

Machine Learning Based Identification of Pathological Heart Sounds

Tanmay Gokhale

Duke University, Durham, USA

Abstract

Automated interpretation of heart sounds holds great promise in increasing the diagnostic accuracy and consistency of cardiac auscultation and allowing for use in remote, tele-health settings. However, existing algorithms for classification of heart sounds have been constrained by limited idealized training sets and methodological issues with validation. As part of the 2016 PhysioNet Challenge competition, we present an algorithm for automated heart sound classification that uses Hilbert-envelope and wavelet features to attempt to capture the qualities of the heart sounds that physicians are trained to interpret. We perform a two-step classification of heart sounds into poor quality, normal or abnormal with sensitivity of 0.7958 and specificity of 0.7459.

1. Introduction

Auscultation of heart sounds is a critical component of the physical exam and can lead to the identification of serious medical conditions. During the physical examination, auscultation allows the physician to gain some insight into the inner workings of the cardiac function without the use of any more complex diagnostic technologies. However, identification of pathological heart sounds by ear is challenging in the most ideal environments, and become exponentially more difficult with ambient noise and other sounds, making the automated classification of heart sounds a powerful tool.

1.1. Normal cardiac heart sounds

The sounds heard during cardiac auscultation are caused by the flow of blood within the heart as the cardiac valves open and shut. The normal cycle of heart sounds consists of two heart sounds (S1 and S2) separated by periods of relatively silence. The first heart sound (S1) is caused by the closure of the mitral and tricuspid valves at the start of cardiac systole (Figure 1). As the heart begins isovolumetric contraction, the increase in ventricular pressure causes the atrioventricular valves to close, preventing backflow of blood into the atria.

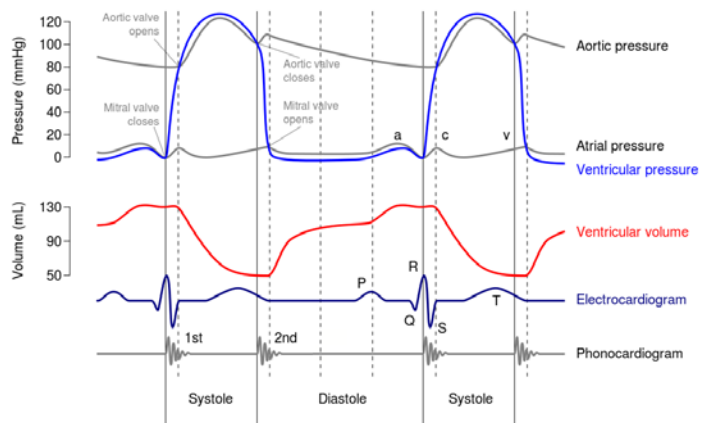


Figure 1. Wiggers Diagram showing cardiac pressures, volume and electrical activity with corresponding phonocardiogram. Modified from figure by Daniel Chang MD, released under CC BY-SA license.

No audible sounds are normally heard during the bulk of the systolic phase where the ventricles eject blood into aorta and pulmonary artery. The second heart sound (S2) occurs due to the closing of the aortic and pulmonary valves at the start of diastole. After the heart has completed ejection, isovolumetric relaxation of the ventricles leads to a pressure gradient between the aorta/pulmonary artery and the ventricle, leading to closing of the semilunar valves.

As heart sounds are caused by abrupt changes in blood flow, opening of the heart valves typically does not produce an observable sound, while closing of the valves through which blood is actively flowing is clearly audible.

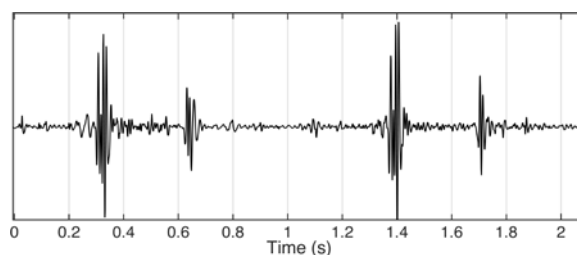


Figure 2. Normal phonocardiogram from two cardiac cycles. Data from recording training-a/a0011

1.2. Pathological heart sounds

A number of cardiac abnormalities can lead to changes in the observed heart sounds. These changes are typically categorized into gallops, murmurs and rubs, each of which have different causes and different physiological implications.

Gallops are extra heart sounds that result in a “galloping” rhythm. The third heart sound (S3) can be observed during early diastole, after the S2. It is often benign in the young and during pregnancy, but in others, especially in the elderly, it is a pathological sign commonly associated with reduced systolic function. While the mechanistic origin of the S3 is not certain it is believed to be due to excess blood volume in the ventricle. The fourth heart sound (S4) occurs during late diastole, shortly before S1, and coincides with atrial systole. Atrial contraction rapidly forces blood into the ventricle, and in the presence of a stiff ventricle caused by diastolic dysfunction, an audible S4 is appreciable.

Cardiac murmurs are caused by the turbulent flow of blood through a diseased valve. Incompetency or insufficiency of a valve can lead to a backflow of blood across the valve, which may produce an audible murmur. In addition, stenosis or narrowing of a valve can occur due to thickening of the valve leaflets. This stenosis leads to an obstruction of flow across the valve, resulting in a murmur. Murmurs are characterized by when they are audible (systolic or diastolic), which depends on the location and type of valvular disease, and by the quality of the audible murmur.

Rubs are a “scratching” like sound caused by friction between layers of the inflamed pericardium in the setting of pericarditis.

1.3. Automated heart sound classification

Because changes in the heart sounds are often very subtle, the automated interpretation and classification of heart sounds has been of interest for many years. Many groups have developed algorithms for this purpose; early work focused on relatively simple thresholding techniques while recent advances in statistical learning have allowed for the application of neural networks and other machine learning methods to this problem [1].

Many existing algorithms appear to have been quite effective at the problem of heart sound classification. However, while these methods excel at interpreting recordings made in idealized settings, many fail to perform with high accuracy when used with ‘dirty’ recordings from realistic clinical settings, or recordings made with different types of recording equipment. In addition, some algorithms have been trained and subsequently validated on the same data set, meaning that high levels of overfitting could be occurring in seemingly effective models of

classification.

In this manuscript, we describe an approach to classify heart sounds using features inspired by the way physicians are trained to interpret heart sounds. We use an extensive previously described data set of heart sound recordings from a variety of environments, perform feature extraction, and train a machine learning algorithm to classify recordings as normal or pathological.

2. Methods

2.1. Data source

Heart sound recordings were obtained from the PhysioNet database, as previously described [1]. Briefly, the database includes 4,430 recordings taken from 1,072 healthy and cardiac disease patients in both clinical and non-clinical environments.

2.2. Base signal features

The PhysioNet Challenge sample entry provided a framework for further feature extraction and classification. The duration dependent logistic regression-based Hidden Markov Model [2, 3], trained using the “a” training set, was used to assign a cardiac cycle state to each segment of the phonocardiogram (PCG) recordings. 20 pre-defined features were extracted based on PCG values and the cardiac cycle state data. These included durations of each cardiac state (S1, S2, systole and diastole), ratios of each cardiac state duration, and ratios of mean signal amplitude in each cardiac state.

2.3. Additional feature extraction

Additional features were selected to attempt to mimic the process via which a physician interprets heart sounds. Physician heart sound interpretation focuses on timing, frequency and intensity of the audible sounds. While the base feature set captures information about state duration and mean amplitude, it does not capture the timing of audible sounds during the state, the quality of the sound or the intensity of individual sounds.

To capture the overall intensity of heart sounds, the PCG signal was segmented based on cardiac cycle state and the Hilbert envelope during each state was found. The maximum value of the Hilbert envelope, and the area under the Hilbert envelope were used as metrics of overall signal intensity. Sound shape is also an important factor in identifying pathological sounds – for example recording with a “crescendo-decrescendo” murmur could have the same mean amplitude and area under the envelope as a low-intensity background noise. To capture this, the area above the Hilbert envelope and below the maximum envelope value was quantified during each segment

(Figure 3).

In order to capture the timing of sounds within each stage as well as the quality of the sound, each state segment was further divided into four equal-duration sub-segments, and the continuous wavelet transform of the signal during each sub-segment using the Coifman fifth order wavelet with integer scales from 1 to 32 was obtained. The mean value of the transform at each scale was included as a feature for each of the sixteen sub-segments (four cardiac cycle states times four sub-segments per state).

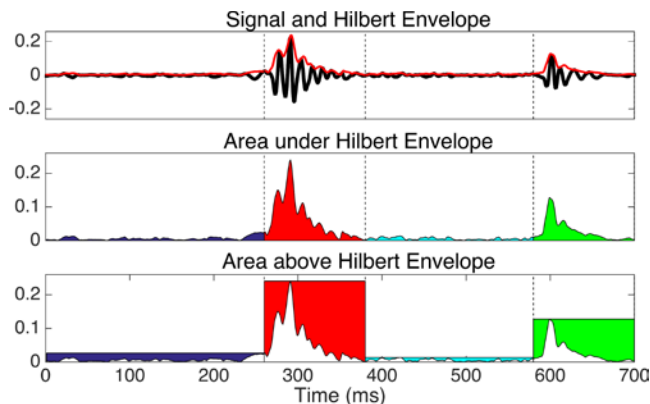


Figure 3. Hilbert envelope based features were used to capture sound intensity and sound shape

2.4. Classifier training

The classification problem was divided into two sub-classification tasks. First, PCG recordings were classified as noisy or adequate for further evaluation. Second, recordings determined to be of adequate quality were classified as normal or pathological.

A number of different classification models were considered for use including support vector machines, decision trees and neural networks. All models were implemented using the built-in functionality in MATLAB 2014b based on the initial requirements of the Challenge. For both tasks, each model was trained and validated using 10-fold cross validation with a balanced training set.

3. Results and discussion

3.1. PCG segmentation

Segmentation of the underlying signal into phases of the cardiac cycle was initially performed using a provided Hidden Markov Model. However, it was noted during the course of the competition that this model failed to accurately segment many recordings. In our experience, segmentation clearly failed to segment signals with murmurs during the early and late stages of systole and

diastole, instead selecting those murmurs as part of the preceding or following S1 or S2. A set of hand annotations was later provided for the competition; however, while the remainder of the feature extraction and classification on the training set could be performed with the manually annotated segmentation, the algorithm for the hidden test set relied on the less reliable Hidden Markov Model segmentation algorithm. Because further processing steps depended on having segmented recordings, it was necessary to continue to use this provided algorithm.

3.2. Wavelet features

Wavelet features were used for detection of timing and quality of heart sounds during each beat. Wavelet features from idealized recordings are shown in Figure 4. Clear differences in the feature set are seen in cases of three different murmurs.

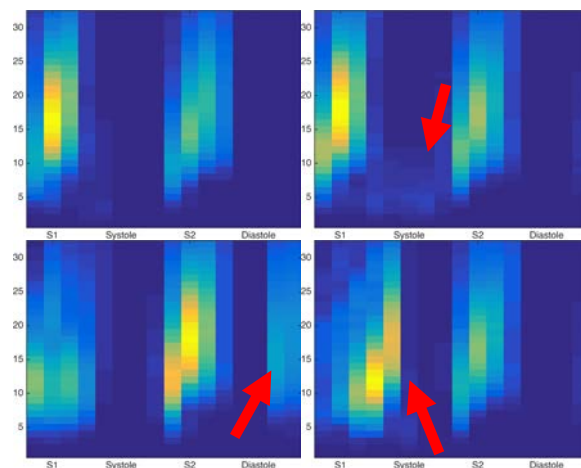


Figure 4. Wavelet values for normal (A), holosystolic murmur (B), mitral stenosis (C) and aortic stenosis (D). Areas of interest marked with arrow.

However, several issues were noted with the use of wavelet based features. First, moderate differences were noted in the character of the S1 and S2 across many recordings. While the wavelet features did recognize the presence of murmurs, they were of relatively low intensity and small changes in the position of highest intensity in S1 or S2 likely masked small changes due to murmurs. Further improvement of the wavelet features could emphasize features during systole and diastole and remove wavelet features from the S1 and S2 phases. Second, incorrect segmentation of the cardiac cycle phases led to the identification of early systolic murmurs as part of the S1 phase. Because the wavelet features divided each phase into 4 equal segments, the murmurs were buried within the overly-extended “S1” segment.

3.3. Classification

The signal quality classification task was performed with a variety of machine learning techniques (Table 1), and bagging trees was determined to be the most effective classification algorithm.

Table 1. Mean training and validation accuracy for signal quality classification task

Type of classifier	TA	VA
Bagging trees	1.00	0.94
Boosted trees (LogitBoost)	0.97	0.93
Logistic classifier	0.87	0.85
Support vector machine	1.00	0.91

The normal/abnormal classification task was then performed using those signals that were identified as being of good quality. Validation accuracy was measured using 10-fold cross validation. The boosted trees classifier using LogitBoost was selected for classification based on its highest validation accuracy (Table 2)

Table 2. Mean training and validation accuracy for normal/abnormal classification task

Type of classifier	TA	VA
Bagging trees (100 learners)	1.00	0.76
Boosted trees (LogitBoost)	0.93	0.76
Boosted trees (AdaBoostM1)	0.86	0.75
Boosted trees (RobustBoost)	0.89	0.75
Logistic classifier	0.80	0.72
Support vector machine	1.00	0.20

3.3. Competition results

The resulting algorithm achieved a sensitivity of 0.735 and a specificity of 0.746, for a score of 0.741 on a random subset of the test set. Classification of the full test set resulted in a sensitivity of 0.7958 and specificity of 0.7459 for a final score of 0.7708

Acknowledgements

This work was supported by National Institutes of Health grant 5T32GM007171 to the Duke Medical Scientist Training Program (T.G).

References

- [1] Liu C Y, Springer D B, Li Q, Moody B, Juan R A, Chorro F J, Castells F, Riog J M, Silva I, Johnson A E W, Syed Z, Schmidt S E, Panadaniil C D, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab M R, Sameni R, Mark R G and Clifford G D 2016 An open access database for the evaluation of heart sound algorithms *Physiol Meas* 37 (in press).
- [2] Schmidt, S.E.; Holst-Hansen, C.; Graff, C.; Toft, E.; Struijk, J.J. Segmentation of heart sound recordings by a duration-dependent hidden markov model. *Physiol Meas* 2010, 31, 513-529.
- [3] Springer, D.B.; Tarassenko, L.; Clifford, G.D. Logistic regression-hsmm-based heart sound segmentation. *IEEE Trans Biomed Eng* 2015, In press.

Address for correspondence.

Tanmay Gokhale
CIEMAS 1427
Box 90281
Durham, NC 27708-0281
tanmay.gokhale@duke.edu