A Multi-Modal Classifier for Heart Sound Recordings

Xulei Yang[1*], Feng Yang[1], Like Gobeawan[1], Si Yong Yeo[1], Shuang Leng[2], Liang Zhong[2], Yi Su[1]

[1]Institute of High Performance of Computing, A*STAR, Singapore
[2]National Heart Centre Singapore, Singapore

Abstract

Information extracted from heart sound signals are associated with valvular heart diseases and other cardiovascular disorders. This study aims to develop a computational framework for the classification of a given heart sound recording. Different techniques have their respective superiority in classifying heart sound recordings with various patterns, and it is difficult to find one technique that outperforms all the others. We hence propose a multi-modal classifier by fusing the classification results from various techniques based on various features.

Using the data obtained from the 2016 PhysioNet/CinC Challenge, we generate two different feature sets: one set was calculated from segmented results by peak-finding method, and the other set was extracted by audio signal analysis. We then assess the performance of two classification techniques – support vector machines (SVMs) and extreme learning machines (ELMs) – by feeding them with the best subset of features selected from these two feature sets. The final heart sound classification result (normal / abnormal) is determined by ensembling the two classifiers with voting. The best performance out of five online entries achieved an overall score of 0.83 with sensitivity=0.70 and specificity=0.96.

1.   Introduction

The heart sound or phonocardiography (PCG), produced by opening and closure of atrioventricular and semilunar valves, carries early pathological information of cardiovascular dysfunction [1]. Traditionally, heart sound is used by the medical specialists for detection of heart diseases through auscultation. The emergence of digital heart sound recording and sophisticated computer-based machine learning techniques allows automatic, fast and reliable evaluation of heart sound signals, thus overcoming the reliance of conventional auscultation on clinical experience and examination skills.

The standard approach for the computer-based automatic heart sound analysis includes the following steps: 1) Filtering low frequency data, this is based on the premise that the heart sound is of low frequency and most energy is concentrated in the region below a certain frequency. 2) Using Hilbert transformation and then calculating the Shannon energy, which borrows the result from some earlier work done [2]. 3) Smoothing on the curve and then using the derivative change to find all possible points of local maxima/minima. 4) Fitting nearby maxima/minima by a parabola and replacing all the nearby points with the peak of the parabola. 5) After the peaks are located, classification of heart sound is then carried out based on the length of the diastolic and systolic period.

In the past years, a large number of approaches have been proposed for each step of the above heart sound analysis process. The interested users may refer to the review papers [3, 4] for further details. Different techniques have their respective superiority in classifying heart sound recordings with various patterns and it is difficult to find one method that consistently performs better than others. In contrast to a silo approach in addressing the problem, we develop a computational framework for the classification of a given heart sound recording by adopting a fusion approach. In this work, we propose a multi-modal classifier by fusing the classification results from various techniques based on various feature sets. Specifically, two classification techniques, i.e., support vector machine (SVM) [5, 6] and extreme learning machine (ELM) [7, 8], as well as two different feature sets, i.e., one from audio signal analysis, and the other from peak segmentation, are integrated into the computational framework. Details of the proposed method with related techniques are discussed in Section 2. Section 3 describes the experiments and results, and Section 4 concludes this paper.

2.   Methodology

The overall workflow of the proposed multi-model classifier is illustrated in Fig.1. The associated classification techniques and feature extraction/selection methods are elaborated below.

2.1. Classification Techniques
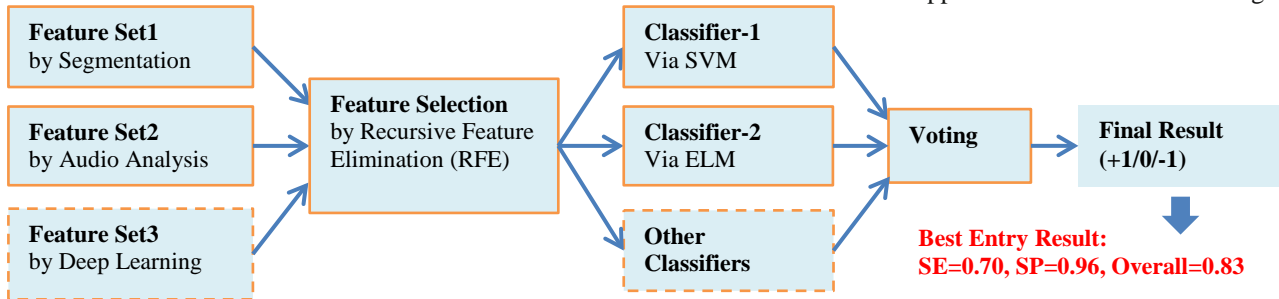
2.1.1. Support Vector Machine (SVM)

Figure 1. The overall workflow of the proposed multi-modal classifier. The best score out of five entries is 0.83. The scope within dotted boxes are reserved for future developments.

The support vector machines (SVMs) model, first proposed by Vapnik and his colleagues [5, 6], is a widely used machine learning technique for pattern recognition. As compared to traditional multivariate methods, linear SVMs has quite good generalization capabilities to unseen data and it has been extended to non-linear implementations by coupling with appropriate kernels such as radial basis function (RBF) kernel. The basic principle of SVMs is to construct an optimal hyperplane, which maximizes the margin on both sides of the hyperplane that separates data points from two classes, and hence minimizes the empirical error, as well as maximizes the geometric margin at the same time, to avoid overfitting problem. In other words, the SVMs constructs a boundary that maximizes the distance between the designated class of each sample (e.g., whether the heart sound recording is "normal" or "abnormal"), which leads to the optimal boundary classifying the samples. In this study, we used the popular libSVM package [9] with RBF kernel function to implement the SVM for classification.

2.1.2. Extreme Learning Machine (ELM)

The ELM is a single-hidden-layer feed-forward neural network (SLFN) algorithm for classification and regression [7, 8]. The essence of the ELM is that the SLFN doesn't need to be tuned, which is different from conventional approach of machine learning. One typical implementation of the ELM is to randomly dispatch computing nodes for the hidden layer, where those nodes could be of independence from the training data. Compared to conventional learning methods for neural networks like back propagation with gradient-based algorithm, ELMs minimise training errors and norms of output weights. ELMs is able to avoid many difficulties faced in pratice by gradient-based learning methods, for example, to determine the stopping criteria, learning rate,

learning epochs as well as local minima [7]. The consequent benefits include shorter computing time for building the model, decreased model complexity, and enhanced predicting performance. ELMs has been used in various fields and applications because of its good generalization ability, robustness, and controllability and fast learning rate. The latest study of ELMs has discovered that ELM's minimal norm of the output weight matches SVM's maximal separating margin. Also, the ELMs tends to exhibit improved scalability and achieves much better generalization performance than traditional SVMs and LS-SVM [11].

2.1.3. K-fold cross validation

In the construction of a real pattern classification system, the available data are generally limited, hence there is a need of a validation technique to estimate how a classification system will perform in practice. We use the *k*-fold cross-validation [10] to evaluate the classification performance. In one round of *k*-fold cross-validation, the data set is randomly partitioned into *k* equal-sized, mutually-exclusive subsets. A SVM classifier is then trained and tested k times, where in each time, one of the subsets is set aside as the testing data and the remaining as the training data. The estimated *k*-fold cross-validation error rate is the total number of incorrectly classified samples in the data set divided by the total number of samples in the data set. In the experimental studies, 10-CV (i.e., *k*=10) was used as the validation method.

2.2. Feature sets

Feature extraction and selection are the most important steps for the heart sound classification task. In this subsection, we will describe the two feature sets that are extracted from a given heart sound recording.

2.2.1. Features by audio signal analysis

This feature set is calculated from the following steps by using standard audio signal analysis techniques:

1): The given heart sound recording is down-sampled to 50 milliseconds non-overlapping frames. After that, short-term features (i.e., spectral roll-off/centroid/flux, short-time energy, energy entropy, and zero crossing rate) are calculated from each frame, leading to 6 feature sequences for the given heart sound recording.

2): For each of these 6 feature sequences, 2 statistical features are calculated, namely, standard deviation and standard deviation divided by mean. This leads to 12 single statistical values that characterize the first feature set.
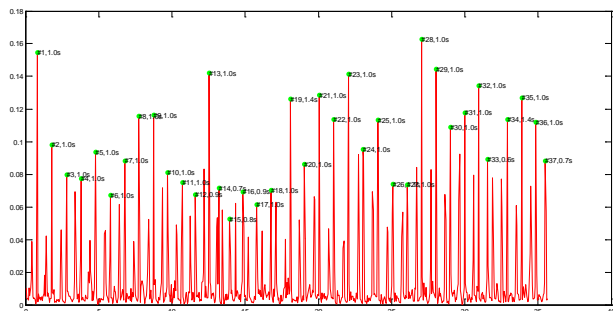

Figure 2. Illustration of the located peaks for sample 'a0001.wav'.

2.2.2. Features by segmentation

This feature set is derived in a two-step approach:

Step-1: The given heart sound recording is down-sampled to 50 milliseconds non-overlapping frames. For each frame, the short-term feature, i.e., Energy Entropy, is calculated. Based on the histogram of the calculated Energy Entropy, the "findpeaks" method from Matlab Signal Processing Toolbox is used to locate the peaks, as illustrated in Fig.2. Duplicated peaks are filtered by using the distances between adjacent peaks.

Step-2: In the sequel, after calculating the distances between adjacent peaks, 4 statistical features are calculated from these distance measurements, namely, the standard deviation, standard deviation divided by mean, mean, and median are calculated for each sample. This step leads to 4 single statistic values that characterize the second feature set.

2.2.3. Feature selection

Feature selection is used to select the most relevant features whilst discarding the noisy and irrelevant ones. Among the many existing feature selection methods, the Support Vector Machine with Recursive Feature Elimination (SVM-RFE) algorithm has been widely used due to its simplicity and good performance in many practical problems [12]. The SVM-RFE algorithm incorporates SVM and RFE to backwardly select feature subsets. In particular, a linear SVM is trained at each step,

after which the features are ranked based on the absolute value of weights from SVM and the worst batch of features are then discarded. Starting with a full candidate set, features are removed sequentially such that the variation of separating boundary is minimized and it stops until the desired number of features is reached. Different desired number of features were evaluated to determine the performance of the various feature combinations.

3.   Experimental results

3.1.   Datasets

The heart sound datasets in this study are provided by the PhysoNet/CinC Challenge 2016 [13]. The training dataset contains 3,126 heart sound recordings, which are lasting from quite short period (5 seconds) to rather long period (120 seconds). There are totally 5 databases indicated by letters A to E. Inside each database, the sound recordings begin with a same letter followed by a number. There is also a test dataset not available to public and for the scoring purpose only. The training and test datasets are divided in such a way that same patient will not be in numerically adjacent, and there are no overlapped samples in both dataset. In this study, the training set is randomly divided into 10 folders for cross-validation purpose.

Table 1. Summary of 5 trials of various combinations on test datasets from PhysoNet/CinC Challenge 2016 [13].

Combinations (Trials 1 to 5)	Results (ACC/TPR/TNR)
1.SVM +Feature Set 1	0.76/0.62/0.90
2.SVM +Feature Set 1&2	0.78/0.69/0.87
3.ELM +Feature Set 1	0.70/0.50/0.89
4.ELM +Feature Set 1&2	0.76/0.69/0.83
5.SVM&ELM +Feature Set 1&2	0.83/0.70/0.96
Summary	
Average	0.77/0.64/0.89
Worst (Trial 3)	0.70/0.50/0.89
Best (Trial 5)	0.83/0.70/0.96

3.2.   Evaluation metric

The accuracy (ACC), true positive rate (TPR), and true negative rate (TNR) are used in this study for the performance evaluation of the proposed multi-modal classifier by the following equations:

$$\text{ACC} = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{fp} + N_{tn} + N_{fn}}$$

$$TPR = \frac{N_{tp}}{N_{tp} + N_{fn}}$$

$$TNR = \frac{N_{tn}}{N_{tn} + N_{fp}}$$

where $N_{\_tp}$ is the number of true positive samples (i.e., correctly predicted abnormal heart sound recordings), $N_{\_fp}$ is the number of false positive samples (i.e., wrongly predicted abnormal heart sound recordings), $N_{\_tn}$ is the number of true negative samples, and $N_{\_fn}$ is the number of false negative samples.

3.3. Testing results

The performance of the proposed framework on heart sound classification is evaluated on test datasets from the PhysioNet/CinC Challenge 2016 [13]. We submitted 5 trials by using various combinations of classifiers and feature sets. Table. 1 summarizes the various combinations and corresponding outcome on the test datasets. The observations from the results are as below:

1) The combination using two classifiers with two feature sets generated the best performance of overall score of 0.83 with sensitivity=0.70 and specificity=0.96.
2) The SVM classifier performed slightly better than the ELM classifier when using the same feature set(s).
3) The combination of two feature sets performed better than each individual feature set by using the same classifier(s).

These results showed the superiority of our proposed multi-model classification approach.

4. Conclusion

Accurate classification of heart sounds is crucial in medical diagnosis. The heart sound corresponding to different heart symptoms, such as murmur and extra-sound, often has subtle differences and robust classifiers are desired to distinguish among them. Machine Learning methods can be applied on these sounds to enable efficient classification. In this study, we developed a multi-modal classifier for this task. It employed a combination of various feature extraction and classification techniques. At this preliminary stage, we achieved a classification accuracy of 0.83, which is promising considering that we used only two classifiers with two feature sets. For the future work, we will continue to improve the performance by fusing more classification techniques as well as different ways to extract feature. On one hand, we will investigate alternative classifiers such as gradient boosting, random forests, and artificial neural networks. On the other hand, we will study the possibility to generate a new feature set by using deep learning approaches [14].

References

[1] S. Yuenyong, A. Nishihara, W. Kongprawechnon, K. Tungpimolrut. A framework for automatic heart sound analysis without segmentation. Biomedical Engineering Online 2011;10:13.

[2] C. N. Gupta. Neural network classification of homomorphic segmented heart sounds. Applied Soft Computing 2007;7-1:286–297.

[3] B.S. Emmanuel. A review of signal processing techniques for heart sound analysis in clinical diagnosis. Journal of Medical Engineering & Technology 2012;36-6:303-307.

[4] U. More Monali, R. Shastri Aparana. Review on heart sound analysis technique. Advances in Intelligent Systems and Computing 2013;216:93-101.

[5] B.E. Boser, I.M. Guyon, V.N. Vapnik. A training algorithm for optimal margin classifiers. The Fifth Annual Workshop on Computational Learning Theory 1992:144–152.

[6] C. Cortes, V.N. Vapnik. Support vector networks. Machine Learning 1995;20-3:273-297.

[7] Q.Y. Zhu, A.K. Qin, P.N. Suganthan, G.B. Huang. Evolutionary extreme learning machine, Pattern Recognition 2005;38-10:1759–1763.

[8] G.B. Huang, Q.Y. Zhu, and C.K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 2006;70-1–3:489–501.

[9] C-C. Chang, C-J. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2011;2-3:1-27.

[10] E. Bradley, T. Robert. Improvements on cross-validation: The 632 + Bootstrap Method. Journal of the American Statistical Association 1997;92-438:548-560.

[11] G.B. Huang, H. Zhou, X. Ding, R. Zhang. Extreme learning machine for regression and multi-class classification. *IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics* 2012;42-2:513-529.

[12] I. Guyon, J. Weston, S. Barnhill, V. Vapnik. Gene selection for cancer classification using support vector machines. Machine Learning 2002;46-1–3:389–422.

[13] PhysioNet/CinC Challenge 2016, https://www.physionet.org/challenge/2016/

[14] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature* 2015;521:436-444.

Address for correspondence.

X. Yang
Institute of High Performance Computing
1 Fusionopolis Way, Connexis (North) #16-16
Singapore 138632
yangx@ihpc.a-star.edu.sg