

Automated Classification of Normal and Abnormal Heart Sounds using Support Vector Machines

Anthony Bouril¹, Darya Aleinikava², Maria S Guillem³, Grace M Mirsky²

¹STATS LLC, Chicago, IL, USA

²Benedictine University, Lisle, IL, USA

³ITACA, Universitat Politècnica de València, Valencia, Spain

Abstract

While a number of researchers have attempted to develop algorithms for automated classification of heart sounds over the last five decades, these studies have been inadequate in terms of clinical utility due to a number of important limitations. The PhysioNet/Computing in Cardiology 2016 Challenge seeks to facilitate the development of highly robust algorithms to perform automatic classification of heart sounds in a manner which overcomes the limitations of previous studies.

The dataset consists of over 3000 phonocardiogram recordings, taken from several locations on the body, from both healthy and pathological adults and children. The classification task requires the algorithm to determine if a recording is normal, abnormal or cannot be scored (due to excessive noise/corruption of the signal). An implementation of a state-of-the-art segmentation algorithm has been provided by the Challenge organizers, leaving the primary focus of the Challenge on the classification task. For this task, we selected a number of features in both the time and frequency domains. For Phase I, our best overall score for the hidden test set was 0.78 (Sensitivity = 0.70, Specificity = 0.87). For Phase II, our best score was 0.7864 (Sensitivity = 0.733, Specificity = 0.8398).

1. Introduction

Auscultation is a technique routinely used for evaluating the mechanical activity of the heart in primary care medicine. Abnormal heart sounds can indicate valvular disorders such as stenosis or regurgitation [1-3]. Audible murmurs arise from turbulence resulting from defects of the valves. Though it is a commonly used technique, the ability to accurately diagnose based upon a standard stethoscope is highly dependent upon the skill of the clinician. In particular, the diagnostic accuracy is strongly influenced by the hearing acuity of the user [4].

As a result, the ability to automatically identify irregular heart sounds could help to improve patient care by eliminating the need for a highly skilled examiner [5]. This analysis is accomplished by digitally processing the phonocardiogram (PCG), which is a recording of the acoustic signal generated by the heart. Advances in processing of the PCG have resulted in additional improvements in the quality of the results [6], making it a useful, low-cost, non-invasive diagnostic modality for certain cardiac pathologies such as regurgitation and stenosis. However, there is still a strong need to enhance the robustness of the automated classification of PCG's in order to allow this technique to be reliably used in a wide variety of applications, including self-evaluation for at-risk patients.

The objective of the PhysioNet/Computing in Cardiology 2016 Challenge is to encourage the development of algorithms to perform automatic classification of heart sounds as normal or abnormal. Furthermore, the Challenge is intentionally designed in a manner which overcomes the limitations of previous studies. These limitations are primarily a result of data that is not representative of the variety of data that may be obtained due to various recording positions and/or the amount of noise present, as well as lack of robust evaluative procedures of the algorithms such as validation on independent test sets.

2. Materials & methods

2.1. Data

Six databases consisting of a total of 3,126 phonocardiograms were provided for training [7,8]. The recordings were acquired from several locations on the body, from both healthy and pathological adults and children. Durations of the recordings varied from five seconds to just over two minutes. All recordings contain a single lead, which was acquired from one of nine locations on the body. Each signal was resampled to

2kHz. The holdout test set was hidden from participants for the duration of the Challenge. The Challenge was divided into two phases. In Phase I, the unofficial phase, participants were allocated five attempts to run their algorithms against the test set; in Phase II, the official phase, twelve attempts were given.

Upon review of the distribution of normal vs. abnormal records in the various databases, we parsed through the recordings to create a balanced dataset. As shown in Table 1, we selected an equal number of normal and abnormal recordings from each database. We also eliminated noisy records to create the training set.

Table 1. Distribution of normal and abnormal records from each database used for the training set.

Database	# Normal Records	# Abnormal Records
A	106	106
B	59	59
C	7	7
D	20	20
E	125	125
F	32	32
Total	349	349

2.2. Preprocessing

The Challenge allows for three different classifications of the recordings: -1 = normal, 0 = unsure due to noise, and +1 abnormal. Prior to classifying as normal vs. abnormal, each recording was analyzed for signal quality to determine if it should be classified immediately as “unsure”, since noisy recordings would likely not reliably be classified correctly as normal or abnormal.

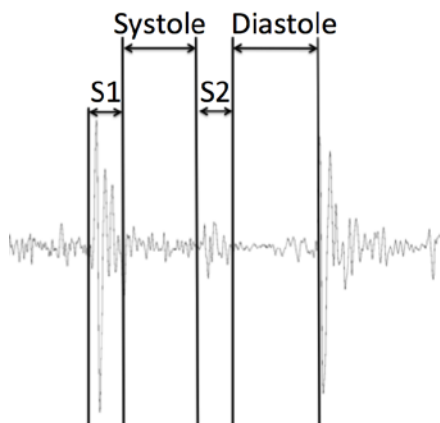


Figure 1. Example of segmentation of PCG into four intervals representing S1, systole, S2 and diastole.

For signals classified as “clean”, the phonocardiogram recordings were segmented using Springer’s improved

version of Schmidt’s method [9,10]. Each cardiac cycle was divided into four parts: S1, S2, systole and diastole. Segmentation is an important step as it allows for the different heart sounds and phases of the cardiac cycle to be analyzed independently. An example of segmentation is shown in Figure 1.

2.3. Classification

A support vector machine (SVM) was created and trained for classifying signals as normal or abnormal. This technique was chosen due to the fact that support vector machines have been shown to be effective in the classification of normal vs. abnormal phonocardiograms [11-13]. Seventy-four features were selected from both the time and frequency domains, with guidance provided by the prior studies as to which attributes would be most effective [11,14]. Sequential feature selection was utilized in order to determine the optimal subset of features used for classification.

The subset of features used in our highest scoring submission included the following:

Time-domain features

- Mean duration of diastole interval
- Mean of the ratio of duration of systole to the RR interval
- Maximum of the mean value of either the region between S1 and S2 or the region between S2 and next cycle’s S1
- 10% trimmed mean of the amplitude of the peaks in S2

Frequency-domain features

- Average of the ratio of the power of each diastolic segment in the frequency band from 100-300Hz to the power of each diastolic segment in the frequency band from 200-400Hz
- 20% trimmed mean of the power of each diastolic segment in the frequency band from 150-350Hz
- 20% trimmed mean of the power of each diastolic segment in the frequency band from 200-400Hz

Filtering in the frequency domain was accomplished using a set of Butterworth bandpass filters at the specified frequency ranges.

The SVM used for classifying signals as normal or abnormal was generated in Matlab® using the *fitcsvm* command [15]. The SVM utilized a Gaussian kernel, which was automatically scaled using a heuristic procedure. The box constraint was set to one, and the

cache size was set to the maximum value in order to allocate sufficient memory to contain the complete Gram matrix.

3. Results

All of our code for the Challenge was written in Matlab® [15] and submitted to the automated scoring system to be evaluated against the hidden test set. Scores for the official phase of the Challenge were calculated according to the following equations [16]:

$$\text{Sensitivity(Se)} = wa_1 \frac{Aa_1}{Aa_1 + Aq_1 + An_1} + wa_2 \frac{Aa_2}{Aa_2 + Aq_2 + An_2}$$

$$\text{Specificity(Sp)} = wn_1 \frac{Nn_1}{Na_1 + Nq_1 + Nn_1} + wn_2 \frac{Nn_2}{Na_2 + Nq_2 + Nn_2}$$

where the capital letters denote the reference label (A = abnormal and N = normal) and lowercase letters denote the output of the classification algorithm (a = abnormal, n = normal, and q = uncertain). The subscripts indicate the signal quality, with 1 indicating a clean signal and 2 indicating a noisy signal. The weights, shown by the variable w, are calculated from the actual distribution of records from the hidden test set. The values for these weights are shown in Table 2.

Table 2. Weight values for the sensitivity and specificity calculations.

Weight	Value
wa ₁	0.7881
wa ₂	0.2119
wn ₁	0.9467
wn ₂	0.0533

For the Challenge, we experimented with different training sets, numerous subsets of features, and various settings for the SVM in order to optimize our design. Using the balanced training set described in Table 1, rather than the entire training set, provided a significant improvement in our scores during the official phase. This improvement is likely a result of reducing the bias induced by the significantly unbalanced distribution of the entire training set that we had initially used for training.

Our scores in the official phase (Phase II) ranged from 0.746 to 0.810, which involved our submissions being evaluated over a randomly selected 20% of the test set. Our best final score achieved in the official phase of the Challenge was 0.7864 (Sensitivity = 0.733, Specificity =

0.8398) when our algorithm was run over the entire dataset rather than the randomly selected subset.

4. Discussion

In this study, support vector machines have been shown to effectively classify phonocardiograms as normal or abnormal. The machine learning technique chosen, support vector machines, was selected due to its proven record of performance in the classification of biomedical signals. The selection of a well-balanced, representative training set as well as the identification of optimal features for classification involved a significant effort. As expected, the resulting performance was sensitive to the particular features chosen. We also observed through the Challenge that our algorithm tended to perform better in specificity as compared to sensitivity. Therefore, the refinement of the selection of features as well as potentially adding features to further improve the accuracy of the method will be the focus of future work.

Acknowledgements

The authors would like to thank the organizers of the PhysioNet/Computing in Cardiology Challenge for facilitating the opportunity to work on a very interesting problem with a large, diverse dataset.

References

- [1] Waider W, Craige E. First heart sound and ejection sounds: Echocardiographic and phonocardiographic correlation with valvular events. *The American journal of cardiology.* 1975;35(3):346-56.
- [2] Mills PG, Brodie BR, McLaurin L, Schall S, Craige E. Echocardiographic and hemodynamic relationships of ejection sounds. *Circulation.* 1977;56(3):430-6.
- [3] Armstrong TG, Mitha AS, Matisonn RE, Weir EK, Chesler EL. Non-invasive observations on initial low frequency vibrations of the first heart sound--correlation with the 'presystolic' murmur in mitral stenosis. *British heart journal.* 1978;40(7):750-7.
- [4] Mangione S, Nieman L. Cardiac auscultatory skills of internal medicine and family practice trainees. *Journal of the American Medical Association* 1997;278:717-722.
- [5] 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* 2009;37(2):337-53.
- [6] Durand, LG, Pibarot, P. Digital signal processing of the phonocardiogram: Review of the most recent advancements. *Critical Reviews in Biomedical Engineering* 1995;22(3/4):163-219.
- [7] Liu C, Springer D, Li Q, et al. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement* 2016;37(9):in press.
- [8] Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, physioToolkit, and physionet components of a new research

- resource for complex physiologic signals. *Circulation*. 2000;101(23):e215-20.
- [9] Schmidt SE, Holst-Hansen C, Graff C, et al. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological Measurement*. 2010;31(4):513.
- [10] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*. 2016;63(4):822-32.
- [11] Maglogiannis I, Loukis E, Zafiroopoulos E, Stasis A. Support vectors machine-based identification of heart valve diseases using heart sounds. *Computer methods and programs in biomedicine*. 2009;95(1):47-61.
- [12] Güraksın GE, Uguz H. Classification of heart sounds based on the least squares support vector machine. *International Journal of Innovative Computing, Information and Control* 2011;7(12):7131-7144.
- [13] Kao WC, Wei CC, Liu JJ, Hsiao PY. Automatic heart sound analysis with short-time fourier transform and support vector machines. *52nd IEEE International Midwest Symposium on Circuits and Systems*. 2009:188-191.
- [14] Singh M, Cheema A. Heart sounds classification using feature extraction of phonocardiography signal. *International Journal of Computer Applications*. 2013;77(4).
- [15] MATLAB and Statistics & Machine Learning Toolbox Release 2016, The MathWorks, Inc., Natick, Massachusetts, United States.
- [16] <https://physionet.org/challenge/2016/>

Address for correspondence.

Prof. Grace M. Mirsky
Department of Computer Science
Benedictine University
5700 College Road
Lisle, IL, USA
gmirsky@ben.edu