# A Lightweight Robust Approach for Automatic Heart Murmurs and Clinical Outcomes Classification from Phonocardiogram Recordings

Hui Lu[1,3], Julia Beatriz Yip[2,4], Tobias Steigleder[2], Stefan Grießhammer[2], Maria Heckel[2], Naga Venkata Sai Jitin Jami[4], Bjoern Eskofier[4], Christoph Ostgathe [2], Alexander Koelpin [3]

[1] Chair of Electronics and Sensor Systems, Brandenburg University of Technology, Cottbus, Germany
[2] Department of Palliative Medicine, Universitätsklinikum Erlangen, Erlangen, Germany
[3] Institute of High-Frequency Technology, Hamburg University of Technology, Hamburg, Germany
[4] Machine Learning and Data Analytics (MaD) Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

## Abstract

*Cardiac auscultation provides an efficient and cost-effective way for cardiac disease pre-screening. The George B. Moody PhysioNet Challenge 2022 aimed to detect heart murmurs and clinical outcomes with heart sound recordings from multiple auscultation locations. Our team HearHeart proposed a lightweight convolutional neural network (CNN) to detect heart murmurs and a random forest model to classify clinical outcomes. 128 Mel-spectrogram features and wide features like the socio-demographic data and statistical features are extracted. Different techniques are employed to migrate the data imbalance and model the overfitting problem. We used two data augmentation methods, noise injection and spectrogram augmentation in time and frequency domain to increase the training samples and avoid overfitting during training. Besides, weighted loss functions are applied to both tasks to deal with data imbalance. In the end, we ensembled the models from cross-validation and used voting for the final classification. We achieved a murmur score of 0.791, and a clinical outcome score of 11731.64 on 5-fold cross-validation in the hidden validation set. While on the hidden test set, we achieved a murmur score of 0.780, and a clinical outcome score of 12110, placing our team 1st and 10th in the challenge tasks, respectively.*

## 1. Introduction

Cardiovascular diseases are the leading cause of death in the world [1]. Cardiac auscultation via stethoscopes is one of the most important and cost-effective tools for pre-screening cardiovascular diseases. Two fundamental heart sound components, the first (S1) and second (S2) heart sounds, can be heard through cardiac auscultation. The S1 results from the closure of the mitral and tricuspid valves, while S2 is caused by the aortic and pulmonary valves. Besides these sounds, murmurs can be heard in the auscultation process and their presence can indicate irregularities of the heart [2, 3].

Therefore, the detection and classification of heart murmurs are essential for an accurate clinical diagnosis. However, auscultation interpretation requires long-year expert knowledge. Computer-aided auscultation systems based on phonocardiogram (PCG) signals have gained increasing interest in recent years [4]. Some studies have investigated computer-aided murmur detection [5–9]. However, some downfalls of these methods are the small size of the datasets used and the need for a reliable heart sound segmentation.

In this context, we developed a lightweight convolutional neural network (CNN) model for murmur and a random forest model for clinical outcome classification in the PCG recordings for the George B. Moody PhysioNet Challenge 2022 [10], while using the largest pediatric heart sound dataset [11] and avoiding primary heart sound segmentation. The main contribution of our work is a robust method that can be translated to the clinical environment, which contains challenges such as diverse types of noise and usually limited resources.

## 2. Methods

### 2.1. Data Introduction

The challenge used 60% of the CirCor DigiScope Dataset [11], a public dataset made of pediatric heart sound recordings. The dataset contains 5,282 recordings of 1,568 patients under 21 years old, which are recorded from 4 main auscultation locations: pulmonary valve (PV), aortic valve (AV), mitral valve (MV), tricuspid valve (TV), and other (Phc). Additionally, the socio-demographic data

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.165

of each patient is provided.

We can observe in Fig. 1 the data imbalance between the heart murmurs classes, in which the absent class contains more samples compared to the present and unknown classes. Besides, we can also confirm that both tasks of murmur and clinical outcome classification are not correlated.

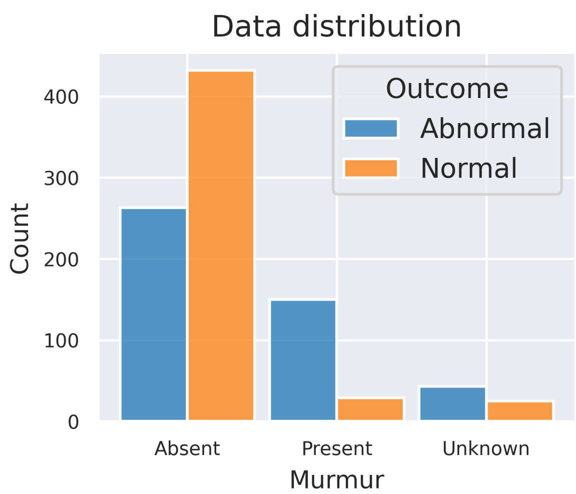


Figure 1. Data distribution in murmur and outcome.

## 2.2. Data Pre-processing

Since the recordings vary in duration from 4.8 to 80.4 seconds (mean=22.9±7.4 s), we selected a fixed duration of 15s for each recording. During training, we randomly cropped recordings that exceeded the defined length, while the shorter ones were zero-padded. As for the validation, zero padding was also applied to the shorter recordings as in the training process. Whereas, the longer recordings were segmented into patches with an overlap of 7.5 seconds to increase the number of validation samples.

Next, we extracted from the selected recording a Mel-spectrogram, with 128 Mel bands within the range of 25-2,000 Hz, followed by a Hamming window with a window size of 50 ms and a frameshift of 25 ms. For each recording, we also extract wide features. One part of the wide features is composed of age, gender, and pregnancy status and are embedded using one-hot vectors. While the other part of the wide features includes statistical features such as zero-crossing rate, spectral centers, and spectral bandwidth.

## 2.3. Data Augmentation

Two data augmentation approaches are applied to reduce overfitting. At first noise injection is used, where

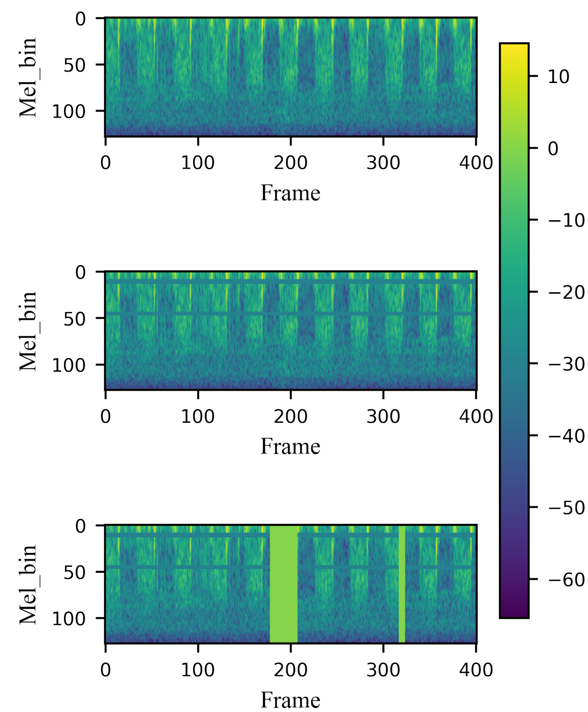


Figure 2. Spectrogram augmentation. Top: original Mel-spectrogram, Middle: Mel-spectrogram after masking in frequency domain, Bottom: Mel-spectrogtram after masking in time and frequency domain

zero mean 15 dB Gaussian noise is added to the recording with the possibility of 0.5 during training. Then spectrogram augmentation on frequency and time domain is conducted, where randomly selected multiply blocks of frequency channels and blocks of time steps are masked. Fig.2 shows one example of spectrogram augmentation in the frequency and time domain.

## 2.4. Murmur Model Structure

We propose a lightweight convolutional neural network (CNN) architecture composed of two branches (Fig. 3) to classify heart murmurs. The CNN uses the Mel-spectrogram as input in the main branch, where it is consecutively fed into convolutional blocks, while the second branch feds the wide features into a fully connected layer (FC). The outputs of the main and secondary branches are concatenated and fed to another FC layer for the final murmur classification.

## 2.5. Clinical Outcome Model

For the clinical outcome classification, we choose the random forest classifier since it was shown to hold stable performance despite the presence of noisy data [12], which

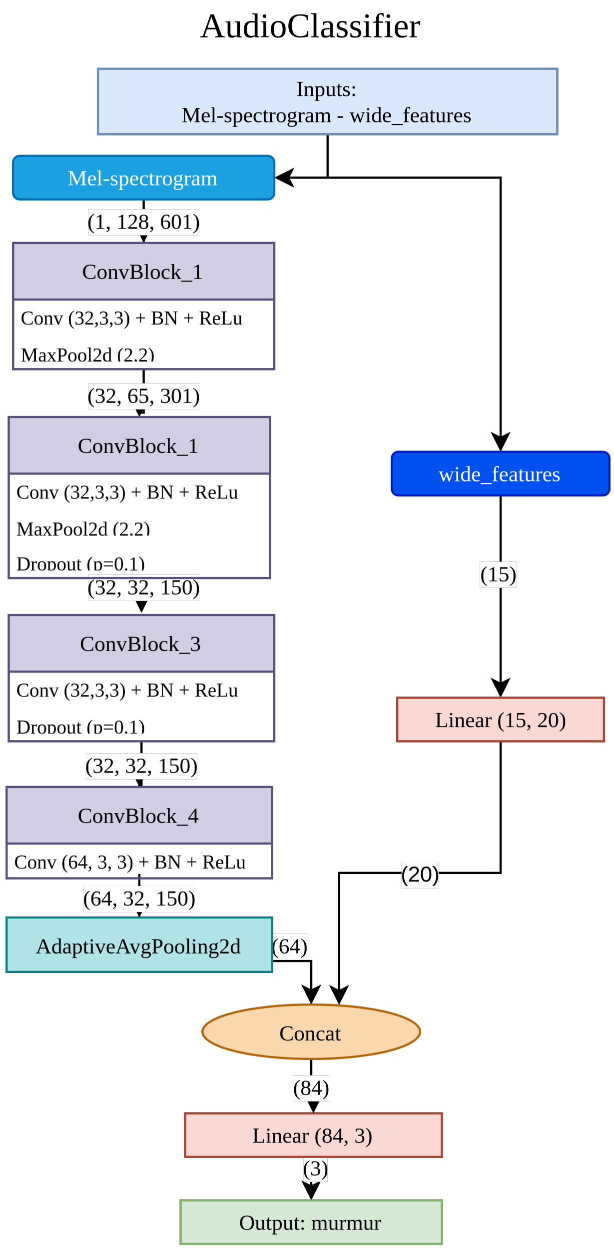


Figure 3. Illustration of the proposed model structure. The network uses the Mel-spectrogram and the wide features as inputs and returns as outputs the heart murmurs classification.

is the case for the used data. The random forest consists of a collection of tree-structured classifiers, where each tree contributes with a unit vote for the final classification [12]. However, as with any ensemble method, its performance also relies on the strength of each tree and the number of trees.

With these considerations, we performed a grid search to optimize hyperparameters in a 5-fold cross-validation style using each fold to evaluate the performance for the clinical outcome classification task.

## 2.6. Training Details

We employed a 5-fold patient-wise cross-validation technique to reduce model variance. We also used a weighted cross-entropy loss function for murmur classification, where we attribute the weights 5, 3, and 1 to the present, unknown, and absent classes, respectively. We used a weighted loss function due to the higher cost of misclassification of the present class.

For the murmur classification task, we used a trained CNN with a batch size of 24. The loss function is optimized with the Adam optimizer with an initial learning rate of 0.001. We also used multi-step learning rate optimization, which reduces tenfold the learning rate at the 30th, 50th, and 80th epochs. Early stopping with the criteria of best challenge score is employed.

While for the clinical outcome classification, the random forest with the best hyperparameters found in the grid search has been used. The hyperparameter optimization resulted in a random forest with 100 trees, where each tree could have the maximum number of leaves in each node of 36. Similar to the murmur classification task, we also attribute weights 5 and 1 to the abnormal and normal classes, respectively, due to the higher cost of misclassification of the abnormal class.

# 3. Results and Discussion

The results of the 5-fold cross-validation are shown in Table1. In addition to murmur and outcome scores, we also report the F1 score and accuracy of each class in two tasks.

The challenge used different scoring systems for ranking the outputs. For the classification task, a weighted accuracy is used as score function. While for the clinical outcome classification task, the metric used was a non-linear function that represents the cost of the treatment. Therefore, for the first task, the score should be as high as possible, while for the second, the score should be as low as possible. More information on the scoring system can be found in the challenge description [10].

Our proposed method achieved an average murmur score of 0.793. Further analysis of the results showed an average F1 score of 0.119 for the unknown class, possibly due to insufficient samples for this class, and could not be solved by only attributing more weights to counterbalance the lack of training samples.

For the clinical classification task, our method achieved an average score of 11731. We could also observe al-

Table 1. Model performance on 5-fold cross-validation, hidden validation, and hidden test sets.

| | Local cross validation | Hidden validation set | Hidden test set | Ranking |
|---|---|---|---|---|
| Murmur score | 0.793±0.03 | 0.747 | 0.78 | 1st |
| Clinical Outcome score | 11731±285 | 9903 | 12110 | 10th |

though the method achieved an average accuracy of 0.914 for the abnormal class, it was only 0.195 for the normal class. Since the clinical outcome annotations are made by medical experts considering extra information, like patient history, physical examination, and echocardiogram report, we believe that only using PCG recordings could not be sufficient for accurate clinical outcome classification.

Moreover, we submitted an ensemble of the trained models from cross-validation for the challenge. The results from all evaluations, as well as the ranking of our solutions in the challenge, can be seen in Table 1. The outstanding performance of our method in the hidden validation and test set demonstrated that our method can generalize and perform well in presence of new data.

## 4. Conclusion

In this paper, we proposed a lightweight CNN model for murmur detection and a random forest model for clinical outcome classification based on PCG recordings. We extracted statistical and frequency-domain features from each recording as part of wide features vector, which is concatenated with features learned from the CNN. In the end, we achieved promising results for both tasks, indicating that our models can be suitable for real-time application in the clinical environment. Furthermore, their computational efficiency and robust performance can improve cardiac auscultation and pre-screening routines.

## Acknowledgments

This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) within the iCampµs2 project, grant number 16ME0420K. Furthermore, this work was also funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant SFB 1483 - Project-ID 442419336.

## References

[1] (WHO) WHO. Cardiovascular diseases (cvds). World Health Organization WHO June 2021;.

[2] Hoeting NM, McCracken CE, McConnell M, Sallee D, Iannucci GJ, Oster ME. Systolic ejection click versus split first heart sound: Are our ears deceiving us? Congenital Heart Disease 2017;12(4):417–420.

[3] Dornbush S, Turnquest AE. Physiology, Heart Sounds. In statPearls. StatPearls, July 2022; .

[4] Li S, Li F, Tang S, Xiong W. A review of computer-aided heart sound detection techniques. BioMed research international 2020;2020.

[5] Chen K, Mudvari A, Barrera FG, Cheng L, Ning T. Heart murmurs clustering using machine learning. In 2018 14th IEEE International Conference on Signal Processing (ICSP). IEEE, 2018; 94–98.

[6] Alam S, Banerjee R, Bandyopadhyay S. Murmur detection using parallel recurrent & convolutional neural networks. Proc ACM KDD workshops 2018;.

[7] Wei W, Zhan G, Wang X, Zhang P, Yan Y. A novel method for automatic heart murmur diagnosis using phonocardiogram. In Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing. 2019; 1–6.

[8] Levin AD, Ragazzi A, Szot SL, Ning T. A machine learning approach to heart murmur detection and classification. In 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2020; 521–525.

[9] Chorba JS, Shapiro AM, Le L, Maidens J, Prince J, Pham S, Kanzawa MM, Barbosa DN, Currie C, Brooks C, et al. Deep learning algorithm for automated cardiac murmur detection via a digital stethoscope platform. Journal of the American Heart Association 2021;10(9):e019905.

[10] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, Perez Alday EA, Sadr N, Sharma A, Mattos S, Coimbra MT, Sameni R, Rad AB, Clifford GD. Heart murmur detection from phonocardiogram recordings: The george b. moody physionet challenge 2022. medRxiv 2022;.

[11] 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.

[12] Breiman L. Random forests. Machine learning 2001; 45(1):5–32.

Address for correspondence:

Hui Lu
BTU Cottbus-Senftenberg
Siemens-Halske-Ring 14
03046, Cottbus, Germany
lu@b-tu.de

Julia Beatriz Yip
Universitätsklinikum Erlangen
Krankenhausstrasse 12
91054, Erlangen, Germany
julia.yip@fau.de