Deep Learning Based Heart Murmur Detection Using Frequency-time Domain Features of Heartbeat Sounds

Jungguk Lee[1], Taein Kang[1], Narin Kim[1], Soyul Han[1], Hyejin Won[1], Wuming Gong[2] and Il-Youp Kwak[1]

[1] Chung-Ang University, Seoul, Korea
[2] University of Minnesota, Lillehei heart institute, Mineapolis, MN, United States

Abstract

The goal of the George B. Moody PhysioNet Challenge 2022 was to use heart sound recordings gathered from various auscultation locations to identify murmurs and clinical outcomes. Our team, CAU_UMN, proposes a deep learning-based model that automatically identifies heart murmurs from a phonocardiogram (PCG). We converted the heartbeat sound into 2D features in the frequency-time domain through feature extraction techniques such as log-mel spectrogram, Short Time Fourier Transform (STFT), and Constant Q Transform (CQT). The frequency-temporal 2D features were modeled using voice classification models such as Convolutional neural networks (CNN) and Light CNN (LCNN). The model using log-mel spectrogram and LCNN was ranked 5th for murmur detection with a weighted accuracy of 0.767 and 5th for clinical outcome detection with a cost of 11933 in the test dataset of the George B. Moody PhysioNet Challenge. We believe that our deep learning based heart murmur detection system will be a promising system for automatic heart murmur detection from PCG.

1. Introduction

Congenital heart disease (CHD), which affects about 1% of live births and has significant morbidity and death, is the most prevalent hereditary birth abnormality. For the diagnosis and treatment of congenital cardiac disorders in children, many underdeveloped nations do not have the necessary infrastructure or cardiology specialists. An affordable solution for non-invasive cardiac disease diagnosis is monitoring phonocardiography. A phonocardiography creates a phonocardiogram (PCG), a particular waveform that accurately depicts the heartbeat intensity over time. The tasks for George B. Moody PhysioNet Challenge 2022 are to design systems that detect murmur and clinical outcome events. Two subtasks are based on weighted accuracy and expected cost[1, 2].

The Challenge recording data were collected from a pediatric population in Northeast Brazil in July-August 2014 and June-July 2015 [3]. Each patient in the Challenge data has one or more recordings from 5 or less auscultation locations. The recordings were collected not simultaneously but sequentially from different auscultation locations using a digital stethoscope. Also, each patient has demographic information such as gender, age, and pregnancy status.

Heart Rate Variability (HRV) has been used as the tool for assessing abnormalities in heart disease in prior competitions and numerous medical studies.[4] The ECG's RR interval is a feature that can effectively represent HRV[5, 6], thus we thought of the 'Peaks Interval' (PI), which corresponds to the RR interval, as an extra feature to express HRV in the PCG.

We evaluated LCNN and ResMax models on waveform data to develop an automated murmur event detection system [7, 8]. The CNN-based models are LCNN (Light CNN) and ResMax, and their basic technique, MFM (Max-Feature-Map), is used in both of these models. MFM can not only separate noisy and useful signals but also operate as the feature selection between two feature maps.

In order to develop a robust deep learning model from a limited quantity of data, we have experimented with several augmentation strategies such as cutout, cutmix, and mixup [9–11].

2. Methods

Figure 1 depicts our murmur and clinical outcome detection system architecture. The structure of the murmur classifier was different from the outcome classifier's, and the main distinction between the two classifiers is whether or not demographic information is added. In common, we extracted 2D features and peaks interval features from the raw data. By passing through a simple embedding, the peaks interval feature was concatenated with the model's embedding.

ISSN: 2325-887X DOI: 10.22489/CinC.2022.071

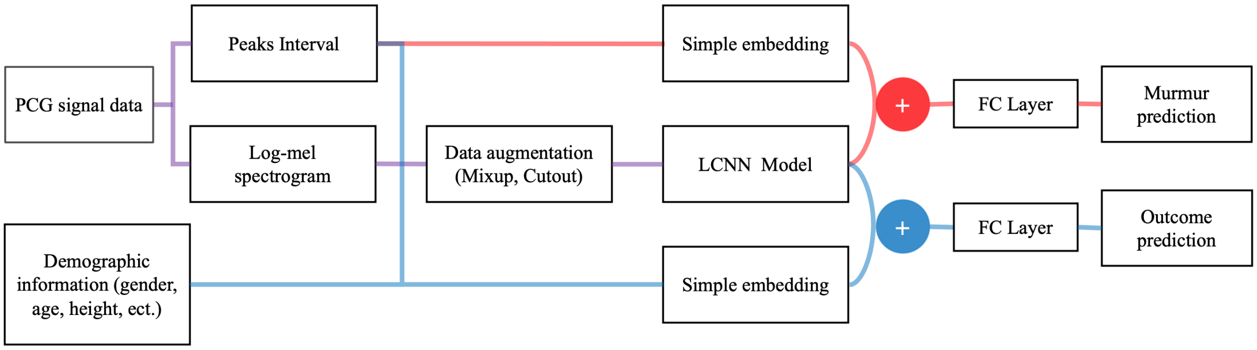
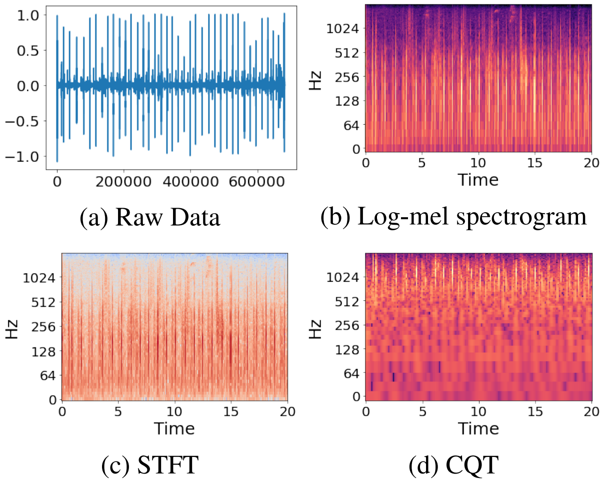


Figure 1: System Architecture

2.1. Feature Extraction

We utilized three methods of converting data, Log-mel spectrogram, STFT, and CQT. Figure 2 visualizes (a) raw, (b) Log-mel spectrogram, (c) STFT, and (d) CQT features of a PCG data from a patient. We also utilized demographic information such as gender, age, height, weight, and pregnancy.



(a) Raw Data (b) Log-mel spectrogram

(c) STFT (d) CQT

Figure 2: Feature engineering

2.1.1. PI feature

Peaks Interval (PI) means the time interval between peak points. The murmur patients have a noise which occurs in the systolic or diastolic phase of the heart[4]. The noise is also a sound, so it generates a wave form. One of the factors of a waveform is that it has a peak point. We thought that if a patient has a murmur, the patient has more peak points than a normal. Having more peak points means that the interval will be shorter. Actually, the average PI interval of normal people was 49% longer than that of murmur patients in the challenge data.

In fact, we wanted to use the value of PI in sequence form. However, due to the noise of the data, PI was not calculated accurately, and thus the Mean of PI was used. This is where further research is necessary.

2.2. Data augmentation

We applied data augmentation to the audio feature to train the model more robustly. Data augmentation improved the generalization performance of the model and prevented overfitting by adding noise to the model trained with a small amount of data. We experimented with various augmentation techniques commonly used in audio data for 2D features (stft, log-mel, cqt, etc.). We implemented augmentation with an online generator, and tried cutmix [10], cutout [9], and mixup [11].

2.3. Models

In this competition, PCG signal data was converted into 2D features like Log-mel spectrogram, STFT, and CQT, and LCNN and ResMax models already been proven in many audio competitions ASVspoof 2017, 2019, and 2021 [8, 12–14] were applied.

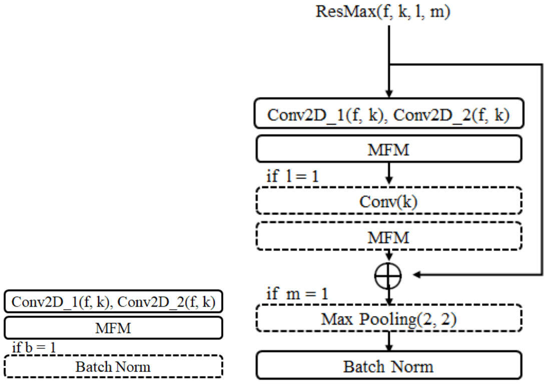
2.3.1. LCNN

Compared with the Light CNN-9 model [7], this paper uses a deeper LCNN model which iterates nine LCNN blocks. The LCNN block consists of convolution, MFM, and an optional batch normalization layer, as shown in Fig. 3 (a) (dotted block applied when $b = 1$). Fig. 4 (a) is the whole LCNN model architecture. Our deeper LCNN model uses 32, 32, 48, 48, 64, 64, 32, 32, and 32 convolution filters. The kernel size of the first convolution layer is set to 5 and the rest of the convolution layers are set to 3 or

1. Global average pooling layer was used instead of fully connected layers.

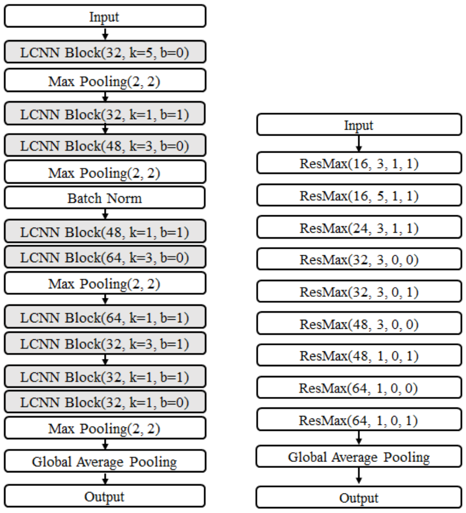
2.3.2. ResMax

ResMax is a model that showed excellent performance in the ASVspoof 2019 competition dataset [8]. The ResMax model consists of four parameters. f is the number of filters, and k is the kernel size. l is an option to apply convolution with kernel size 1 and element-wise maximum to convolution layers (dotted block applied when $l = 1$). m is an option that optionally applies the 2 by 2 MaxPooling. The ResMax block is defined in Fig. 3 (b) and the whole ResMax model composed of 9 ResMax blocks is shown in Fig. 4 (b).



(a) LCNN block (b) ResMax block

Figure 3: Model Blocks



(a) LCNN model (b) ResMax model

Figure 4: Model Architectures

Training	Validation	Test	Ranking
0.828	0.734	0.767	5/40

Table 1: Weighted accuracy metric scores (official Challenge score) for our final selected entry (team CAU_UMN) for the murmur detection task, including the ranking of our team on the hidden test set.

2.4. Model training

In the Challenge data, there is a single or multiple file depending on the stethoscope position for each patient. Each of the files was considered as one sample in our training step. However, in the evaluation process, results had to be derived for each patient. Therefore, we performed the process by combining individual samples. There are some differences depending on the model in the evaluation process. For the murmur detection track, we used the highest probability among the values calculated for each stethoscope position. For the outcome detection track, the probability value calculated for each stethoscope position was averaged.

We divided our training set as 8 to 2 for a separate validation set in our model training. We used cost-sensitive learning because the importance of the murmur class and the outcome class are different in the evaluation metrics.We trained the model by integrating the unknown class into the absence class because the distribution of values in the final model for the unknown class did not converge well in murmur detection. In the evaluation, not detecting the unknown class showed higher performance of the weighted accuracy, so we made the system to detect only by 'absent' or 'present'.

3. Results

The models that showed the best performance through experiments were submitted. We used an LCNN with PI feature for Murmur detection and an LCNN with PI and demographic information for outcome detection.

Tables 1 and 2 summarize our result. Weighted accuracy metric scores for our proposed model were 0.828, 0.734, and 0.767, respectively, on the training, validation, and test sets. Cost metric scores for our proposed model were 8097, 9493, and 11933, respectively, on the training, validation, and test sets.

4. Discussion and Conclusions

The top teams of the challenge also applied many interesting methods. Impressive techniques included two-stage classification (Present vs Unknown and Absent, Unknown vs Absent) and a relabeling method to clarify the Unknown

Training	Validation	Test	Ranking
8097	9493	11933	5/40

Table 2: Cost metric scores (official Challenge score) for our final selected entry (team CAU_UMN) for the clinical outcome identification task, including the ranking of our team on the hidden test set.

class. It was also interesting to build a hierarchical model with multiple mel spectrograms. When modeling using machine learning techniques and demographic data, the performance was usually high on cost metric. However, most of the top teams had the disadvantage of taking a long time to train models.

We have the advantage of having a fast training speed of 1 hour 40 minutes 45 seconds by building a murmur detection system using the LCNN model, a CNN-based deep learning model. Nevertheless, the accuracy is only 1.3% different from that of the winning solution in the murmur detection track. In addition, our proposed model is robust by several augmentation techniques, so the performance difference among the train, validation, and the test set is not significantly high. Our novel spectrogram based deep learning model achieved 0.767 weighted accuracy (5 out of 40 submitted systems) for murmur detection, and achieved a cost of 11933 (5 out of 40 submitted systems) for clinical outcome detection in the official phase of the George B. Moody PhysioNet Challenge.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. NRF-2020R1C1C1A01013020).

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, Physiotoolkit, And Physionet: Components Of A New Research Resource For Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[2] 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;URL https://doi.org/10.1101/2022.08.11.22278688.

[3] 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.

[4] El-Segaier M, Lilja O, Lukkarinen S, Sörnmo L, Sepponen R, Pesonen E. Computer-based Detection And Analysis Of Heart Sound And Murmur. Annals of biomedical engineering 2005;33(7):937–942.

[5] Tsipouras MG, Fotiadis DI, Sideris D. An Arrhythmia Classification System Based On The RR-interval Signal. Artificial intelligence in medicine 2005;33(3):237–250.

[6] Faust O, Shenfield A, Kareem M, San TR, Fujita H, Acharya UR. Automated Detection Of Atrial Fibrillation Using Long Short-term Memory Network With RR Interval Signals. Computers in biology and medicine 2018; 102:327–335.

[7] Wu X, He R, Sun Z, Tan T. A Light CNN For Deep Face Representation With Noisy Labels. IEEE Transactions on Information Forensics and Security 2018;13(11):2884–2896.

[8] Kwak IY, Kwag S, Lee J, Huh JH, Lee CH, Jeon Y, et al. Resmax: Detecting Voice Spoofing Attacks With Residual Network And Max Feature Map. In 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021; 4837–4844.

[9] DeVries T, Taylor GW. Improved Regularization Of Convolutional Neural Networks With Cutout. arXiv preprint arXiv170804552 2017;.

[10] Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y. Cutmix: Regularization Strategy To Train Strong Classifiers With Localizable Features. In Proceedings of the IEEE/CVF international conference on computer vision. 2019; 6023–6032.

[11] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. Mixup: Beyond Empirical Risk Minimization. arXiv preprint arXiv171009412 2017;.

[12] Lavrentyeva G, Novoselov S, Malykh E, Kozlov A, Kudashev O, Shchemelinin V. Audio Replay Attack Detection With Deep Learning Frameworks. In Proc. Interspeech 2017. Stockholm: ISCA, 2017; 82–86.

[13] Lavrentyeva G, Novoselov S, Tseren A, Volkova M, Gorlanov A, Kozlov A. STC Antispoofing Systems For The ASVspoof2019 Challenge. In Proc. Interspeech 2019. Graz: ISCA, 2019; 1033–1037.

[14] Tomilov A, Svishchev A, Volkova M, Chirkovskiy A, Kondratev A, Lavrentyeva G. STC Antispoofing Systems For The ASVspoof2021 Challenge. In Proc. 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge. Brno: ISCA, 2021; 61–67.

Address for correspondence:

Il-Youp Kwak
Department of Applied Statistics
College of Business & Economics
84, Heukseok-ro, Dongjak-gu,
Seoul 06974
Republic of Korea
ikwak2@cau.ac.kr