Two-Stage Multitask-Learner for PCG Murmur Location Detection

Maurice Rohr, Benedikt Müller, Sebastian Dill, Gökhan Güney, Christoph Hoog Antink

KIS*MED – AI Systems in Medicine,
Technische Universität Darmstadt, Darmstadt, Germany

Abstract

Heart murmurs are a potential symptom for cardiac diseases , which can be captured easily by smartphones or similar devices. Hence, the automated analysis of murmurs in heart sound recordings may provide a cost-efficient pre-screening method for heart conditions. In this study, we present an approach for detecting heart murmurs that utilizes a Pooling-based Artificial Neural Network (PANN) structure to extract features from audio waveforms of arbitrary lengths. It can classify single recordings based on recording location and the extracted features in an end-to-end manner. The approach is inspired by the multiple instance learning framework.

We performed a 10-fold stratified cross-validation on our training set and show that the results are consistent with the evaluation on the hidden test set of the PhysioNet challenge 2022. Our team ***Heart2Beat*** *was ranked 12th in the murmur detection task and 11th in the clinical outcome task and achieved a weighted accuracy metric score of 0.751 and a clinical outcome cost of 12244 respectively.*

1. Introduction

Cardiovascular diseases are the cause of approximately one third of all deaths globally [1] and a major focus of risk factor analysis and screening. Heart murmurs are indicators of heart diseases and have a high prevalence, yet recognizing them requires strong cardiac auscultation skills, which among many physicians are sub-optimal [2]. Therefore, technical and automated solutions for heart murmur detection are required.

Heart murmurs are essentially audible vibrations caused by perturbations of the blood flow such as strong pressure gradients or velocity changes. Mostly, they arise when heart valves are not opening or closing correctly. Our aim which is also the goal of the George B. Moody Challenge 2022[3], thus was to predict the presence of heart murmurs from Phonocardiograms (PCG).

PCGs are recordings of all sounds of the heart during a cardiac cycle. This includes sounds such as the first (S1) and second (S2) heart sound, but also murmurs. Commonly, a segmentation based on S1/S2 is performed before computing features using e.g. the (mel-)spectorgram for PCG analysis.

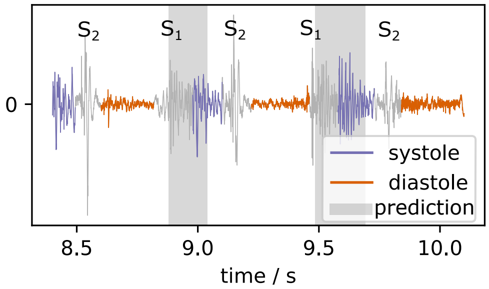

Figure 1. Murmur prediction (grey bar) generated by MILU-Net on recording 9979_AV [4] where diastolic murmur is present.

By formulating the detection of heart murmurs as a two stage multiple instance learning (MIL) problem with "weakly labeled" data, as known from sound event detection [5], we can show that a machine learning model can find the relevant segments for PCG analysis on its own, without the need for prior segmentation. Weakly labeled in this context means that for each sound recording, only a single tag is provided without knowing the exact onset and offset times of the relevant sound bites. More specifically, in MIL [5], a weakly labeled dataset $D = \{B_n, \mathbf{y}_n\}^N$ consists of a set of bags $B_n = \{\mathbf{x}_{n,1}, \dots, \mathbf{x}_{n,T_n}\}$, where each bag is a collection of instances. T_n is the number of instances in the n-th bag and each instance has a length (duration) d. Given a particular sound class k, a bag (PCG recording) is considered *positive* (e.g. "murmur present", $y_{k,n} = 1$) if it contains at least one positive instance and *negative* ($y_{k,n} = 0$) if it contains no positive instance at all. The problem formulation heavily implies that, given a prediction for each instance of a bag, the condensed label is given by the maximum of these predictions. For both hierarchically ordered sub-problems, predicting murmur at time instance and location level, "weak labels" are available from the CirCor dataset [4].

Our two-stage approach aims to guide physicians to the

the most audible murmur location, while being able to enhance murmur sounds.

2. Methods

We split the methods part in two problems: Primarily, we want to construct a model (*MILU-Net*) which, based on weak labels, produces a fine resolution murmur detection which manages to explain the final decision about the presence of murmur for a single recording. Secondly, based on that structure, we design a model (*PANN*) that achieves good murmur prediction accuracy for a set of multiple recordings of a particular patient at the cost of losing interpretability with respect to a single recording. Both approaches rely on the same *pre-processing*.

Pre-processing. The recordings are pre-processed independently of location by removing segments with low signal quality based on signal-to-noise ratio and saturation [6] and applying bandpass filtering (10th order Butterworth filter, 10 to 800 Hz). Only during training, all signals are cut or zero padded to a fixed length of 8.2 s to increase efficiency.

MILU-Net. The MILU-Net model consists of a simple U-Net [7] structure for feature generation and a part which facilitates MIL. The U-Net guarantees that its output is the same dimension as the input and thus provides a "strong" label for each sample of the input. It consists of 5 down- and up-sampling convolutional layers with batch norm and ReLU activations and a final output convolutional layer that summarizes the features into a 1D signal. Most importantly, the output of the U-Net is used in a softmax-pooling layer [5] with sigmoid activation to obtain a scalar output murmur probability. By using the softmax-pooling layer, we loosen the MIL assumption, which implies max-pooling. In return, we achieve a more robust training that depends less on the initialization, because the output depends on all instances instead of single particular instances that are chosen randomly due to parameter initialization. The model is overfitted to the available data in order to verify if it can learn the unique attributes of murmur. The excitation of the U-Net is then directly related to the relevance and "murmurness" of the respective signal part. An example prediction is shown in Fig. 1.

PANN. The PANN model in Fig. 2 follows the MIL idea loosely by widening the single output signal to an array of feature signals. It employs a convolutional encoder consisting of 6 blocks of 1d-convolutions (kernel size=5, padding=same), batch norm, 1d-convolution, dropout and max pooling (stride=2) to encode the signals in a feature-rich presentation. These features, which can be thought of as time signals, are then processed by an adaptive pooling layer which produces a fixed-size output of 30 features in total (max-pooling=15, average-pooling=10, min-pooling=5). The intuition of the convolution block is that it gets activated by murmurs in the respective parts in each segment. The approach then takes advantage of the periodicity of murmurs by employing the pooling layer that collects and summarizes the information about murmur appearance from all segments of the signal, rendering the output feature dimensions independent of the input length. The output features of the pooling layer are then combined in a fully connected layer (30x64+5 input, 100 hidden and 20 output neurons) with the one-hot encoded recording locations. These outputs are then evaluated in a linear decision layer with softmax activation function. By processing each recording location separately, we enable the user to verify the suspected murmur origin.

As depicted in Fig. 3, in a second stage a multi-label model is fed with the features and the encoded output of the PANN model which includes (1) features summarizing the relative median energy in five frequency bands, (2) the one-hot encoded recording location, and (3) demographic features (age,sex,weight). The multi-label model is a simple feed forward neural network with 4 hidden layers with (123, 492, 246, 20) neurons, batch norm after each hidden layer and leaky ReLU activations and one dropout layer at the end. The output layers consist of 3 neurons for murmur prediction and 2 for outcome prediction with softmax activation each.

Augmentation. Due to the small training dataset we employ data augmentation. During training and after pre-processing, one or multiple augmentations are performed at random: *scaling*, *gaussian noise*, *drop*, *cutout*, *shift*, *resampling*, *random resampling*, *sine wave*, *bandpass filtering*. Scaling randomly rescales the signal. Gaussian noise adds gaussian noise to the signal. Drop randomly sets signal values to zero. Cutout randomly sets signal intervals to zero. Shift randomly shifts the signal in time (creating zeros at either end). Random resampling creates smooth time offsets simulating a changing heart rate. Resampling linearly resamples the signal to another sampling frequency, simulating another heart rate. Sine wave adds a random sine wave to the signal. Bandpass filtering randomly applies a bandpass filter between 0.2 and 45 Hz.

Training. During both training stages we employ *weighted cross entropy loss*. The weights are chosen as the inverse of the relative frequency of each class to counteract class imbalance. For multi-label model training these weights are multiplied by the respective class weights of the weighted accuracy score employed by the challenge. First the PANN model is trained using a learning rate of 0.001 which is then decreased by a factor of 0.3 after each 30 epochs. The best model is picked based on the mean of training and validation accuracy.

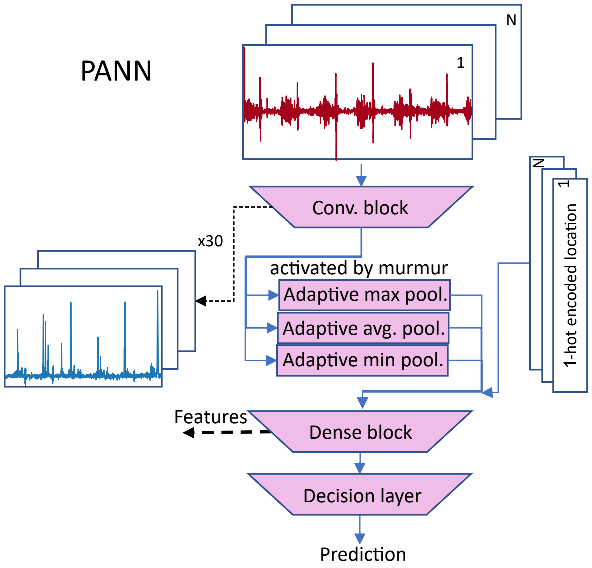

Figure 2. General Structure of PANN Model for murmur detection

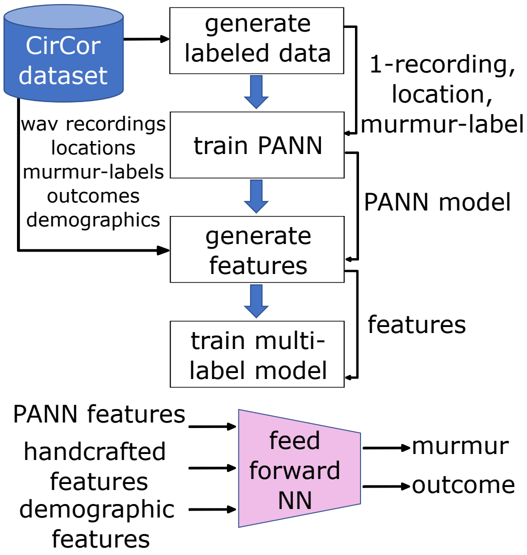

Figure 3. Training of multi-label model

3. Results

We test the PANN model by computing the accuracy for each auscultation location separately. The labels are created by considering all recordings of a subject with absent/unknown murmur as "Absent"/"UNK". If murmur is present, the locations marked as hearable are labeled "Present", the rest "UNK". The accuracy of the intermediate decisions of PANN model (Tab. 1) for a confidence threshold of 0.8 show that murmur is detected with an accuracy of $>$ 90% for all recording locations separately. Given the true label for the subject is "Present", the accuracy in predicting the correct murmur locations is reduced.

Table 1. PANN prediction for presence of murmur per auscultation location [4] given the experts label says "Present" and the total of high signal quality recordings (n=2803).

accuracy	AV	PV	TV	MV	total
present	0.733	0.754	0.800	0.756	0.761
total	0.909	0.911	0.911	0.921	0.913

Table 2. 5-fold cross-validation for different augmentations ordered descendingly by improvement on murmur prediction performance on training.

augmentation	AUROC	AUPRC	F-measure
g. noise	0.839 (0.047)	0.664 (0.043)	0.546 (0.110)
scaling	0.840 (0.020)	0.659 (0.050)	0.537 (0.068)
shift	0.835 (0.029)	0.647 (0.035)	0.536 (0.047)
resample	0.832 (0.046)	0.652 (0.036)	0.532 (0.057)
bandpass	0.803 (0.062)	0.632 (0.068)	0.545 (0.043)
cutout	0.791 (0.095)	0.620 (0.082)	0.557 (0.090)
none	0.797 (0.077)	0.616 (0.064)	0.532 (0.080)
sine wave	0.808 (0.062)	0.617 (0.062)	0.514 (0.047)
drop	0.814 (0.031)	0.609 (0.037)	0.500 (0.041)
r. resample	0.782 (0.112)	0.598 (0.127)	0.505 (0.121)

We performed a 5-fold stratified cross-validation for each of the signal augmentations we used during training, by applying each singular augmentation with a probability of 0.25. The effect on the final murmur predictions of each augmentation as well as the validation measures for no augmentation ("none") is shown in Table 2.

We performed a 10-fold stratified cross-validation of the final model (Fig. 3) resulting in the following scores listed as mean (standard deviation): **Murmur** AUROC 0.831 (0.038), AUPRC 0.657 (0.059), F-measure 0.572 (0.076), Accuracy 0.722 (0.088), Weighted Accuracy 0.715 (0.077); **Outcome** AUROC 0.628 (0.047), AUPRC 0.634 (0.047), F-measure 0.580 (0.056), Accuracy 0.590 (0.057), Cost 13640 (2401).

The official results on the validation set and hidden test set are listed in Tables 3 and 4.

Table 3. Weighted accuracy metric scores (official challenge score) for our final selected entry (team Heart2Beat) for the murmur detection task, including the ranking of our team on the hidden test set.

Training	Validation	Test	Ranking
0.803	0.72	0.751	12/40

Table 4. Cost metric scores (official challenge score) for our final selected entry (team Heart2Beat) for the clinical outcome identification task, including the ranking of our team on the hidden test set.

Training	Validation	Test	Ranking
10810	9135	12244	11/39

4. Discussion

The goal was to learn high resolution murmur detection on weakly labeled recordings. As can be seen in Fig. 1 this works in principle. The main problem is that generalizability is difficult to achieve and oftentimes recurrences of murmur in the same recording do not lead to the same model activation, which is in part due to the fact, that one case of murmur is enough for the network to be correct about a complete recording. This renders the application for murmur enhancement impractical. While the accuracy for murmur prediction on the recording level of PANN is quite good ($> 90\%$), for the subject level, a rule-based system that followed the MIL idea (regarding murmur as "Present" if there was murmur in at least one location, "Absent" if murmur is absent in all locations and "UNK" otherwise) turned out to score badly (weighted accuracy of 0.56).

Random resampling augmentation appears to reduce the performance of the model significantly in one of the folds while providing no improvement in the others. Rescaling the signals or adding Gaussian noise during training are consistently improving generalization of the model during training. All other augmentation techniques did not result in a significant difference, although more variance in the data certainly helps training more generalized models. Thus, we decided to use them either way but to a lesser extent. Surprisingly, while rescaling reverts normalization of the signals which is standard procedure during training of most machine learning models, it improves generalization. Adapting the probability of an augmentation being applied based on ranking table 2 leads to an improvement in all relevant metrics.

The results of the 10-fold stratified cross-validation are consistent with the evaluation on the test set. However, the training of the model has shown to be highly sensitive to initialization and the selection of training data. Both are an immanent problem of pooling based neural networks [5]. Training based on cost functions for both classification tasks simultaneously reduced the classification accuracy for murmur only by 3.6 %.

5. Conclusion

We present a model based on the MIL network to create fine-granular murmur detection in PCG recordings. While we can show that a basic model generates good results in finding specific murmur locations only from training on weak labels, this learning does not yet translate to good prediction accuracy on unseen data. An adapted model based on the same pooling ideas but with decreased level of interpretability achieves competitive results in both murmur detection and clinical outcome prediction.

References

[1] Roth GA, Mensah GA, Fuster V. The Global Burden of Cardiovascular Diseases and Risks: A Compass for Global Action. Journal of the American College of Cardiology 2020; 76(25):2980–2981.

[2] Vukanovic-Criley JM, Criley S, Warde CM, Boker JR, Guevara-Matheus L, Churchill WH, Nelson WP, Criley JM. Competency in Cardiac Examination Skills in Medical Students, Trainees, Physicians, and Maculty: A Multicenter Study. Archives of internal medicine 2006;166(6):610–616.

[3] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, Perez-Alday EA, Sadr N, Sharma A, Mattos S, Clifford GD. Heart Murmur Detection from Phonocardiogram Recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv 2022;.

[4] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C, Jorge A, Mattos S, Hatem T, Tavares T, Elola A, Rad AB, Sameni R, Clifford GD, Coimbra MT. The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification. IEEE Journal of Biomedical and Health Informatics 2022;26(6):2524–2535.

[5] McFee B, Salamon J, Bello JP. Adaptive Pooling Operators for Weakly Labeled Sound Event Detection. IEEEACM Transactions on Audio Speech and Language Processing 2018;26(11):2180–2193.

[6] Plesinger F, Viscor I, Halamek J, Jurco J, Jurak P. Heart Sounds Analysis Using Probability Assessment. Physiological measurement 2017;38(8):1685.

[7] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, 2015; 234–241.

Address for correspondence:

Maurice Rohr
KIS*MED, TU Darmstadt
Merckstr. 25, 64283 Darmstadt, Germany
rohr@kismed.tu-darmstadt.de