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

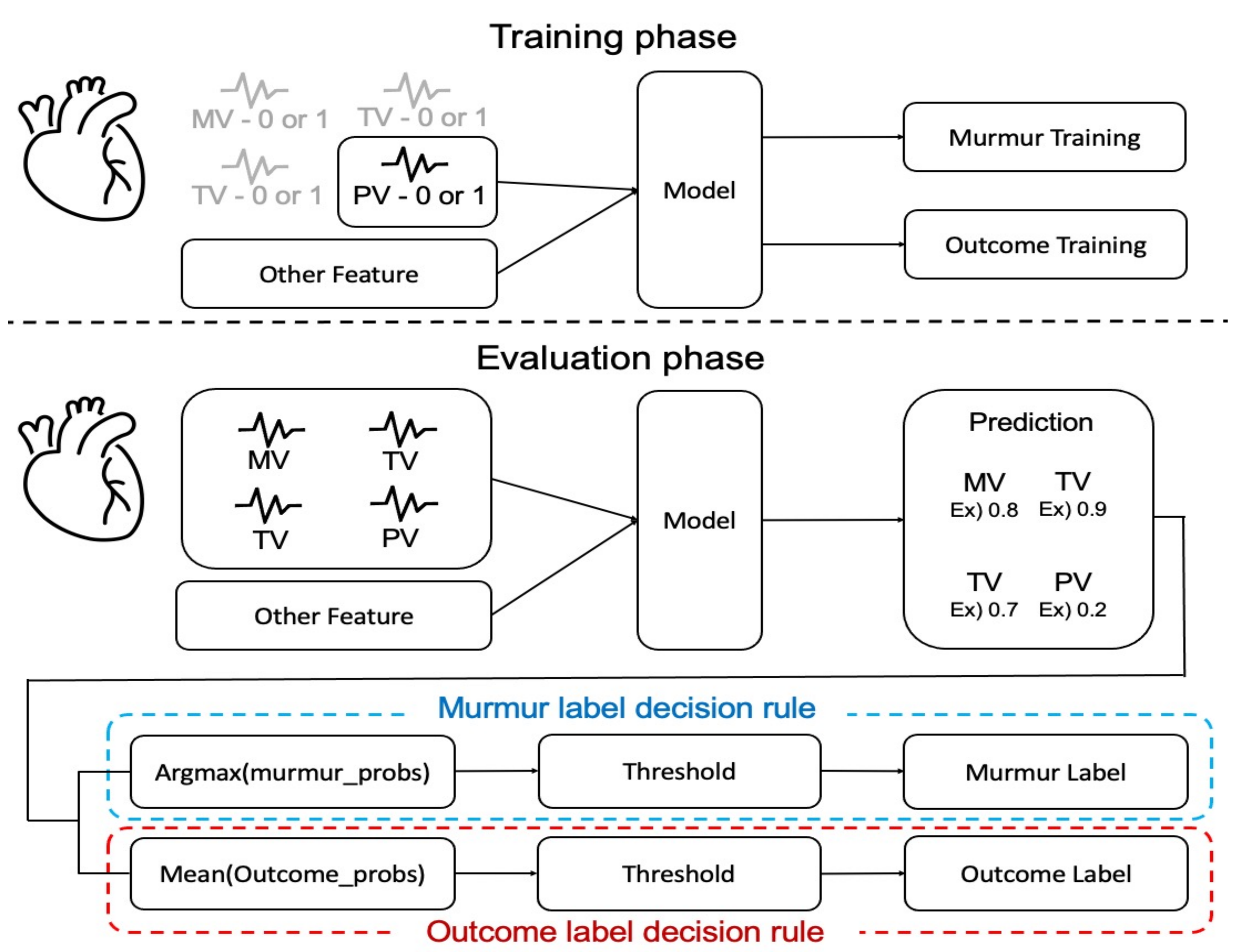
Jungguk Lee[1], Taein Kang[1], Narin Kim[1], Soyul Han[1], Hyejin Won[1], Wuming Gong[2] and Il-Youp Kwak[1]

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

Introduction

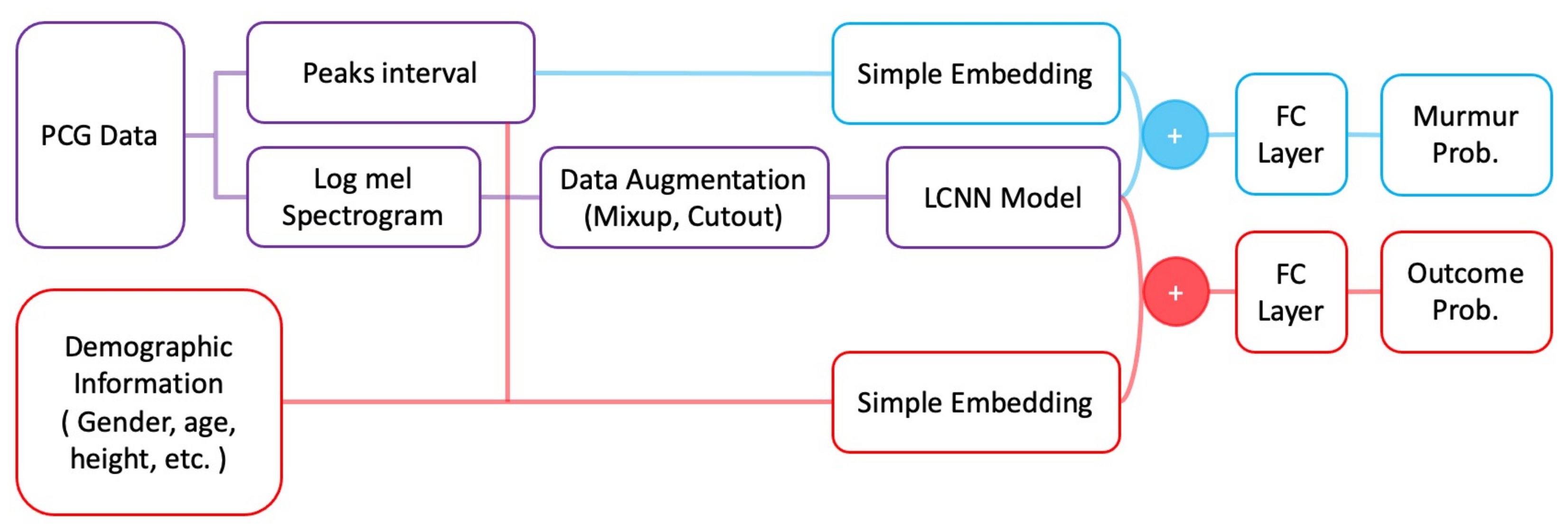
- Automated murmur detection(AMD) system detects murmur and abnormal heart movements in phonocardiogram (PCG) data.
- The AMD system will present a chance to enhance public health in medically underserved regions, such as developing nations, by detecting congenital and acquired heart disease and providing early diagnosis for clinical evaluation of patients.
- We compared LCNN and ResMax models for the 2D feature transformation of waveform data in order to develop the system and “LCNN” was chosen.
- Given that the RR interval of an ECG is a feature that may be used to describe HRV (Heart Rate Variability) efficiently, the "Peaks Interval" (PI) that corresponds to the RR interval was taken into consideration as an extra feature to express HRV in PCG.
- We explored several augmentation strategies in order to build a robust deep learning model on a small amount of data and finally “Cutout” & “Mixup” were chosen.
- we decided on the way of classifying labels in accordance with our label decision rule using threshold.

Figure 1. Overview of the AMD system



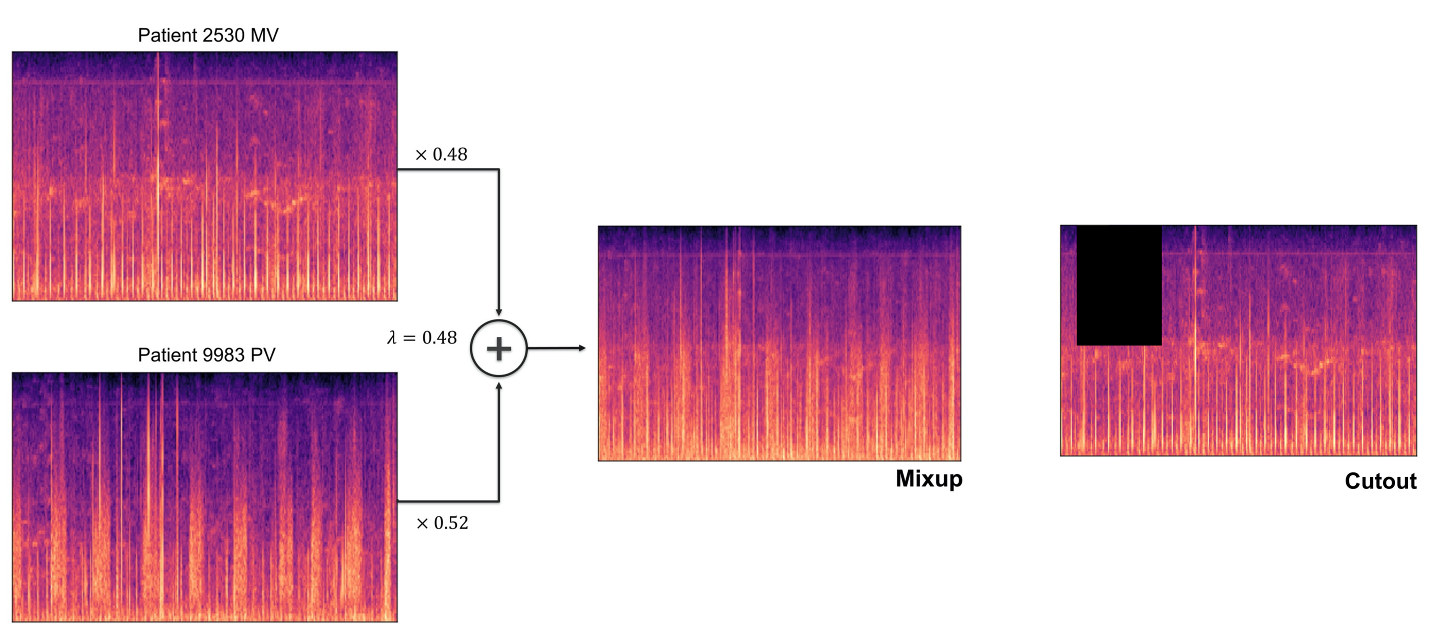
- We handled the PCG for each patient's auscultation site as a separate set of data and learnt the outcome and murmur classifiers individually.
- We had to predict the label for the patient during the evaluation phase, not the label for the patient’s auscultation site.
- The patient's label was predicted after processing the auscultation label probability of the auscultation position in accordance with the decision rule in the figure.

Figure 2. Proposed system architecture



- The purple line indicates that it was applied to both the outcome and the murmur classifier.
- The Murmur classifier uses the blue line, and the outcome classifier uses the red line.
- To generate spectrogram, we used 512 fast Fourier transform basis functions, a hop-length of 256 samples, 120 frequency bins, and converted Hz to mel. Subsequently, we extracted a log-mel spectrogram.
- In common, We extracted 2D log mel-spectrogram and peaks intervals from the 20 seconds of raw data for each patient's auscultation site , added data augmentation to the 2D features, and used these features as input to the LCNN model.
- Main distinction between the two classifiers is whether or not demographic information is added.
- After passing through a simple embedding, embedding was concatenated with the model’s embedding.

Figure 3. Data Augmentation

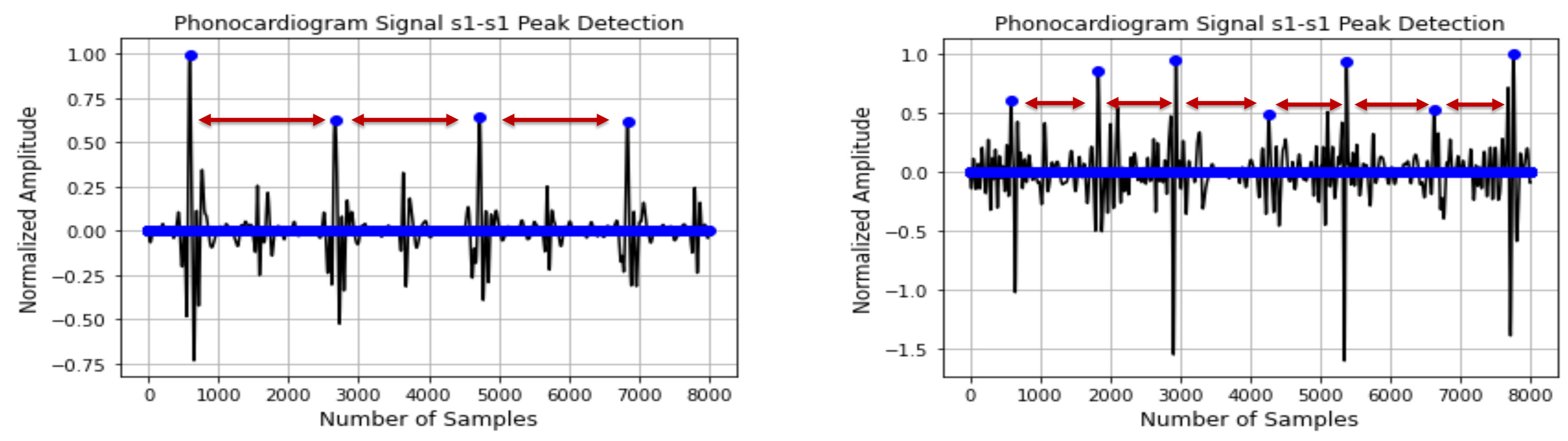


- Data augmentation improved the generalization performance of the model and prevented overfitting.
- We applied data augmentation to the audio feature & label to train the model more robustly.
- We used cutout and mixup augmentation to the 2D features.
- The mixup is as follows:

$$X = \lambda X_i + (1-\lambda)X_j$$
$$y = \lambda y_i + (1-\lambda)y_j$$

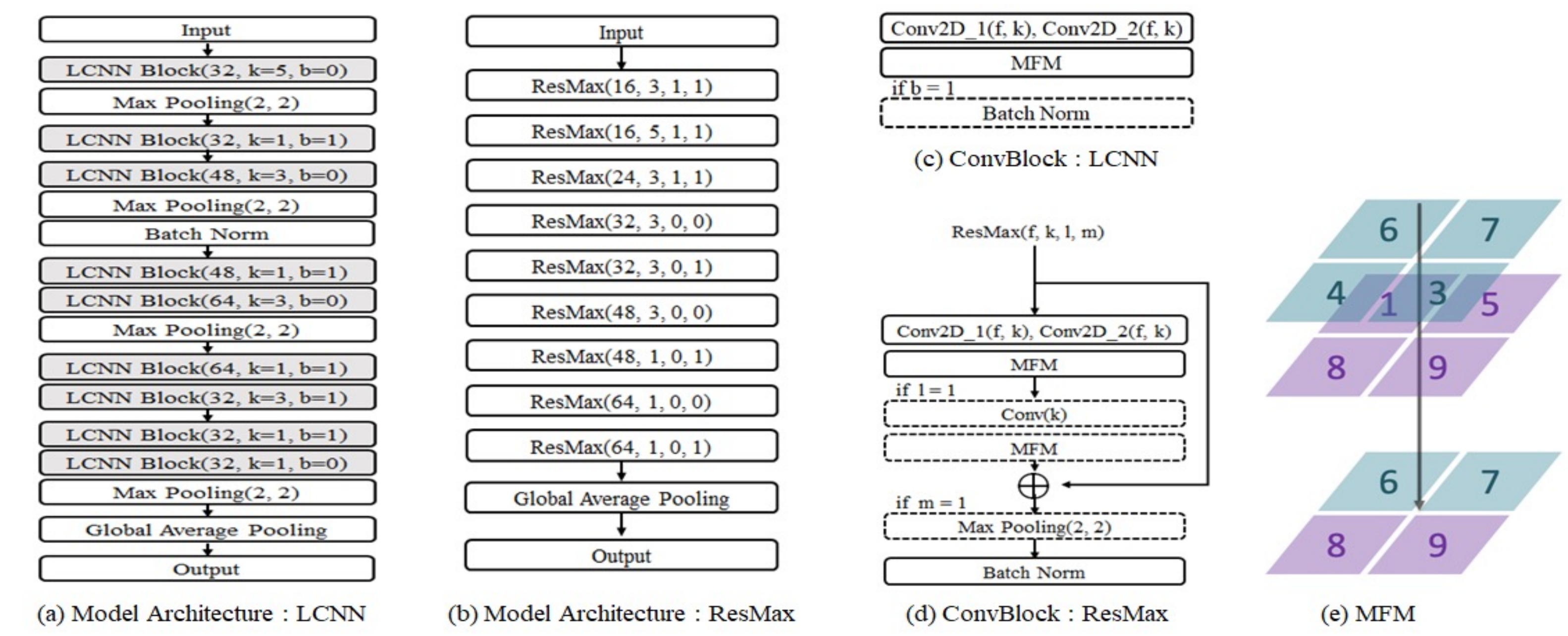
- Mixup : Where $\lambda \in [0,1]$ and is acquired by the sampling of the beta distribution with parameter α, β (0.7). X_i and X_j are different data samples; y_i and y_j are their corresponding labels.
- The cutout technique is to mask a part of the 2D feature with zero in a box shape.

Figure 4. PI Feature



- Peaks Interval means the time interval between peak points.
- Because the murmur patients have a noise, the patient has more peak points than a normal.
- Having more peak points means that the interval will be shorter.
- The Mean of PI of normal people was 49% longer than that of noisy patients in Challenge data.
- As a result, applying PI feature to the model was better for both the murmur detect and the outcome results.

Figure 5. Model Architecture



- (a) Deep LCNN model iterates 9 LCNN blocks.
- (b) The overall model of ResMax consists of 9 ResMax blocks.
- (c) The LCNN block consists of convolution, Max-Feature-Map (MFM), and an optional batch normalization layer (dotted block applied when b = 1).
- (d) Uses the skip-connection network in ResMax. The dotted boxes are optional.
- (e) Considers element-wise maximums in two layers with the same dimension.

Abbreviation: f, # filter; k, kernel size; b, option to batch norm; l, option to convolution and MFM; m, option to maxpooling;

Table 1, 2 Results

Table 1

Number	Model Configuration	Track 1 Weighted Accuracy	Track 2 Cost
1	LCNN - Mel	0.79	11,446
2	LCNN - Mel & CQT	0.76	11,489
3	LCNN - Mel & STFT	0.78	11,647
4	ResMax - Mel	0.77	11,354
5	ResMax - Mel & CQT	0.76	11,523
6	ResMax - Mel & STFT	0.77	12,074

- Our team trained the model comparing the LCNN and the ResMax.
- For each model, the Mel-spectrogram was used as a basis. In addition, STFT and CQT were added respectively.
- The selected model was the LCNN-Mel.

Table 2

Number	Model Configuration	Track 1 Weighted Accuracy	Track 2 Cost
1	LCNN - Mel	0.79	11,321
2	LCNN - Mel & demographic info	0.78	12,714
3	LCNN - Mel & PI	0.80	10,899
4	LCNN - Mel & demographic info & PI	0.79	9,711

- We also applied demographic information and PI feature to the models.
- For the weighted accuracy, there was the best performance when only PI feature was used. The accuracy was 0.80.
- In terms of cost metrics, both of PI and demographic information were used in outcome model. The model cost score was 9,711.

Summary

- We proposed a novel spectrogram based deep learning models.
- The Log-mel spectrogram was better than other 2D features of sound.
- Method of data augmentation were Mix-up and Cutout. We applied the method to Log-mel spectrogram.
- We considered PI and demographic information. As a result, it was possible to obtain higher performance than before.
- The performance of LCNN was better more than ResMax.
- Our LCNN model with PI feature achieved 0.734 weighted accuracy (31 out of 303 submitted systems)
- The model for outcome detection achieved a cost of 9,493 (24 out of 303 submitted systems) in the official phase of the George B. Moody PhysioNet Challenge.

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