Outcome Prediction and Murmur Detection in Sets of Phonocardiograms by a Deep Learning-Based Ensemble Approach

Sven Festag[1,2], Gideon Stein[3], Tim Büchner[3], Maha Shadaydeh[3], Joachim Denzler[3], Cord Spreckelsen[1,2]

[1] Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Germany
[2] SMITH consortium of the German Medical Informatics Initiative, Germany
[3] Computer Vision Group, Institute for Computer Science, Friedrich Schiller University Jena, Germany

Abstract

We, the team UKJ_FSU, propose a deep learning system for the prediction of congenital heart diseases. Our method is able to predict the clinical outcomes (normal, abnormal) of patients as well as to identify heart murmur (present, absent, unclear) based on phonocardiograms recorded at different auscultation locations. The system we propose is an ensemble of four temporal convolutional networks with identical topologies, each specialized in identifying murmurs and predicting patient outcome from a phonocardiogram taken at one specific auscultation location. Their intermediate outputs are augmented by the manually ascertained patient features such as age group, sex, height and weight. The outputs of the four networks are combined to form a single final decision as demanded by the rules of the George B. Moody PhysioNet Challenge 2022. On the first task of this challenge, the murmur detection, our model reached a weighted accuracy of 0.567 *with respect to the unknown validation set. On the outcome prediction task (second task) the ensemble led to a mean outcome cost of* 10679*. By focusing on the clinical outcome prediction and tuning some of the hyper-parameters only for this task, our model reached a validation score of* 9031*.*

1. Introduction

An early diagnosis of congenital heart diseases is essential for effective treatment and to prevent irreversible effects [1]. Unfortunately, especially in developing countries, providing a diagnosis is often logistically infeasible [2], e.g. due to a lack of medical professionals in combination with large geographical distances. In the interest of improving on this reality, the George B. Moody PhysioNet 2022 challenge [3, 4] calls for an algorithmic prediction of clinical outcomes and the detection of possible heart murmurs based on phonocardiograms. Such a system has the potential to be deployed to perform prescreening of patients, reducing the number of required experts. Machine learning and especially deep learning techniques have proven successful in solving similar tasks, as was shown by the submissions to the 2016 PhysioNet challenge [5]. However, contrary to the PhysioNet 2016 challenge, this year's challenge provides multiple recordings of the same patient from different locations, which opens up the field for a new set of algorithmic approaches. Our team proposes a deep learning ensemble approach. Our method consists of four temporal convolutional networks with identical architectures but each one is exclusively trained on a single recording location. Their intermediate outputs are augmented by manually ascertained patient features and combined according to the rules of the George B. Moody PhysioNet Challenge 2022.

The remainder of the paper is organized as follows. Subsection 1.1 briefly describes the challenge data sets. Section 2 details our method and experimental setup. Finally, Section 3 summarises the results of our method, before the article is ended with some concluding remarks in Section 4.

1.1. PhysioNet Challenge 2022

The challenge organizers provide participants with a public training set consisting of multiple phonocardiograms with a sampling frequency of 4 kHz from 942 patients. Moreover, phonocardiograms of 626 different patients are kept secret by the organizers and used as validation and test data [6]. The auscultation locations at which the recordings were produced are the pulmonary (PA), the aortic (AA), the mitral (MA), the tricuspid (TA) or an unspecified area. Typically, recordings from multiple locations are available for every patient. Additionally, metadata of the patients, such as age, height, weight, sex, etc., are provided. The goal of the challenge is twofold. Firstly,

the developed algorithm should identify whether a heart murmur is present in the phonocardiograms. To generate labeled training data, medical experts analyzed the data of all patients and associated every patient with exactly one of the following classes: "murmur presen", "murmur absent" or "unknown". Secondly, the algorithm needs to predict the clinical outcome ("normal" or "abnormal") of each patient. Again, medical experts were asked to diagnose the patients accordingly and define the correct labeling. For further information regarding the challenge, we refer to [3].

2. Method

The proposed ensemble comprises four networks of identical topology, each specialized in identifying murmurs and predicting patient outcome from a phonocardiogram taken at one specific auscultation location. They are trained on recordings from either the PA, AA, MA, or TA. The motivation behind this division is that most murmurs are best audible at one location, and some murmurs are only audible at a specific location [7]. The networks are trained in a supervised fashion using one expert label for every patient and each task: murmur present, absent, or unknown, and normal or abnormal outcome.

The input to a network consists of two parts, a phonocardiogram recorded at the corresponding location and a vector of the patient's meta-data such as age group (~ 0.5 months, ~ 6 months, ~ 72 months, ~ 180 months, ~ 240 months), sex (1: female, 2: male), height in cm and weight in kg. If one of the features is unknown, it is set to 0.

The phonocardiograms are preprocessed before they are evaluated by the networks. To suppress some of the recorded background noises, a Butterworth low-pass filter of order 6 and cutoff frequency of 400 Hz is applied as most murmurs fall in this frequency range. Afterwards, the series is downsampled to a sampling frequency of 1 kHz, the systolic peak of the first heart cycle is identified, and an interval of 5 seconds starting at this position is cut out. In the last step, the preprocessed window is normalized by a min-max-normalizer.

For the identification of systolic peaks, the following approach is used. At first, another low-pass filter (cutoff frequency 4 Hz) and a min-max-normalizer are applied to obtain the envelope of the rectified signal. Afterwards, the peaks are detected in this low-frequency signal by finding all local maxima that lie at least 0.5 seconds apart from each other. These positions are assumed to mark systolic peaks.

Moreover, a spectrogram summarizing the changes in the frequency domain of the series is computed. To this end, a short-time Fourier transform with a frame length of 0.04 seconds, a frame step of 0.016 seconds, a Fourier transform with a size of 200, and a von-Hann-windowing is applied to the preprocessed five-second interval of the phonocardiogram. After preprocessing, we end up with an audio signal of 5 seconds, a corresponding spectrogram, and a 4-dimensional feature vector.

The topology of one network is presented in Figure 1. At first, the audio signal is processed by two atrous 1D convolutional layers (size: 5, dilation rate: 2, stride: 1, activation: ELU), each followed by max-pooling (size: 2, stride: 2) and batch normalization. The first layer has 64 filters, while the second has 32. Atrous convolution refers to the filtering with kernels that have holes (weights are zero at these positions). It was introduced to reduce the number of computations without reducing the receptive field of a filter [8]. The intermediate result is handed to a normal convolutional layer without holes/dilation (filters: 2, size: 5, stride: 1, activation: ELU) that is again followed by max-pooling (size: 2, stride: 2) and batch normalization. In parallel, the input spectrogram is processed by two 2D convolutional layers (filters: 16 and 8, sizes: 20×2 and 5×5, stride: 1, activation: ELU), two max-pooling layers (size: 4×4, stride: 4) and batch normalizers.
The results of the two previously described convolutional branches are flattened and concatenated into a one-dimensional vector. During training, dropout with a rate of 0.3 is applied to this vector before it is further processed by a dense layer with 512 neurons applying the ELU function. At this stage, the last part of the input, the patient features, comes into play. They are combined with the 512 activations by concatenation. To get a decision for the murmur task, this vector is fed to the murmur head of the network. This head consists of two plain feed-forward layers of sizes 256 (ELU activation) and 3 (linear activation). By applying softmax to the last 3 results, one receives a "probability distribution" $y_{pr}^{(m)}$ over the three murmur classes. For the outcome prediction, the outcome-head is used. It also comprises two dense layers with 30 neurons (ELU activation) or one neuron (sigmoid activation), respectively. The final output is interpreted as the predicted probability of a normal outcome.

For the training, the set of training data needs to be subdivided into four parts, one for each ensemble member. E.g., the subset for the aortic network consists of phonocardiograms (and corresponding patient features) that are recorded in the AA. Phonocardiograms that are marked as "related to a patient with murmurs" but also as "murmur not audible in this location" are discarded from the training sets since the true class is ambiguous. Every set is further split into a training (80%) and a validation (20%) part, where both parts have a similar distribution of murmur labels. During the first phase of each individual training, the convolutional layers and the murmur head are trained based on the murmur labels for a fixed number of epochs. The aim is to minimize the mean of the categorical focal

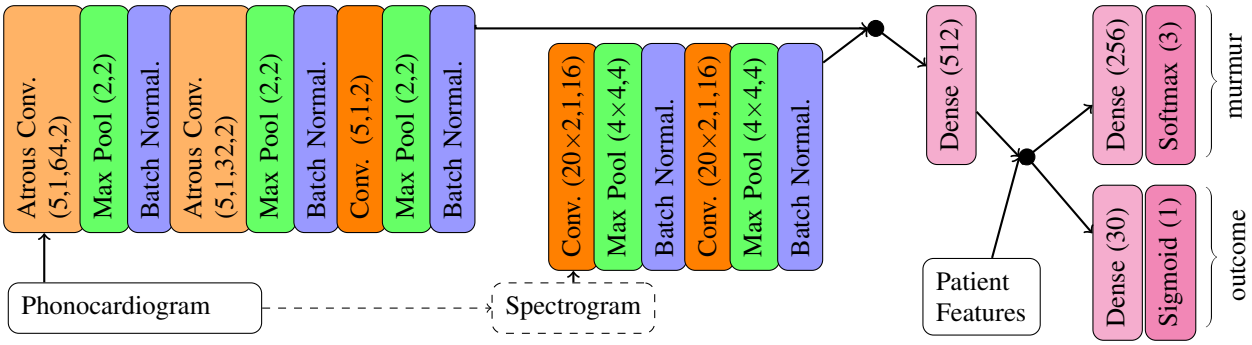

Figure 1. Topology of one ensemble network. The numbers in the blocks define the size, the stride, the number of filters and the dilation rate (if applicable). ELU activation function is used in all layers except for the decision layers.

losses $cfl(y_{tr}, y_{pr}^{(m)})$ [9] with the help of the RMSProp optimizer, where y_{tr} is a one-hot encoding of the true murmur class and

$$cfl(y_{tr}, y_{pr}) = -\alpha \left(1 - (y_{tr} \circ y_{pr})\right)^{\gamma} \log(y_{tr} \circ y_{pr})$$

In this formula, "$\circ$" symbolizes the dot product and α, as well as γ, are hyper-parameters.

Since the challenge score, which is the weighted accuracy (present: 5, unknown: 3, absent: 1) for the murmur detection task, weights the correct classification of murmur occurrences higher than the classification of other classes, we weight the individual cfls similarly during training. After the first training phase, the weights that led to a maximum weighted accuracy on the validation set are restored.

During the second phase, only the outcome head is updated. It is believed that the early layers can already extract and combine useful features from the phonocardiogram, its spectrogram, and the patient data. The weights of the outcome head are updated with respect to the true outcome labels, the predicted ones, and the simple binary cross entropy loss (cfl with $\alpha = 1, \gamma = 0$). Again, an RMSprop optimizer is used for a fixed number of epochs before the weights that lead to the lowest average challenge outcome cost (cf. [3]) on the validation set are restored.

During inference, the outputs of several of the independently trained networks are used. Every recording available for a patient is processed by the corresponding network. Since a murmur might not be audible in all locations, the global murmur decision is set to "present" if for at least one recording the local decision is "present". In all other cases, the probability vectors are averaged to get the two global decisions (murmur and outcome).

2.1. Experiments

In the first experiment, we conducted a five-fold cross-validation on the public training data. The ensemble nets were trained for 500 epochs on the murmur task and additional 100 epochs on the outcome task. The experiment was performed exclusively with data from the public training set. The parameters of the categorical focal loss function were set to $\alpha = 3$ and $\gamma = 5$, the learning rate for the first task to $\eta = 0.0001$ ($\beta = 0.9$) and for the second task to $\eta = 0.001$ ($\beta = 0.9$). During the loss computation for the first task, the errors were weighted individually with respect to the correct label (present: 5, unknown: 3, absent: 1).

For the second experiment, our system was executed by the challenge organizers and trained on the public training set. Afterwards, the network was evaluated against the secret validation set.

Intermediate results suggested that shortening the input window to 1.25 seconds leads to better validation outcome scores while impairing the murmur detection. Hence, we also asked the organizers to evaluate this adapted version against the validation set in a third experiment.

3. Results

Metric	5-CV	Official Valid.
Weighted Accuracy (m)	0.524	0.567
Accuracy (m)	0.574	
F1 (m)	0.401	
AUROC (m)	0.573	
Outcome score (o)	13225	10679
Accuracy (o)	0.568	
F1 (o)	0.486	
AUROC (o)	0.524	

Table 1. Results achieved during the first two experiments. Rows marked with (m) correspond to the murmur detection task. The second column corresponds to the averaged results of the five-fold cross-validation and the third to the evaluation on the secret validation set.

The results achieved during the first two experiments are summarised in Table 1. On the secret validation set, our

model reached a weighted accuracy of 0.567 concerning the murmur detection task, while it led to an average outcome cost of 10679.

During the third experiment, the model with shorter input windows reached a weighted accuracy of 0.398 and an outcome score of 9031 on the validation set.

4. Discussion

In this work, we presented our approach for the PhysioNet 2022 challenge and its preliminary results. Our approach performs better on the outcome prediction task than on the murmur detection task. We assume that the features extracted by our models from phonocardiograms and corresponding spectrograms are especially useful for the clinical outcome prediction and thus, are a broad summary of the cardiac health of patients. So far, we did not perform an extensive hyper-parameter search or added any kind of transfer learning which could increase the performance of our method in the future, especially on the murmur detection task. Despite this, we believe that our approach represents a step toward the automatic detection of congenital heart diseases.

Acknowledgements

The work by SF was supported by the Google Cloud Research Credits program with the award 209639286.

The work by SF and CS was supported by the German Federal Ministry of Education and Research (grant number 01ZZ1803B) in the context of the Smart Medical Information Technology for Healthcare consortium.

References

[1] Silove ED. Assessment and management of congenital heart disease in the newborn by the district paediatrician. Archives of Disease in Childhood Fetal and Neonatal edition 1994; 70(1):F71–F74.

[2] Bernier PL, Stefanescu A, Samoukovic G, Tchervenkov CI. The challenge of congenital heart disease worldwide: epidemiologic and demographic facts. Seminars in Thoracic and Cardiovascular Surgery Pediatric Cardiac Surgery Annual 2010;13(1):26–34.

[3] 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, August 2022.

[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation June 2000;101(23).

[5] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AEW, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37(12):2181–2213.

[6] 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 June 2022;26(6):2524–2535.

[7] Alpert MA. Systolic Murmurs. In Walker HK, Hall WD, Hurst JW (eds.), Clinical Methods: The History, Physical, and Laboratory Examinations, 3rd edition. Boston: Butterworths, 1990; ch. 26.

[8] Giusti A, Cireşan DC, Masci J, Gambardella LM, Schmidhuber J. Fast image scanning with deep max-pooling convolutional neural networks. In 2013 IEEE International Conference on Image Processing. September 2013; 4034–4038.

[9] Lin T, Goyal P, Girshick RB, He K, Dollár P. Focal loss for dense object detection. CoRR 2017;abs/1708.02002. URL `http://arxiv.org/abs/1708.02002`.

Address for correspondence:

Sven Festag
Institute of Medical Statistics, Computer and Data Sciences,
Jena University Hospital
Bachstraße 18, 07743 Jena, Germany
sven.festag@med.uni-jena.de