An LSTM-based Listener for Early Detection of Heart Disease

Philip Gemke[1,2], Nicolai Spicher[3], Tim Kacprowski[1,2]

[1]Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany
[2]Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Braunschweig, Germany
[3]Department of Medical Informatics, University Medical Center Göttingen, Germany

Abstract

As a contribution to the George B. Moody PhysioNet Challenge 2022 we (team listNto_urHeart) propose a phonocardiogram classifier. Based on the assumption that these recordings bear similarity to music, we borrow methods from the field of computational music analysis. In contrast to end-to-end machine learning approaches, we propose a carefully-crafted processing pipeline for automatically detecting single heartbeats in phonocardiogram recordings which are then classified by a bi-directional long short-term memory network. Our approach has the advantage of not requiring manual annotations during training, therefore alleviating the lack of annotated training data. In murmur detection, we reached a weighted accuracy of 0.68 *in validation,* 0.668 *in test (rank:* $25/40$*) and* 0.64 ± 0.08 *during training. In predicting patient outcome, we reached* $10,362$ *in validation,* $13,866$ *in test (rank:* $27/39$*) and* $11,386 \pm 2,108$ *during training. The results indicate that borrowing algorithms from computational music analysis could bear the potential to address challenges in phonocardiogram classification successfully.*

1. Introduction

Congenital heart disease (CHD) and valvular heart disease (VHD) can be identified early by abnormal heart sounds. Early diagnosis can avoid medical complications due to disease progression and the financial burden of more expensive treatments. As part of the George B. Moody PhysioNet Challenge 2022 [1], we (team *listNto_urHeart*) tackle the problem of algorithmic prescreening of phonocardiograms (PCG) to detect CHD and VHD in the CirCor DigiScope dataset [2].

Physiological heart cycles show two distinct sounds: The *S1* sound resulting from the atrioventricular valves closing and *S2* sound resulting from the semilunar valves closing either together or the aortic valve closing before the pulmonary valve. Especially the S2 sound results in a complex morphology, showing one or two peaks depending on the patients physiology. In pathologic heart cycles, additional peaks, i.e. heart sounds, can appear. The automatic segmentation of the different sounds has a poor performance which makes annotated segmentation for every patient by doctors the gold standard [3]. However, this annotation is time-consuming and expensive.

To overcome the limited data availability for training machine learning (ML) models, we propose an fully automatic PCG segmentation method by borrowing techniques from the field of computational music analysis. In contrast to end-to-end ML approaches in which the ML model learns all the steps between input and final output [4], we propose a carefully-adjusted processing pipeline before feeding the signals to a long short-term memory (LSTM) network (see Fig. 1). We hypothesize that heart sounds, measured over a sequence of multiple cardiac cycles, show a base rhythm in the same sense as musical tracks do. Moreover, similar to different instruments, PCG data are composed of multiple sound sources besides the heart sounds, e.g. breathing of the subject or speaking of other persons in the room.

2. Material and Methods

The challenge provides a training data set consisting of $3,163$ recordings measured at different locations from 942 patients. During training, we select the recordings labeled as *best audible location* for each patient. For patients without this label, the longest recording was selected for training. Regarding predictions on the hidden test and validation sets, we did not take location into account.

In order to enable others to reproduce our results, we provide complete source code online[1].

[1]https://github.com/PhilipGemke/listNto_urHeart_Entry3.2.git

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.151

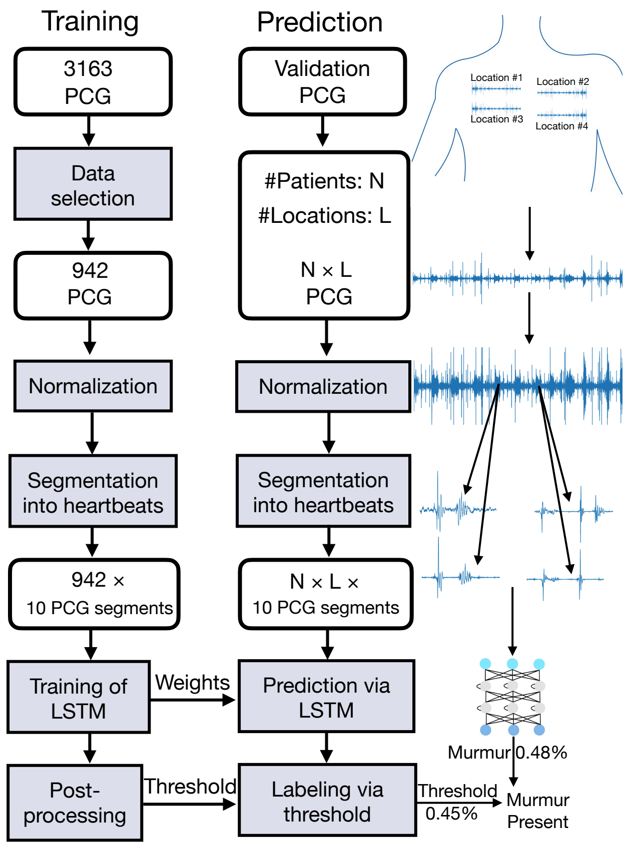

Figure 1: Schema of processing pipeline

2.1. Preprocessing

Although PCG is a proven diagnostic tool, a typical issue in the analysis is a critically low signal-to-noise ratio (SNR), especially in uncontrolled environments [5]. Therefore our preprocessing strategy includes filtering recordings to reduce noise (sec. 2.1.1), splitting the signal into different segments containing single heartbeats and selecting only the 10 segments with highest SNR (sec. 2.1.2).

This approach is based on the assumption that murmurs and their corresponding pathologies are detectable in every heartbeat of a PCG, since the CirCor DigiScope dataset focuses on systolic (S1) murmurs (96.7%). Only in diastolic (S2) pathologies, the inspiratory split affects the quality of heart sounds which might lead to the pathology not being detectable in a single segment [2].

2.1.1. Normalization

PCG recordings were normalized using volume-based and percussion-based approaches depicted in Fig. 2. **Volume-based normalization:** We hypothesized that physiological differences such as the size of chest wall, the amount of fat and muscle tissue, or the size of the heart and vessels, determining the resonance space of the heart sound, lead to different amplitudes in the PCG recordings. Furthermore, issues during measurement, such as changing pressure on the stethoscope influence the volume and frequency of the recording directly.

Therefore, we normalized all recordings to -20dBFS (decibel relative to full scale) using the method `apply_gain()` provided by the open-source library pydub[2]. We estimated a threshold based on the root mean square RMS $= \sqrt{A^2/2} = 0.707 * A$. Here, A represents the amplitude of a sine wave fitted to the PCG signal as this approach is more robust to outliers than direct amplitude estimation [6]. The RMS value is then used to increase the amplitude of all samples which are below -20dBFS to this level and to decrease all other samples.

Percussion-based normalization: We hypothesized that PCG data are composed of multiple sound sources similar to different instruments in an audio track, therefore we used percussion filters which are commonly used in sound-engineering to separate vocals from drums.

Hence, we used an audio filter called `harmonic percussive separation` provided by the open source library librosa[3] to separate harmonic and percussive parts. Subsequently, a short-time Fourier transformation was performed using `stft()` to obtain a power spectrogram. This was used as an input to a median-filtering harmonic percussive separation using `decompose()` (parameters: *sr=4000; hop_length=32; n_fft=128; win_length=128*) followed by `istft()`, converting the PCG power spectrograms back to time domain.

2.1.2. Segmentation into Heartbeats

After normalization, PCG recordings were processed by `beat.beat_track` *(tightness=128, units=samples, trim=False)* provided by librosa to obtain indices of peaks. The tightness is an option that allows some irregularity of detected beats to enable peak detection in recordings with physiological irregularities such as extra systoles. The indices were used to extract short segments from the whole recordings by placing windows of 600ms duration on each index and storing samples covered by the window.

We devised a two-stage procedure to select the segments representing single heartbeats with highest SNR for a single patient. First, single segments with standard deviation (SD) larger than the mean of all SD values were removed to exclude segments showing high amplitude noise stemming from crying or talking (see Fig. 3). Second, for each remaining segment we computed Approximate Entropy (ApEn) using antropy library[4]as a measure of regularity. We expect segments with high SNR to show repeti-

[2]https://pydub.com/
[3]https://librosa.org/
[4]https://pypi.org/project/antropy/

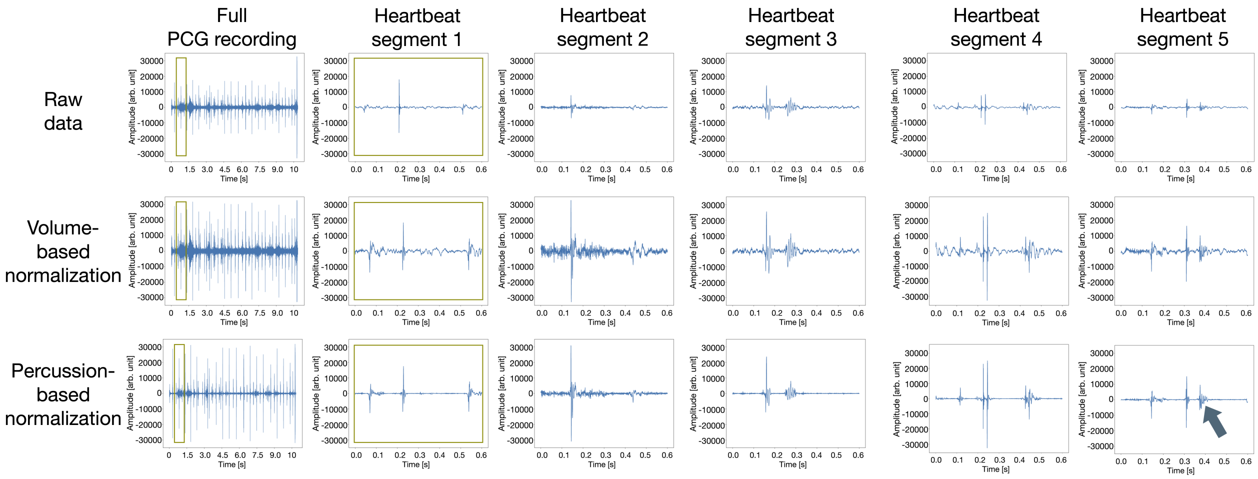

Figure 2: In the first row a whole PCG recording was processed. In the second row a segment from this recording is shown. The other rows show segments extracted from different patients. The arrow in the last row points to a heart sound which was made more pronounced due to the percussion normalization by removing surrounding noise.

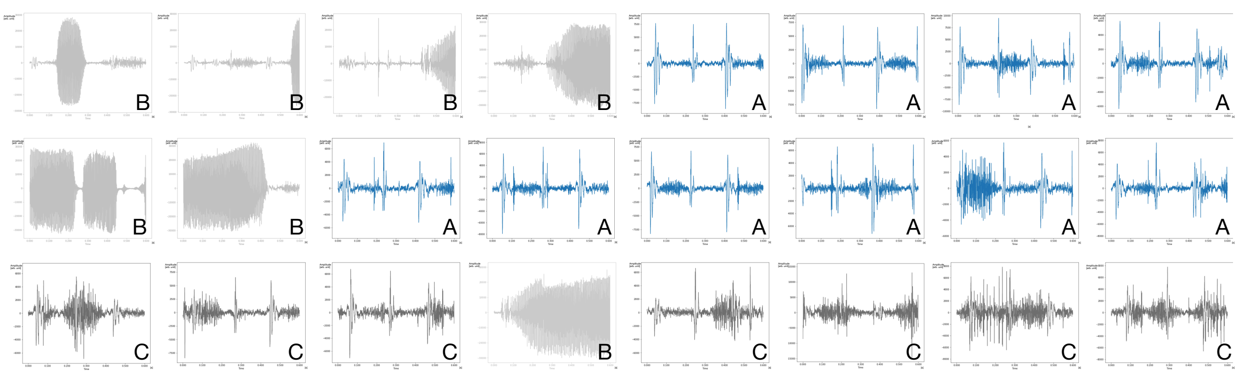

Figure 3: Selection of PCG segments: After volume-normalization (sec. 2.1.1), peaks in the PCG signals resulting from heart sounds were detected using a bpm-song-classifier (sec. 2.1.2). 600 ms segments were extracted with the 10 most suitable being selected using a two-stage procedure, neglecting segments with too high ApEn or SD values. The shown segments stem from a single recording and are sorted ascending w.r.t. ApEn values. **A:** Segments kept for further analysis **B:** Segments excluded based on SD **C:** Segments excluded based on ApEn value.

tive patterns of signal fluctuations due to physiological effects while in segments with low SNR random patterns appear. ApEn has already been applied successfully to audio signals, e.g. for speech quality measurement [7].

We kept the 10 segments showing the lowest ApEn values for each recording. In rare cases ($\approx$ 3%) with less then 10 detected segments, remaining segments were duplicated.

2.2. Classification

The 10 extracted audio segments were used for training of a bi-directional LSTM network. The network was built using the open source libraries keras[5] and tensorflow[6]

The bi-directional LSTM architecture was chosen for its ability to remember sequences from both directions. As both tasks of the challenge (*murmur* detection and *outcome* prediction) have different underlying ground truth, we designed a separate LSTM model for each. During testing, for each patient, we took the 10 segments extracted during preprocessing and predicted murmur and outcome for every segment. As there are up to 5 different measurement locations for each patient, this results in $10 - 50$ probabilities per task.To obtain only one probability value

[5]https://keras.io/
[6]https://www.tensorflow.org/

Table 1: Weighted accuracy and cost metric scores (official Challenge score) for our best entry (*team listNto_urHeart)*

	Training	Validation	Test	Ranking
Weighted accuracy	0.64 ± 0.08	0.685	0.668	25/40
Cost [$]	$11,386$ $\pm 2,108$	10,362	13,866	27/39

per patient, we averaged the probabilities, resulting in a mean probability. For the determination of a classification threshold for both tasks, we used the training data and performed an iterative minima (maxima) search over the costs (accuracy) in the interval $]0,1[$ in steps of 0.1.

3. Results

Preprocessing and segmentation extraction resulted in segments containing heartbeats with similar properties in individual patients. Fig. 2 shows clearly that the normalization adjusted segments to similar amplitude levels. As can be seen in Fig. 3, segments containing high levels of noise were removed. Furthermore, the percussion-filter was able to remove the percussive elements of the heart sounds (e.g. arrow in Fig. 2).

Tbl. 1 shows the results of the official phase and cross validation after classification, yielding a 25th place in the accuracy task.

4. Discussion and Conclusion

In this work, we developed a method for classification of PCG recordings based on the principle of segmenting the full recordings into single heartbeats which were then fed to an LSTM network. The final prediction was made by averaging the LSTM output for each heartbeat and comparing the resulting value to a threshold.

The provided PCG data was recorded in a clinical environment and thereby suffers from noise from various sources, resulting in low SNR. This makes it hard to automatically detect heartbeats in a PCG recording, which is why many state-of-the-art approaches use annotated heartsounds, manually picked by cardiologists [8]. In contrast, we proposed a method for fully automatic segmentation, which could help to overcome the lack in databases.

Our methods are based on the idea of interpreting PCG recordings as multi-instrumental audio tracks. In computational music analysis the rhythm of a song is calculated, sounds are decomposed into different components, and tracks are classified into genres [9]. This corresponds to the tasks associated with PCG classification. Additionally, the rhythmic basis of the heartbeat and the signal decomposition are important features for the heart sound analysis [10]. The achieved results indicate that borrowing algorithms from computational music analysis could bear the potential to address challenges in PCG processing successfully. Similar approaches were also applied to ECG [11].

In conclusion, our results indicate the potential of a manually-crafted preprocessing pipeline using audio processing techniques with few heuristic parameters for PCG classification in contrast to end-to-end approaches.

References

[1] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, et al. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv 2022;URL https://doi.org/10.1101/2022.08.11.22278688.

[2] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C, et al. The CirCor DigiScope dataset: from murmur detection to murmur classification. IEEE Journal of Biomedical and Health Informatics 2021;26(6):2524–2535.

[3] Chen Y, Sun Y, Lv J, Jia B, Huang X. End-to-end heart sound segmentation using deep convolutional recurrent network. Complex Intelligent Systems 2021;7(4):2103–2117.

[4] Gjoreski M, Gradišek A, Budna B, Gams M, Poglajen G. Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. IEEE Access 2020;8:20313–20324.

[5] Leal A, Nunes D. Noise detection in phonocardiograms by exploring similarities in spectral features. Biomedical Signal Processing and Control July 2018;44:154–167. ISSN 17468094.

[6] Fowler CA. Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. The Journal of the Acoustical Society of America 1990; 88(3):1236–1249.

[7] Metzger RA, Doherty JF, Jenkins DM. Using approximate entropy as a speech quality measure for a speaker recognition system. In 2016 Annual Conference on Information Science and Systems (CISS). IEEE, 2016; 292–297.

[8] Zeinali Y, Niaki STA. Heart sound classification using signal processing and machine learning algorithms. Machine Learning with Applications 2022;7:100206.

[9] Poonia S, Verma C, Malik N. Music genre classification using machine learning: A comparative study. IITM Journal of Management and IT 2022;13(1):15–21.

[10] Arvin F, Doraisamy S, Safar Khorasani E. Frequency shifting approach towards textual transcription of heartbeat sounds. Biological procedures online 2011;13(1):1–7.

[11] Idrobo-Ávila E, Loaiza-Correa H. Can the application of certain music information retrieval methods contribute to the machine learning classification of electrocardiographic signals? Heliyon 2021;7(2):e06257. ISSN 2405-8440.

Address for correspondence:

t.kacprowski@tu-braunschweig.de