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, propose 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-melspectrogram and LCNN was ranked 31 out of 303 submitted methods with a murmurs score of 0.734 and 24 out of 303 submitted methods with a outcomes score of 9493 in the official phase 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 and acquired 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 and monitoring is the phonocardiograph. A phonocardiograph creates a phonocardiogram (PCG), a particular waveform that accurately depicts the heartbeat intensity over time. The tasks for George B. Moody PhysioNet Challenge 2022 is to design a system that that detect murmur event. Two subtasks are based on weighted accuracy and expected cost of using murmur detection system.

The PhysioNet/Computing in Cardiology Challenge 2016 uses PCG data to distinguish between normal and abnormal in a similar manner. abnormal ones didn't provide information on a case-by-case basis and came from individuals with a known heart diagnosis. The ensemble model of the CNN-based classifier, the Adaboost classifier using segmentation features, and the Artificial NN model without segmentation features all outperformed the competition in 2016 [1–3].

We evaluated LCNN and ResMax models on waveform data to develop an automated murmur event detection system [3–5]. 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. Additionally, ResMax made an effort to use skip connection, which is the main concept of ResNet, to send more prior information.[6]

Heart Rate Variability has been used as the tool for assessing abnormalities in heart disease in prior competitions and numerous medical studies.[7] The ECG's RR interval is a feature that can effectively represent HRV[8, 9], thus we thought of the 'Peaks Interval' (PI), which corresponds to the RR interval, as an extra feature to express HRV in the PCG.

In order to develop a robust deep learning model from a limited quantity of data, we have experimented with several augmentation strategies. For masking methods, there are ffm and cutout.[10, 11] For combining input and label from other samples, there are cutmix and mixup.[12, 13] Cutout was chosen as the final augmentation technique after multiple tests.

After conducting multiple trials, we finally established a threshold as the final label decision rule, and we decided on the way of classifying labels in accordance with this threshold.

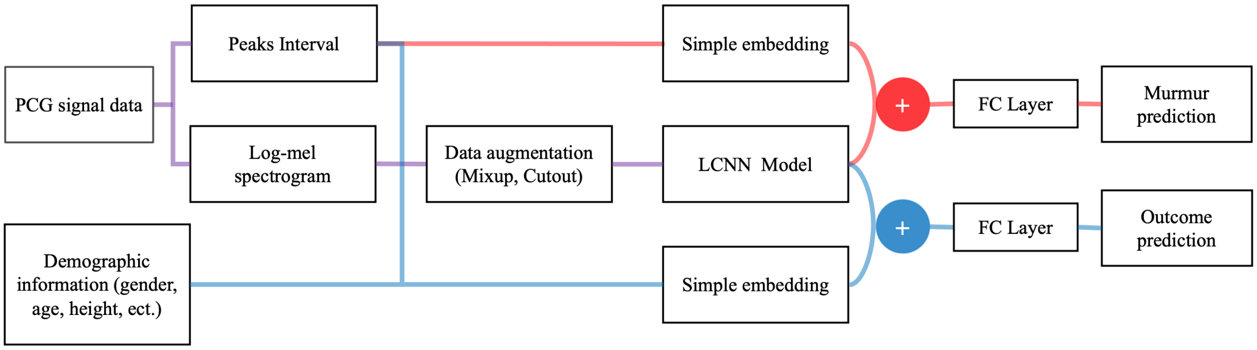


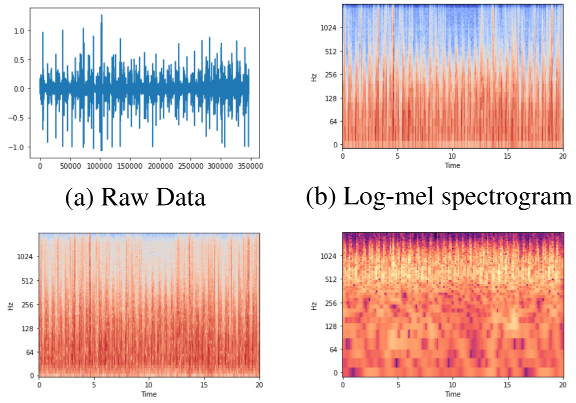
Figure 1: System Architecture

2. Methods

Figure 1 depicts our automatic detection architecture. The structure of murmur classifier was different from the outcome classifier's, and the main distinction between the two classifiers is whether or not demographic infomation is added. In common, We extracted 2D features and peaks intervals from the raw data for each patient site, added data augmentation to the 2D features, and used these features as input to the model. By passing through a simple embedding, the peaks interval feature was concatenated with the model's embedding.

2.1. Feature Extraction

We trained our models using 2D features which the were converted. There are three methods of converting data. Log-mel spectrum, STFT, and CQT. The figures are visualizations of 2d features about PCG data of a patient.



(a) Raw Data

(b) Log-mel spectrogram

(c) STFT

(d) CQT

Figure 2: Feature engineering

Consequently, Log-mel had the best performance. In order to compare performance, our model was fixed and only the feature was changed. Log-mel's performance is 16.3% better than CQT Under Weight Accuracy. In another case, the performance is 4.3% better than STFT. Because the log-mel spectrogram scales well the low frequency that humans can hear well, PCG has a low frequency band, so it is considered to fit well.

There are the demographic infomation. They are gender, age, height, weight, and pregnancy. For gender, there are two categories. Man and Woman. For Age, It is classified into six categories. The six categories are Neonate, Infant, Child, Adolescent and Young Adult. For height and weight, we use real-value, if they are nan-value, we use mean of those value. The features were used for our outcome model and not used the murmur detection model.

2.1.1. PI feature

Peaks Interval means the time interval between peak points. the murmur patients have a noise which occurs in the systolic or diastolic phase of the heart.[7] 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 are interested in the factor. We thought that if a patient was a murmur, the patient has more peak points than a normal. Having more peak points means that the interval will be shorter. Actually,the Mean of PI of normal people was 49% longer than that of noisy patients in Challenge data.

In fact, we wanted to use the value of PI in sequence form. However, due to the noise of the data, accurate PI difficultly was cauculated. and thus Mean of PI was used. This is where further research is needed.

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 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 cumix [12], cutout [11], and mixup [13].

2.3. Models

In this competition, PCG signal data was converted into Log-mel, STFT, and CQT spectrograms, and LCNN and ResMax models already been proven in many audio competitions ASVspoof 2017, 2019, and 2021 [5, 14–16] were applied.

2.3.1. LCNN

Compared with the Light CNN-9 model [4], this paper uses a deeper LCNN model which iterates six convolution layers and five network-in-network (NIN) layers. The LCNN block consists of convolution, MFM, and an optional batch normalization layer, as shown in Fig. 3(a) (dotted block applied when $b = 1$). The LCNN full model applied no batch normalization only on the 3rd convolution layer and applied it to the rest of the convolution layers and the NIN layer as shown in Fig. 4(a). Global average pooling layer, batch normalization, and dropout layer with probability 0.5 were used instead of fully connected layers.

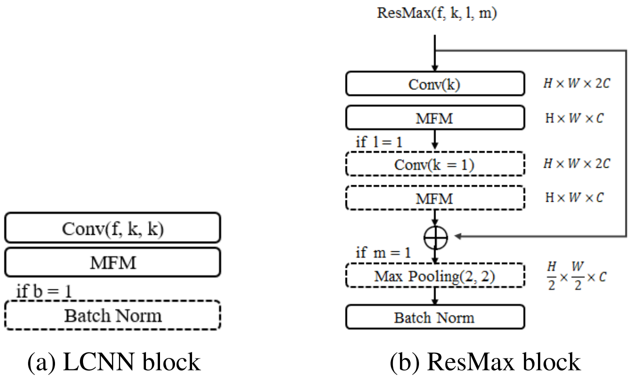
2.3.2. ResMax

ResMax is a model that showed excellent performance in the ASVspoof 2019 competition dataset [5]. 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 overall model of ResMax is shown in Fig. 4(b). There are a total of 9 ResMax blocks, and ResMax blocks are shown in Fig. 3(b).

3. Experiments

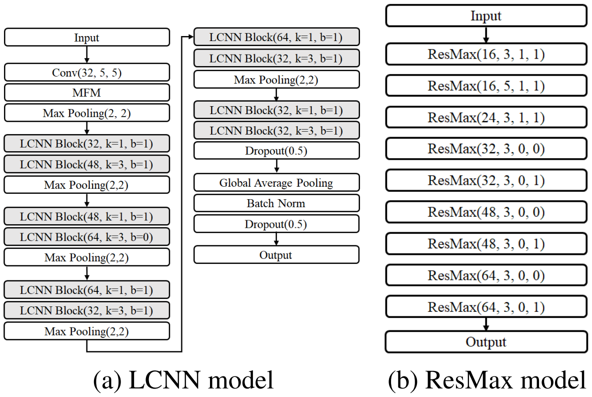
3.1. Dataset

As mentioned in 2.1, the demographic infomation include Age, Sex, Height, Weight, and Pregnancy status. The Murmur label has three categories. Presnet, Avocent and Unknown. Murmur Location is the PCG location where Murmur is existed. Among Murmur locations, the most audible location is called the Most Audible location. There were features related to Systolic & Diastolic, but we did not use the variable in the training and evaluation process.



(a) LCNN block (b) ResMax block

Figure 3: Model Blocks



(a) LCNN model (b) ResMax model

Figure 4: Model Architectures

In the Challenge data, there is a single or multiple file depending on the stethoscope position for each patient. We trained our model, the each of files was considered as one sample. 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 Murmur track, Use the highest probability among the values calculated for each stethoscope position. Classify using this probability value and a threshold.For Outcome track, the probability value calculated for each stethoscope position was averaged. The next process is the same as murmur.

3.2. Experimental setup

we set the ratio of train set to be used for learning and test set for test and evaluation as 8 to 2 of the 942 patients.[17] We judged that the distribution of patients with murmur class ’Present’ was important in both evaluation of murmur and outcome, so we distributed most audible locations equally for train set and test set, and ’Unkown’ and 'Absent' of murmur were randomly distributed 8 to 2.Since the distribution of 'outcome' class was similar, we used the train/test set divided as above for both murmur and outcome evaluations.

The competition evaluated the model in two aspects, and two different evaluation metrics were proposed for this purpose.These are the weighted accuracy that gave more weight to discriminating murmur and the cost metric that are evaluated from a perspective of cost when actually using the model. We conducted experiments for model development with training data and evaluating in terms of weighted accuracy and cost in test data for model advancement .

We used cost-sensitive learning because the importance of the murmur class and the outcome class are different in the evaluation metrics.[18] 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'.

3.3. Random search for LCNN and ResMax

In order to find hyperparameters in the LCNN and ResMax model fit, random search was used. The use of the mel-spectrogram feature, the stft feature, the cqt feature, the PI feature, the weights in cost-sensitive learning, the mixup augmentation, and the cutout augmentation were among the parameters that were taken into consideration as hyperparameters. By letting the parameters be chosen at random and running the model randomly 100 times, the best parameters were discovered. Several significant findings from random searches were described.

3.3.1. Additional use of CQT and STFT

Early trials showed that log-mel spectrogram feature is effective, and we experimented with performance changes while taking additional CQT and STFT features into account. In Table 1, both LCNN and ResMax models have similar or worse performance considering additional features. This led to the usage of solely log-mel spectrogram features.

Table 1: Evaluation on test data for the additional use of CQT and STFT.

Model	Added features	Weighted Accuracy	Cost
LCNN	-	0.79	11446
LCNN	CQT	0.76	11489
LCNN	STFT	0.78	11647
ResMax	-	0.77	11354
ResMax	CQT	0.76	11523
ResMax	STFT	0.77	12074

3.3.2. Additional use of other variables

We evaluated the impact of demographic infomation (age, sex, height, weight, pregnancy, recording location) and PI on the basic LCNN model. As seen in Table 2, the PI feature helped the LCNN model perform better. The model using both demographic information and PI features in the LCNN model demonstrated the best performance in terms of cost, whereas the model applying solely PI features to LCNN demonstrated the best performance in terms of weighted accuracy.

Table 2: Evaluation on test data for the additional use of other variables.

Model	Added Features	Weighted Accuracy	Cost
LCNN	-	0.79	11321
LCNN	demographic info.	0.78	12714
LCNN	PI	0.80	10899
LCNN	demographic info., PI	0.79	9711

3.4. Submitted system

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 infomation for outcome detection.

4. Conclusion

This paper covers the models used by our team participating in George B. Moody PhysioNet Challenge 2022. In order to distinguish murmur or outcome from heartbeat waveform data (PCG), weighted accuracy and cost were provided as evaluation metrics in this year's competition.

We proposed a novel spectrogram based deep learning models. Our LCNN model with PI feature achieved 0.734 weighted accuracy (31 out of 303 submitted systems), and our LCNN model for outcome detection achieved a cost of 9493 (24 out of 303 submitted systems) 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] Potes C, Parvaneh S, Rahman A, Conroy B. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In 2016 computing in cardiology conference (CinC). IEEE, 2016; 621–624.

[2] Zabihi M, Rad AB, Kiranyaz S, Gabbouj M, Katsaggelos AK. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In 2016 computing in cardiology conference (CinC). IEEE, 2016; 613–616.

[3] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. volume 86. Ieee, 1998; 2278–2324.

[4] 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.

[5] Kwak IY, Kwag S, Lee J, Huh JH, Lee CH, Jeon Y, Hwang J, Yoon JW. 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.

[6] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.

[7] 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.

[8] 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.

[9] 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.

[10] Kwak IY, Choi S, Yang J, Lee Y, Oh S. Cau_ku team's submission to add 2022 challenge task 1: Low-quality fake audio detection through frequency feature masking. arXiv preprint arXiv220204328 2022;.

[11] DeVries T, Taylor GW. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv170804552 2017;.

[12] 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.

[13] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. mixup: Beyond empirical risk minimization. arXiv preprint arXiv171009412 2017;.

[14] 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.

[15] 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.

[16] 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.

[17] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C, Jorge A, Mattos S, Hatem T, Tavares T, et al. The circor digiscope dataset: from murmur detection to murmur classification. volume 26. IEEE, 2021; 2524–2535.

[18] Elkan C. The foundations of cost-sensitive learning. In International joint conference on artificial intelligence, volume 17. Lawrence Erlbaum Associates Ltd, 2001; 973–978.

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