Heart Murmur Detection in Phonocardiographic Signals Using Breathing Noise Suppression

Kristóf Müller[1], Márton Áron Goda[1]

[1] Pázmány Péter Catholic University Faculty of Information Technology and Bionics, Budapest, Hungary

Abstract

The automated heart murmur detection has become increasingly relevant to aid the medical diagnosis. The George B. Moody PhysioNet Challenge 2022 proposed a similar problem to encourage research in this field, by providing a manually labeled dataset of 942 patients.

Our approach focuses on an accurate classification method using segmented S1, S2, systole, and diastole sections. A custom segmentation process was developed for detecting the heart sounds, then with further refinement the systole and diastole regions were also determined. This method is based on wavelet decomposition and Shannon entropy (WT-SE). Along with this algorithm, a Hidden Semi-Markov Model (HSMM) was trained for comparing the results. Multiple types of features were obtained for each segment, such as time domain, frequency domain, and other features (e.g. Lyapunov exponent). Dimensionality reduction on the extracted feature-space was performed with two approaches, principal component analysis (PCA) and analysis of variance (ANOVA). For classification a support vector machine (SVM) and random forest (RF) was tested.

The recordings also contained breathing noises in several cases, these signals could cause problems in the segmentation stage and could be misclassified. To mitigate these problems, a breathing noise suppression phase was implemented based on the periodicity of the detected event. Locally a 10-fold cross-validation was performed and the best scoring one was submitted for evaluation.

1. Introduction

In the 2022 George B. Moody PhysioNet challenge a difficult task was presented, to classify heart murmurs and clinical outcomes of patients based on only phonocardiographical (PCG) data. The teams were provided with over 3500 segmented recordings from 942 patients, labeled for containing murmurs and clinical outcomes of the patients [1, 2]. This was similar to the 2016 PhysioNet challenge where only normal and abnormal classification had to be performed. For this previous challenge multiple entries received an accuracy score above 0.80, one example being by Goda and Hajas [3]. Here the authors utilized most notably frequency and wavelet decomposition based features for classification with a support vector machine (SVM). This work was notable since it achieved a relatively high score without using any neural-networks (NN). Another high scoring submission which did not employ NNs was by Homsi *et al.*, in which the authors implemented an ensemble classifier with extracted features from the segmented signal [4]. This model used a random forest (RF) with LogitBoost, and a cost-sensitive classifier (CSC). These results inspired us to work with more conventional machine learning methods, although NNs were the most popular machine learning model in this past challenge.

2. Methods and materials

Our process consisted of five major stages, preprocessing, segmentation, feature extraction, dimensionality reduction, and classification as seen in Figure 1.

2.1. Preprocessing

All recordings were resampled to 1000 Hz, to reduce the amount of memory used by the following steps. The first and last second of each signal was rejected, since in most cases these regions contained noise bursts coming from the recording equipment. After this a wavelet-based denoising was performed with the *coif5 family* and SURE algorithm, based on the results of [5]. At the breathing suppression phase the fact that these noises contain higher frequencies and they appear with a different periodicity (0.5-1 Hz) compared to heart sounds (1-2 Hz) was utilized. First a nonlinear dynamic range expansion was performed to accentuate the breathing and the heart sounds with the `expander` Matlab function (threshold -25 dB, ratio 7, attack 100 ms, hold 300 ms, release 500 ms). Figure 2b shows a result from this expansion. The expanded signal

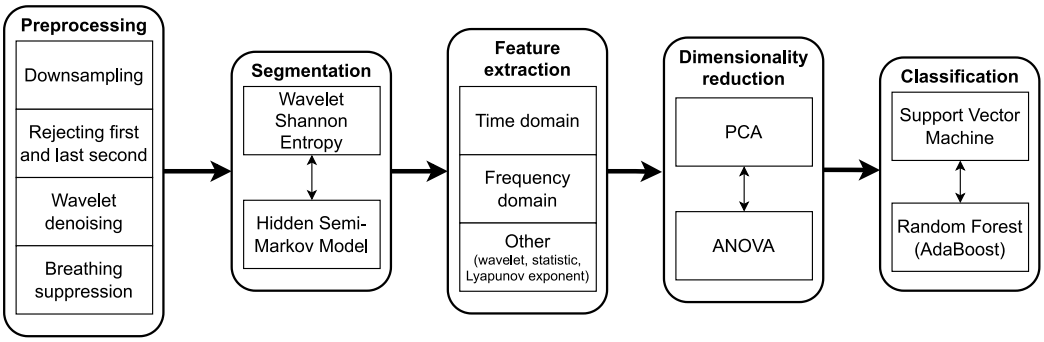

Figure 1: The main stages of our process with their respective steps. Interchangeable steps are shown with a double arrow

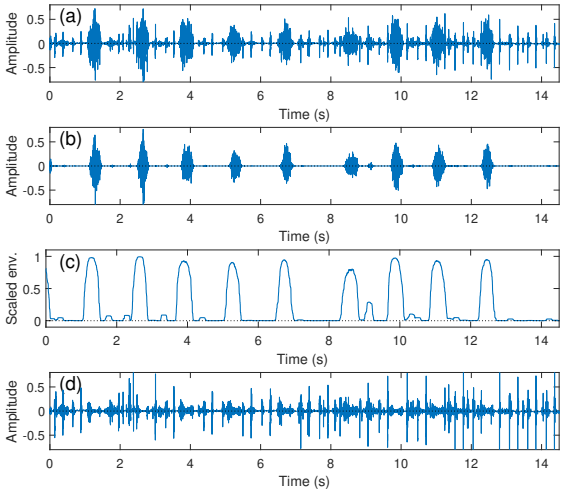

Figure 2: Main stages in breathing suppression. (a) Signal with breathing noises (b) Dynamic range expanded signal (c) Envelope after nonlinear scaling (d) Signal with reduced breathing noises

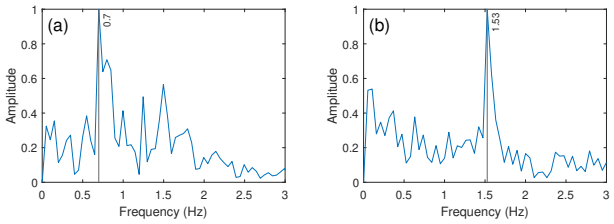

Figure 3: Comparison between the envelope spectra, with their maxima labelled. (a) Signal with breathing noises (b) Signal without breathing noises

was filtered with a 10th order high-pass Butterworth filter at 200 Hz. Next the root mean square (RMS) envelope was calculated and scaled to have an amplitude close to the original signal values. This result can be seen in Figure 2c. After subtracting its mean the Fourier transform of the envelope was obtained, to find the rough periodicity estimate of the high frequency events. In case the measurement was short enough not to contain a breathing event, or the respiratory sounds was too subtle, a higher overall periodicity was observable on the spectrum, as shown in Figure 3. This way noisy signals could be selected, and the suppression could be performed. The scaled envelope values were used as an amplitude reduction factor, meaning at higher values the amplitude of the original signal was lowered corresponding to the envelope value. Then finally a dynamic range compression was performed on the amplitude reduced signal, so that the incidental murmurs could be enhanced without the breathing noises. The compression was done with the `compressor` Matlab function (threshold -25 dB, ratio 5, attack 100 ms, release 500 ms). The final result is shown in Figure 2d. This suppression method was most useful as a preprocessing step for the various segmentation methods but this process was also performed on the original signals before feature extraction, although with a high pass filter set to 250 Hz instead.

2.2. Segmentation

The preprocessed signal was then decomposed with *db6* wavelets and the fourth level was chosen for further processing. The Shannon entropy (SE) envelope was calculated and smoothed out with a moving mean filter with a 60 ms window [6] on this decomposition level. Next, the instantaneous phase of this envelope was computed by using its Hilbert transform, and positive slope zero-crossings were considered as a sound event. Then, a predefined time interval centered around each of these events was taken, and the cetroid of their RMS was labeled as the timestamps of the events. After this refinement, certain events moved closer to each other enough to be considered the same event. With a minimal time difference threshold these dense labels were filtered out. The main steps of this process are illustrated in Figure 4, along with the

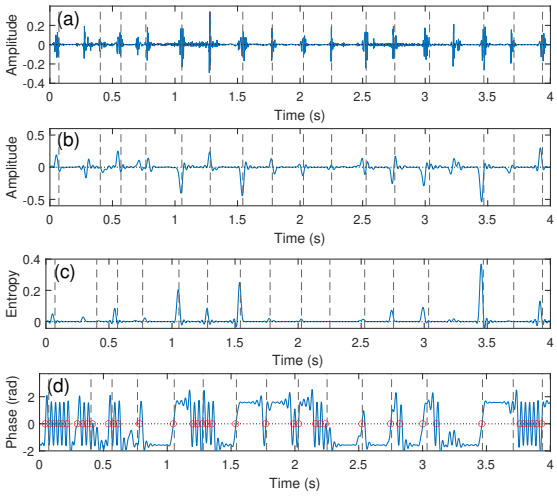

Figure 4: Shannon entropy based segmentation, dashed lines mark the detected heart sounds. (a) Input signal (b) Fourth wavelet detail (c) Shannon entropy (d) Instantaneous phase, red circles mark the positive slope zero-crossings

detected heart sounds. Using this method the S1 and S2 sounds could be detected with acceptable accuracy. However, the given sounds were not differentiated, and in order to categorize them, the systole and diastole intervals of the heart cycle were localized. For this a simple length based rule was implemented, because on average the length of a systole is shorter than that of the diastole. The next step consisted of finding the start and end of each event, so that these intervals could be segmented from the original signal. This was achieved by finding the left and right first locations where the RMS value dropped below a set percentage (60%) of the one at the centroid. This process was used for the heart sounds and the sections between them were considered to be systole and diastole regions. For further accuracy the length of these inter-sound regions were kept consistent by an outlier removal and a maximal and minimal accepted length. A Hidden Semi-Markov Model (HSMM) was also trained to segment the recordings based on Springer's implementation [7]. These results were then compared with the previously described wavelet and Shannon entropy (WT-SE) method.

2.3. Feature extraction

For each patient only the recordings with known locations were used, meaning the AV, MV, PV, and TV labeled recordings. Each recording was processed the same way, after preprocessing and segmentation S1, systole, S2, and diastole sections were separated and the same features were calculated for each heart cycle segment along with certain features for the entire signal. The features were averaged along the heart cycle segments and their standard deviations were also included. The final feature vector for each patient consisted of different statistical measures between these averages and standard deviations for each recording. These measures include: minimum, maximum, mean, median, variance, and standard deviation. The time domain features consisted of the length of the segment, the root mean square (RMS) value, kurtosis, zero cross rate, and heart rate variability (HRV). For HRV the RMS of successive differences method (RMSSD) was used. This feature was calculated only by using the onset times of the S1 sounds. All other calculations were done the conventional way. In the frequency domain two intervals in the spectrum of the segment was selected, one between 5-30 Hz for low frequency content and another between 45-70 Hz for high frequencies. In these regions the frequency with the largest amplitude was obtained, along with the standard deviation at that point. A different partitioning was chosen for low and high frequency energy calculation, being a simple threshold at 35 Hz. The energy ratio of these sections to the total amount of energy was obtained. Spectral kurtosis was calculated from the kurtosis of the frequency spectrum of the segment. Other features include the wavelet energy and Lyapunov exponent of the segment [8, 9]. From the wavelet decomposition of the signal with the *db6 family*, the first detail coefficients were chosen and the energy was calculated. Lyapunov exponent was calculated with a time delay of 5 ms and an embedding dimension of 3, for both the original signal and the first wavelet detail. The final set of extracted features used the entire recording for calculation. These included the kurtosis, three frequency energy ratios (24-144 Hz, 144-200 Hz, 200-500 Hz), and the maximum peak of the autocorrelation between 0.3 s and 2.5 s based on [10]. All other available biometrical data was also included.

2.4. Classification

Principal component analysis (PCA) and feature selection based on analysis of variance (ANOVA) was used separately as dimensionality reduction methods. With PCA the threshold of explained variance was set to 99%, while with ANOVA based feature selection was set to include the 110 best scoring features. For classification two main methods were considered: support vector machines (SVM) and random forests (RF) with AdaBoost, both have shown good performance previously [3, 4]. While training the models custom misclassification costs were set. In murmur classification, misclassified "Present" cases were set to a cost of 5 while misclassified "Unknown" cases were set to 3. In clinical outcome classification, misclassified "Abnormal" cases had a cost of 2. The SVM used standardized features and a linear kernel, with the box constraint parameter set to 1. The boosted RF had a learning

Classifier	PCA			ANOVA		
	HSMM	WT-SE	GT	HSMM	WT-SE	GT
SVM	0.38	0.46	0.38	0.66	0.66	0.59
RF	0.47	0.46	0.47	0.64	0.61	0.58

Table 1: Murmur classifier validation scores

Classifier	PCA			ANOVA		
	HSMM	WT-SE	GT	HSMM	WT-SE	GT
SVM	13065	14714	12548	12404	12890	12740
RF	13690	13884	14193	12284	11759	11498

Table 2: Outcome classifier validation scores

rate of 0.1 and 60 learners in its ensemble, with maximum branching set to 20.

3. Results

For local results 10-fold cross-validation was used to get an estimate of the performance of the given model. For the murmur classifier the weighted accuracy, and for clinical outcome classifier the mean cost score was used. The resulting validation scores based on these measures are shown in Table 1 and Table 2 comparing the different methods tested for segmentation (HSMM, WT-SE) with the ground truth (GT), dimensionality reduction (PCA, ANOVA), and classification (SVM, RF). Our final submission received a 0.574 weighted accuracy and 9178 mean cost score. However, the murmur classifier in a previous submission achieved a score of 0.696, this is shown in Table 3. The resulting drop in accuracy could be caused by the later implemented features which, while increased the performance in the outcome prediction, reduced the overall ability for more accurate murmur classification.

4. Conclusion and discussion

Based on these results, an SVM classifier was used in murmur classification and RF was used in clinical outcome prediction, both with HSMM segmentation and ANOVA feature selection. Interestingly, in most cases both tested segmentation methods outperformed the ground truth segmentation in terms of classification. This is most likely due to that not all heart cycles were labeled manually in a recording and that resulted in slightly biased features. The WT-SE segmentation based on our local results could perform comparably to the HSMM method. The disadvantage being that it produced less accurate segments in a lower amount, but this could be optimized by choosing its appropriate parameters or implementing a parameter estimation stage. While its obvious advantages come in that it does not need training, it is more robust to breathing artifacts, and the segmentation step requires less time to perform.

	Murmur	Outcome
Final submission	0.574	9178
Best scores	0.696	9178

Table 3: Scores received during official phase. The best scores come from two different submissions

References

[1] 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.

[2] 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;.

[3] Goda MA, Hajas P. Morphological determination of pathological pcg signals by time and frequency domain analysis. In 2016 Computing in Cardiology Conference (CinC). 2016; 1133–1136.

[4] Homsi MN, Medina N, Hernandez M, Quintero N, Perpiñan G, Quintana A, Warrick P. Automatic heart sound recording classification using a nested set of ensemble algorithms. In 2016 Computing in Cardiology Conference (CinC). 2016; 817–820.

[5] Gradolewski D, Redlarski G. Wavelet-based denoising method for real phonocardiography signal recorded by mobile devices in noisy environment. Computers in Biology and Medicine 9 2014;52:119–129. ISSN 0010-4825.

[6] Nivitha Varghees V, Ramachandran KI. Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope. IEEE Sensors Journal 2017; 17(12):3861–3872.

[7] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 2016;63(4):822–832.

[8] Delgado-Trejos E, Quiceno-Manrique AF, Godino-Llorente JI, Blanco-Velasco M, Castellanos-Dominguez G. Digital auscultation analysis for heart murmur detection. Annals of Biomedical Engineering Feb 2009;37(2):337–353. ISSN 1573-9686.

[9] Kumar D, Carvalho P, Couceiro R, Antunes M, Paiva R, Henriques J. Heart murmur classification using complexity signatures. In 2010 20th International Conference on Pattern Recognition. 2010; 2564–2567.

[10] Tang H, Wang M, Hu Y, Guo B, Li T. Automated signal quality assessment for heart sound signal by novel features and evaluation in open public datasets. BioMed Research International 2021;2021. ISSN 23146141.