcome detection. Each patient's recordings are segmented and predicted by a trained neural network and the statistic of prediction score distribution is derived. These features are used for patient condition (murmur or outcome status) prediction.

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 [2], 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

The pre-processing includes segmentation, spike removal, resampling, and normalization. Since the typical heart rate of children is in the range of 75-180, we set the length of the segment as 1 seconds. Therefore, each segment contains at least 1 complete heart-beat sounds including S1 and S2, systole and diastole sound. We used 75 % of overlap to increase the number of samples and to allow smooth transition between each segment. We used Schmidt spike removal introduced in [3] to remove unwanted spike signals. We tested various sampling rates and found that 2-time downsampling did not affect performance of the model. Therefore, we used 2 time down sampling from 4,000 Hz to 2,000 Hz to reduce processing time. After segmentation, spike removal and resampling, each segmented signal is normalized by its absolute maximum value. 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), 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.

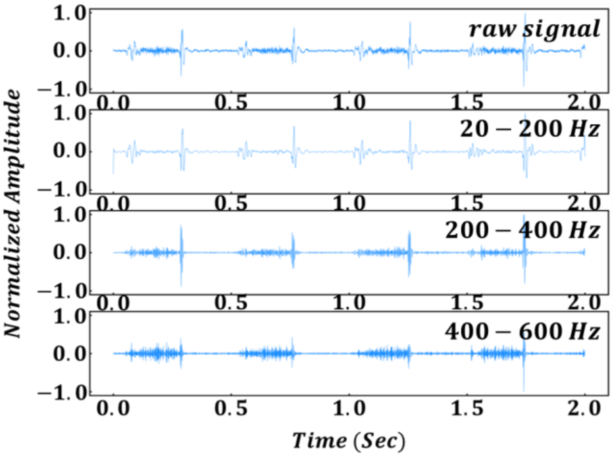

Figure 3. Example of holosystolic murmur with grade 2 in different frequency bands. (Raw signal, 20~200 Hz, 200~400 Hz and 400~600 Hz band-pass filtered signals respectively.

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

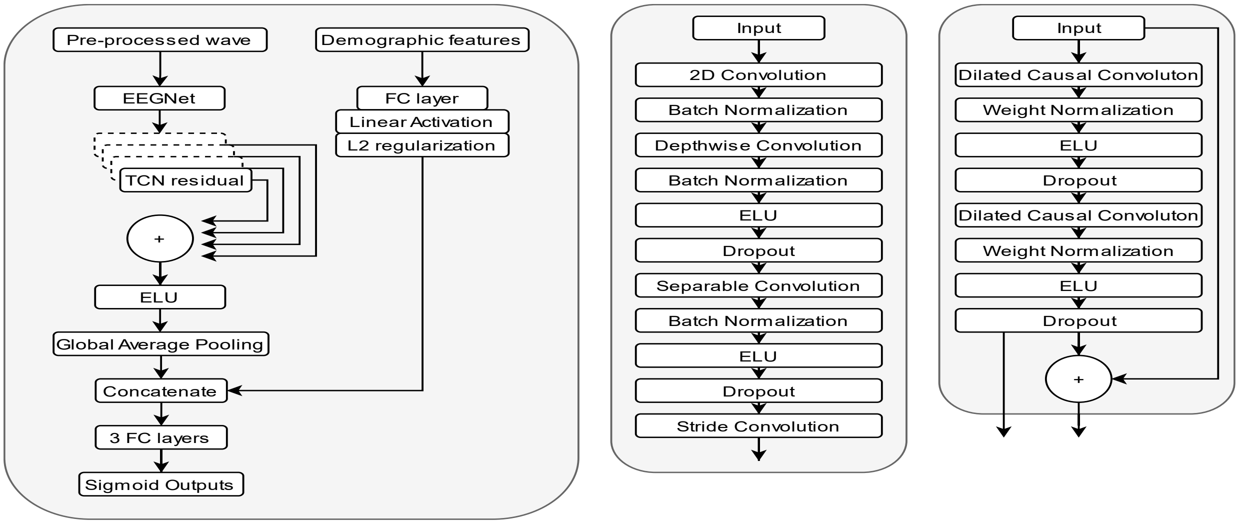

Figure 3. Model architecture. Left: Overall architecture. Middle: EEGNet. Right: TCN residual block

different feature maps. In this study, EEGNet employ to learn murmur-specific band-pass filtering.

After feature extraction through modified 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]. Its formula is as followings.

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

Where α_t is balancing factor for class imbalance and γ is modulating factor to classify between easy samples and hard samples. In this study, we set the parameters of loss function as default values ($\alpha_t = 0.25,\ \gamma = 2$).

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 (murmur detection) or 10 (outcome detection) residual blocks and set the dilation rate as $(1, 2, 4, \ldots, 2^n)$, 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. We selected the number of residual blocks by considering length of heartbeat events.

We used two separate random forest (RF) models as our diagnostic model (scikit-learn). One is for 3 classes of murmur classification, and another is for the outcome classification. We validated diagnostic models with 5-fold CV. Two parameters, maximum depth of tree and number of estimators are optimized by max_depth and n_estimator parameters in *RandomForestClassifier* function in scikit-learn package.

For training diagnosis model, features were extracted from sigmoid outputs of EEG-TCN models. We utilized statistics such as mean, skewness, kurtosis from distribution of EEG-TCN classifier's outputs for all segmentations. We additionally utilized proportion of murmur positive segmentation over total segments.

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 worked same processes.

3. Results

Training	Validation	Test	Ranking
0.727	0.689	0.688	22/40

Table 1. Weighted accuracy metric scores (official

Challenge score) for our final selected entry (team amc-sh) 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
11,886	9,203	13,002	20/39

Table 2. Cost metric scores (official Challenge score) for our final selected entry (team amc-sh) 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.

4. Discussion and Conclusion

For the outcome prediction challenge, our team achieved 11,886 challenge cost in 5-fold CV of public training dataset and 9,203 challenge cost for the validation dataset. Our method showed 13,002 challenge cost for the test dataset and ranked 20/39 for the outcome prediction challenge. We had lowest challenge cost in validation dataset but had highest cost in test dataset. This may be explained by population shift so we investigated this effect by additionally considering demographic features on our EEG-TCN model. When we concatenated demographic features on embedding features of proposed models, the model showed improved murmur classification performance. For each demographic feature (age, BMI, sex, pregnancy status), the improved performance was 0.663, 0.663, 0.527, 0.572, respectively while baseline precision was 0.517. Considering all 4 features at the same time, the improvement was 0.555. Therefore, age and BMI feature were the most informative feature for improving murmur detection capability of our model. Therefore, BMI feature was the most informative feature for improving murmur detection capability of our model. This result indicates that importance of incorporating demographic features to classify murmur or patient outcome to regularize distribution shift due to different cohort. Further investigation on the performance of outcome prediction is absolutely needed and our group is planning to pursue this direction as our future work.

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] M. A. Reyna *et al.*, "Heart Murmur Detection from Phonocardiogram Recordings: The George B. Moody Physionet Challenge 2022," m*edRxiv,* 2022.

[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.

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