A Novel Feature Extraction Method for Heart Sounds Classification

Yucel Kocyigit

Celal Bayar University, Electrical and Electronics Eng. Dep. Manisa Turkey
yucel.kocyigit@cbu.edu.tr

Abstract. The experimental results showed that the proposed method efficiently classifies heart sounds. Heart sound analysis is a basic method for heart examination, which may suggest the presence of a cardiac pathology and also provide diagnostic information. In this study, a novel feature extraction method based on Independent Component Analysis is applied to classify nine different heart sound categories. The extracted features are subjected to classification by Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) using 5 Cross Validation and by Artificial Neural Network. The experimental results showed that the proposed method efficiently classifies heart sounds.

Keywords: Heart Sound Classification, Independent Component Analysis, Linear Discrimination Analysis, Support Vector Machine, Artificial Neural Network.

1 Introduction

Auscultation is a technique in which a stethoscope is used to listen to the sounds of a body. The structural defects of the heart are often reflected in the sounds the heart produces. Physicians use the stethoscope as a device to listen to the patient's heart and make a diagnosis accordingly. They are particularly interested in abnormal sounds, which may suggest the presence of a cardiac pathology and also provide diagnostic information. For instance, a very important type of abnormal sound is the murmur, which is a sound caused by the turbulent flow of blood in the cardiovascular system. The timing and pitch of a murmur are of significant importance in the diagnosis of a heart condition, for example, murmurs during diastole are signs of malfunctioning of heart valves but murmurs during systole may correspond to either a pathological or healthy heart, depending on the acoustic characteristics of the murmurs.

In the literature, it is observed that time–frequency/scale methods have been applied to characterize heart sounds [1,2]. In previous publications, the authors have discussed the characterization of heart murmurs using time–frequency methods over a number of cardiac cycles [3,4]. It is observed that wavelet transforms have been frequently used to extract features from heart sounds [5-7]. Recently, there have been

studies about determination of HS features using segmentation methods for diagnosing Cardiovascular disease[8-9].
In this study, a novel feature extraction method based on Independent Component Analysis (ICA) is applied to analyze nine different heart sound categories. Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Artificial neural networks (ANNs) have been used to classify heart sounds [3,6,10]. The type of ANNs used in these papers was called Multi-Layer Perceptrons.

2 Methods

2.1 Independent Component Analysis (ICA)

ICA can be used for a feature extraction method that transform multivariate mixture signal into a signal having components that are mutually independent [11,12]. Independent components can be extracted from the mixture signals by using this method . Thus, independence denotes the information carried by one component cannot be gathered from the others.
Mixture signals consist of independent original signals which overlap under different conditions. ICA can solve the problem by which observation of the mixture signals decomposed independent signals.
Assume that we observed an M-dimensional observation vector expressed as;

$$x(t) = [x_1(t), x_2(t), \ldots\ldots, x_M(t)]^T (\text{¡ Error! Marcador no definido.})$$

And an N-dimensional original vector s(t) can written as;

$$s(t) = [s_1(t), s_2(t), \ldots\ldots, s_N(t)]^T (\text{¡ Error! Marcador no definido.})$$

Without loss of generality, we can assume that both the mixture variables and the independent components have zero mean: If this is not true, then the observable variables x(t) can always centered by subtracting the sample mean, which makes the model zero-mean. Using this vector matrix notation, to give the the relation between x(t) and s(t) the above mixing model is written as

$$x(t) = a_1 s_1 + a_2 s_2 + \ldots\ldots + a_N s_N = As(t)(\text{¡ Error! Marcador no definido.})$$

It is implicit that there exists a linear relationship between the signals x and s, so the vector A is given as;

$$A = [a_1, a_2, a_3, \ldots . a_N](\text{¡ Error! Marcador no definido.})$$

The aim of the use of ICA is to find the estimation of the mixing matrix A, consequently the independent source vectors from the observed mixed vector x. This aim is the same to find a separating matrix W that satisfies,

$$\hat{s} = Wx(\text{¡ Error! Marcador no definido.})$$

where $\hat{s}$ is the estimation of s.

There are two quite standard preprocessing steps in ICA. First, the mean of the data is usually subtracted to center the data on the origin. This is to ensure that the components have a zero mean. The second step is to whiten the data. This means that the data is transformed so the components are uncorrelated and have unit variance.
Having preprocessed the data, the goal of ICA is a transform W which minimizes the statistical dependencies between the estimated sources. The independence criteria chosen depend on the data to be analyzed. There are several methods which include ICA by maximization of nonGaussianity, ICA by Kurtosis Maximization or Minimization, Negentropy and the FastICA algorithm. For the ICA in the study, the FastICA software package for Matlab is applied [13-14]. The output obtained on applying the independence criteria is multiplied with the whitening output. This result is multiplied with the transpose of the input data.

2.2 Linear Discriminant Analysis (LDA)

For a two-class classification problem, LDA tries to find one hyper-plane to separate one group from another one. There are various parameters that affect performance of LDA including distance metric such as Euclidean and Mahalanobis distances [15]. For the sake of simplicity, the Euclidean distance metric was selected for this study. Since the classes are more than two in this study, majority voting is used for decision making.

2.3 Support Vector Machine (SVM)

SVM based classifier is one of the widely utilized classifiers in the post-genome era due to its learning and generalization ability for high-dimensional data sets [16]. It is claimed for the SVM classifier that a two-group data set can always be separated by a hyper-plane provided that a suitable non-linear mapping to a sufficiently high dimension is found. In order to achieve it, polynomials, Gaussians or other basic functions (now called "kernels") may be used. Furthermore, one of the main tasks during the construction of SVMs is to find separating hyper-plane(s) with the largest possible margin. It should be noted that a SVM classifier with the larger margin generally results in a classifier with better generalization ability.

The support vectors derived from the samples during training are the data points that highly represent the hyper-planes. These points can then be regarded as the most representative data samples that could help construct a robust classifier [15]. SVM can classify more than two classes in MATLAB.

2.4 Artificial Neural Network (ANN)

The ANN is an important information processing paradigm. The ANN is represented by weighted interconnections between processing elements (Figure 1). These synaptic weights are the parameters that actually define the non-linear function performed by the ANN. The process of determining such parameters is called training or learning, relying on the presentation of many training patterns.

The Back-Propagation (BP) algorithm is the most widely used training algorithm because of its relative simplicity and universal approximation capacity. The BP algorithm defines a systematic way to update the synaptic weights of multi-layer perceptron (MLP) networks. The supervised learning is based on the gradient descent method, minimizing the global error on the output layer.

The definition of the network topology (the number of hidden layers and of neurons in each layer) is a compromise between the generalization and convergence. The convergence is the capacity of the network to learn the patterns on the training set, and the generalization is the capacity to respond correctly to the new patterns.

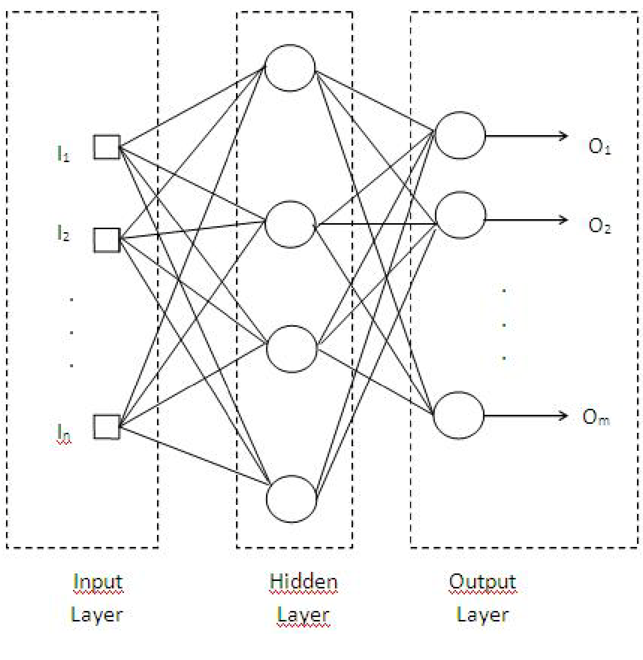

Fig. 1. ANN Structure

3 Results

In this study, the features are determined by analyzing nine different heart sounds: opening snap (OPS), aortic stenosis (AST), midsystolic click and late systolic Murmur (MCC+LSM), normal FCG (NFC), third heart sound (S3), fourth heart sound (S4), ventricular septal defect (VSD), patent ductus arteriosus (PDA), mitral stenosis (MST). Figure 2 shows the heart sounds examined in the study.

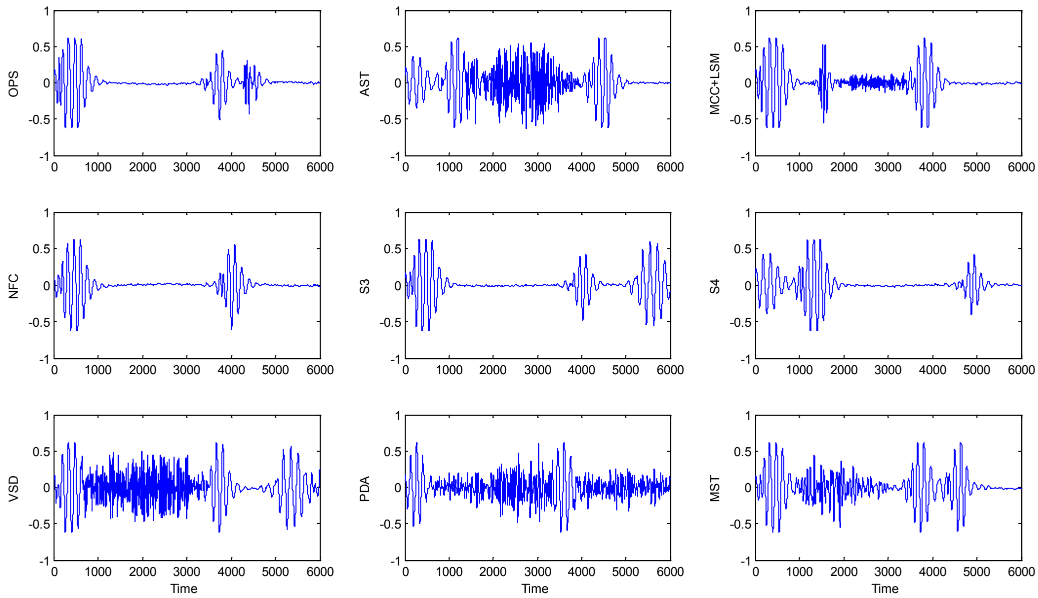

Fig. 2. One period of nine different HSs.

HS signals are sampled at 7500 frequency and analyzed within windows that contain 6000 discrete data. Each heart record contains 17 periods of heart sounds which are used 15 periods of them for training, and the other 2 periods for testing. The extracted features are subjected to classification by LDA and SVM using 5 Cross Validation and ANN.

Feature extraction is one of the most important processing steps in decision making. In this study, a novel feature extraction method based on Independent Component Analysis is applied to analyze nine different heart sound categories. For the ICA, the FastICA software package for MATLAB is applied to the HS signals and obtained 20 independent components. The number of components are determined by a criteria which the components present the HS signals substantially.

LDA and SVM have 95.5% and 84.4% training performances using 5 Cross Validation. On the other hand, the classification results of both methods are 89% (Table 1).

Table 2 and Table 3 shows the classification performances by LDA and SVM respectively. As can be seen, both methods give same results.

The other method, multi-layer perceptron (MLP), also used for classification. MLP is trained by the data set including 135 heart sounds periods. Moreover, a test set is formed by 18 heart sounds periods of nine different categories. Optimal MLP topologies are determined after several trials for each feature extraction process. It is observed that 94% performance (the ratio of the true positives to the total vectors in the test set) is obtained by using MLP of 20-20-1 topology. Table 4 shows the classification performances by MLP for nine different heart sounds.

As it can be seen from Table 1, all results are encouraging for applying ICA to the signal to obtain the features. Though LDA is a simple method for classification, it yields good results. On the other hand, the training performance of SVM is worse than LDA, but it also produces good result. ANN gives better results because of its generalization.

Table 1. All results for classification methods

	LDA (%)	SVM (%)	ANN (%)
Training Performance	95.5	84.4	
Test Performance	89	89	94

Table 2. Classification performances by LDA

	OPS	AST	MCC	NFC	S3	S4	VSD	PDA	MST
OPS	2								
AST						2			
MCC			2						
NFC				2					
S3					2				
S4						2			
VSD							2		
PDA								2	
MST									2

Table 3. Classification performances by SVM

	OPS	AST	MCC	NFC	S3	S4	VSD	PDA	MST
OPS	2								
AST						2			
MCC			2						
NFC				2					
S3					2				
S4						2			
VSD							2		
PDA								2	
MST									2

Table 4. Classification performances by MLP network

	OPS	AST	MCC	NFC	S3	S4	VSD	PDA	MST
OPS	2								
AST	1	1							
MCC			2						
NFC				2					
S3					2				
S4						2			
VSD							2		
PDA								2	
MST									2

4 Conclusions

Differentiating from the other studies, this study has only used ICA for the feature extraction of the HSs and succeeded the classification without utilizing any time/frequency extraction methods. The results of applying the ICA are more encouraging, even though simple classification methods like LDA are used. The classification result of ANN shows that more complex methods can yield better performance. In future works, it may be able to try classification with vectors contained the powers of the signal in time and frequency added to ICA coefficients.

5 Acknowledgment.

The author thanks Prof. Z. Dokur and Prof. T. Ölmez for providing the data sets used in this study.

6 References

1. F. Debjais, L.G. Durand, Z. Ouo, R. Guarlo, "Time–frequency analysis of heart murmurs. Part II: Optimization of time–frequency representations and performance evaluation", *Med. Biol. Eng. Comput*. Vol. 35, pp. 480–485, 1997.
2. P.M. Bently, Time–frequency analysis of native and prosthetic heart valve sounds, Ph.D. thesis, Electrical and Electronics Department, University of Edinburgh, 1996.
3. T.S. Leung, P.R. White, W.B. Collis, E. Brown, A.P. Salmon, "Classification of heart sounds using time frequency method and artificial neural networks", Proceedings of the 22nd Annual EMBS International Conference,2000, pp.988-991
4. H. Yoshida, H. Shine, K. Yana, "Instantaneous frequency analysis of systolic murmur for phonocardiogram", Proc. of the 19th Annual International Conference of the IEEE-EMBS, 1997, pp. 1645–1647.

5. E. Comak, A. Arslan, I. Turkoglu, A decision support system based on support vector machines for diagnosis of the heart valve diseases,Comput. Biol. Med. 37 (2007) 21–27.
6. Z. Dokur, T. Ölmez, “Feature determination for heart sounds based on divergence analysis”, *Dig. Signal Proc.* Vol.19, pp. 521–531, 2009.
7. H. Uğuz, “Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy”, *Neural Comput & Applic* Vol.21, pp.1617–1628, 2012.
8. Z. Yan, Z. Jiang, A. Miyamoto, Y. Wei, “The moment segmentation analysis of heart sound pattern”, *Comp. Meth. And Prog. In Biomedicine* Vol.98, pp. 140–150, 2010.
9. H. Tanga, T. Li, T. Qiua, Y. Park, “Segmentation of heart sounds based on dynamic clustering”, *Biomedical Signal Processing and Control* , Vol.7, pp. 509– 516, 2012.
10. S. Ari, G. Saha, “In search of an optimization technique for Artificial Neural Network to classify abnormal heart sounds”, *Applied Soft Computing* Vol.9, pp. 330–340, 2009.
11. A. Hyvärinen, “Survey on independent component analysis” Neural Computing Surveys, Vol.2, pp.94–128, 1999.
12. A. Chagnaa, C. Ock, C. Lee, P. Jaimai, “Feature Extraction of Concepts by Independent Component Analysis”, Int. J. of Inf. Proc. Sys, Vol.3, No.1, pp.33-37, June 2007.
13. A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis”, *IEEE Trans. on Neural Networks*, Vol. 10, No. 3, pp.626–634, 1999.
14. FastICA MATLAB Package online at http://www.cis.hut.fi/projects/ica/fastica [last access date: 01 June 2013]
15. Larrañaga, P. , Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J.A., Armañanzas, R., Santafé, G., Pérez, A., Robles, V. , “Machine learning in bioinformatics”, *Briefings in Bioinformatics*, vol.7, no.1, pp. 86-112, 2006.
16. Vapnik V. ,The Nature of Statistical Learning. Springer-Verlag, 1995.