Machine Learning Based Identification of Pathological Heart Sounds

Tanmay Gokhale
Duke University, Durham, North Carolina, USA

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 hearts 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

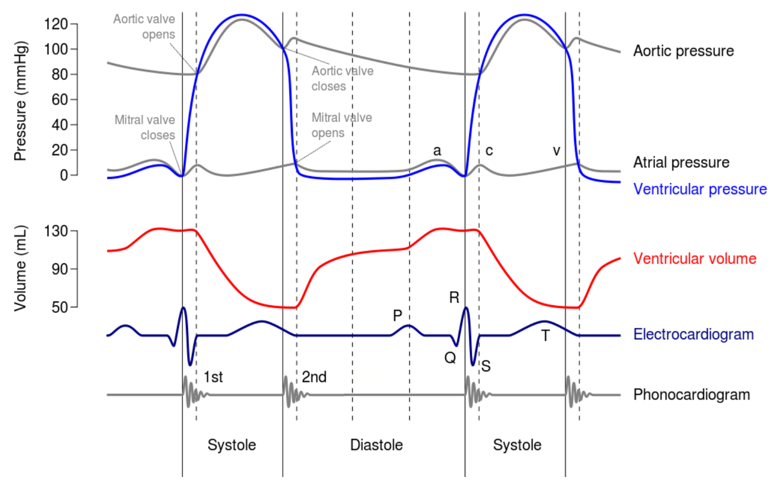

Figure 1. Wiggers Diagram showing cardiac pressures, volume and electrical activity with corresponding phonocardiogram.
Modified from figure by Daniel Chang MD, released under CC BY-SA license.

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 while the second heart sound (S2) occurs due to the closing of the aortic and pulmonary valves at the start of diastole. 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.

Pathological heart sounds are typically divided into murmurs, gallops and rubs. Murmurs are caused by turbulent flow of blood through a diseased valve that is either insufficient or stenotic. They 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. Gallops are extra heart sounds (similar to S1 and S2) occurring during diastole while rubs are caused by friction between layers of the pericardium in the setting of pericarditis

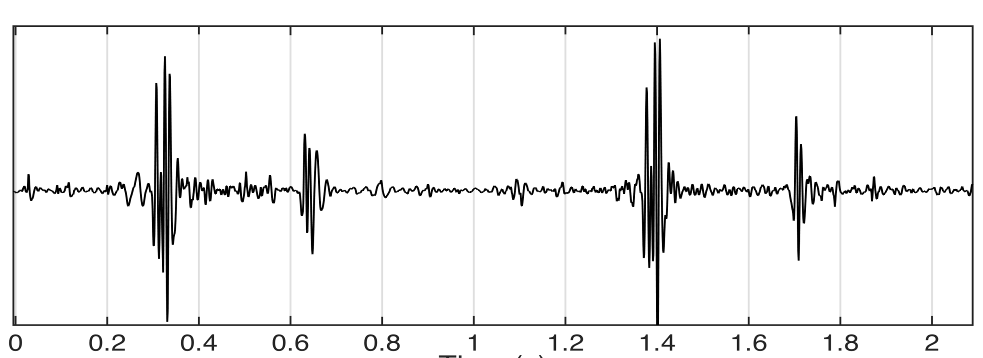

Figure 2. Normal phonocardiograph from two cardiac cycles.
Data from recording training-a/a0011

Many groups have developed algorithms for the automated interpretation of heart sounds. However, while these methods excel of interpreting recordings made in idealized settings, many fail to perform with high fidelity 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.

Here, 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 [1] 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.

Methods

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, 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.

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.

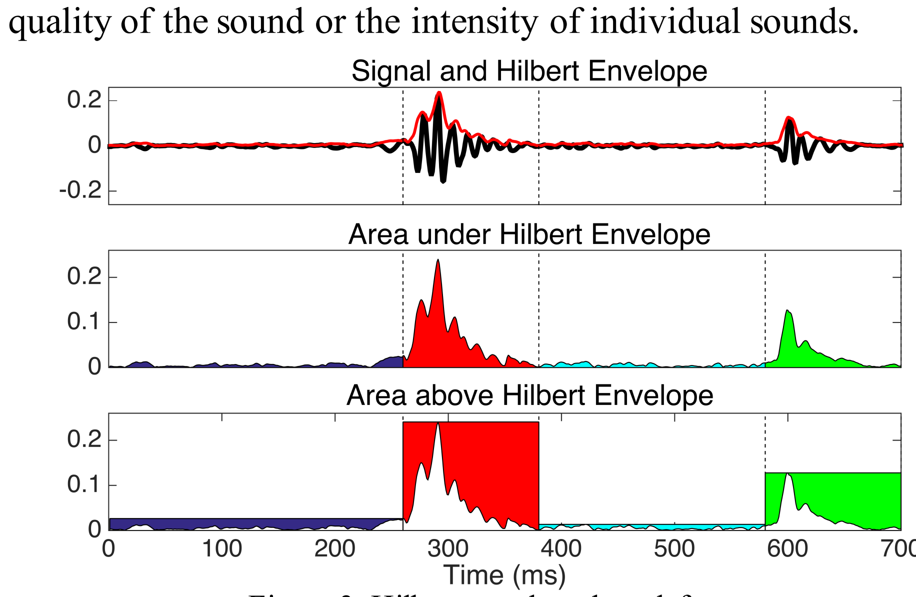

Figure 3. Hilbert envelope-based features

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.

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).

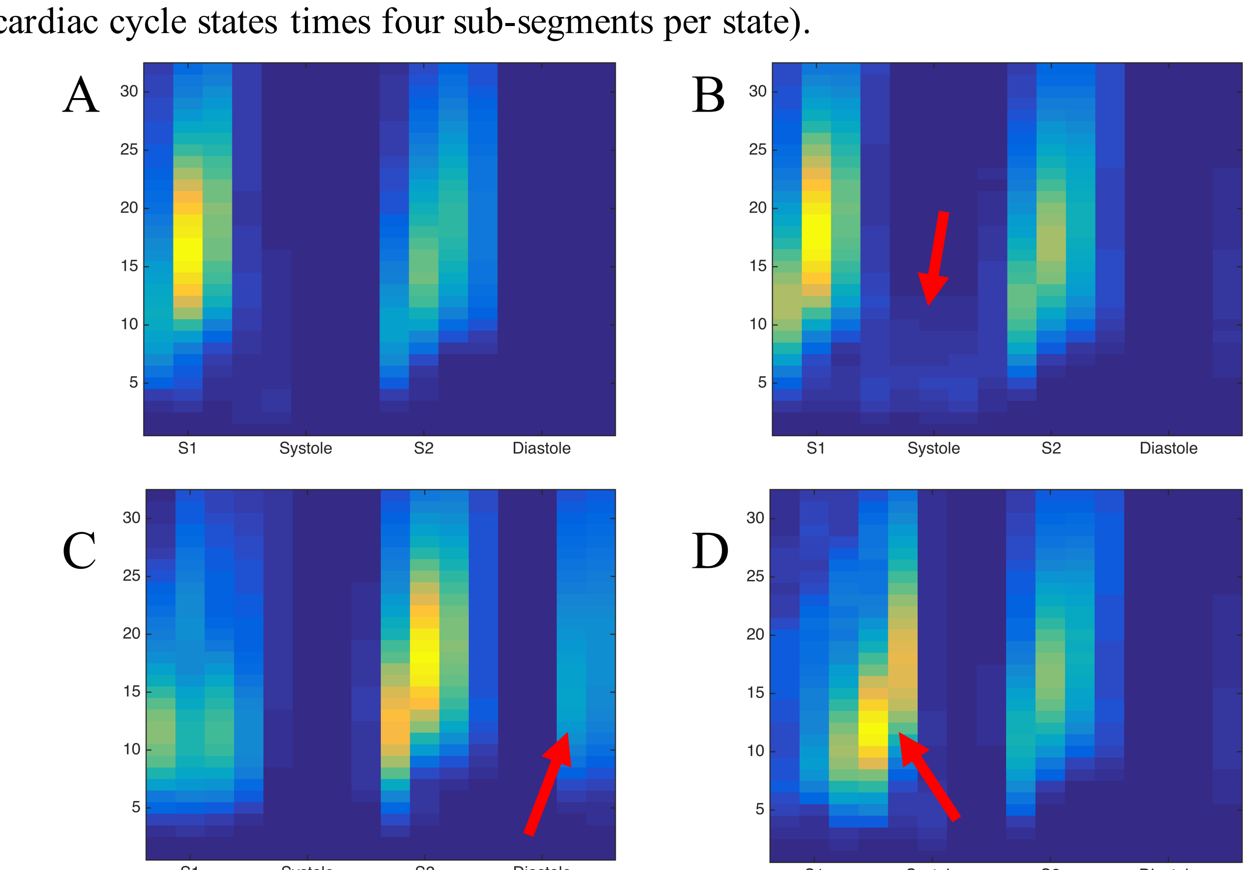

Figure 4. Wavelet values for normal (A), holosystolic murmur (B), mitral stenosis (C) and aortic stenosis (D). Areas of interest marked with arrow.

Results

Signal Quality Classifier

A classifier was first trained to identify signals of poor quality. Each classifier type was trained and validated using 10-fold cross validation with a balanced training set

Classification Method	Training Accuracy	Validation Accuracy
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 bagging trees classifier was selected for classification of signal quality.

Normal/Abnormal Classifier

Classification Method	Training Accuracy	Validation Accuracy
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	0.73
Support vector machine	1.00	0.20

The boosted trees classifier with LogitBoost was selected for classification of normal versus abnormal. The combined classifier achieved a sensitivity of 0.735 and a specificity of 0.746 in the PhysioNet Challenge.

Conclusions

Machine learning holds great promise in the automated interpretation of heart sounds. Our algorithm, which attempted to identify features of the heart sounds that physicians used during auscultation, was able to correctly classify recordings with 73.5% sensitivity and 74.6 specificity. Further work is needed in modifying segmentation algorithm to more accurately segment recordings with systolic pathological sounds, and in modifying the wavelet features such that minor differences in the normal S1 and S2 do not overpower detected differences caused by pathology.

References

[1] Liu C, Springer D, Li Q et al. An open access database for the evaluation of heart sound algorithms. *Physiological Measurements*. 37: 11. 2016. In Press

Acknowledgements

This work was supported by the Duke Medical Scientist Training Program (T32 GM007171)