A Study on Classification of Sounds

J. Stephy Francisca[1],V. Ramalingam[2]
Department of CSE
[1] PG Scholar,Annamalai UniversityChidambaram, Tamil Nadu
[2] Professor,Annamalai UniversityChidambaram, Tamil Nadu

Abstract-*Nowadays, the classification of sounds is a challenging area of research in signal processing field. In this paper, sounds are classified as bird sound or non-bird sound. A total of eighty (80), wave files of sounds are collected from the Internet. Out of these 80 wave files, 40 files are from bird sounds and remaining 40 files are from the non-bird sounds. The sound signal is then analyzed in order to extract the Mel Frequency Cepstral co-efficient (MFCC) which is identified as main features. The MFCC features are identified and extracted from input sounds. Training and test data prepared with four–fold cross validation. Then classification of sounds into bird or non-bird is implemented using Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN) classifiers. Finally performance in the classification of sounds is computed by finding confusion matrix, Receiver Operating Characteristic (ROC), Error histogram. Experimental results show that the average accuracy of bird or non-bird sound classification using SVM is 70%, and that of BPNN is 95%.*

Keywords- MFCC, SVM, ROC, BPNN, Error Histogram.

I. INTRODUCTION

The sound is the important form of signal. Generally sound is basically a pattern formed in the vibration or movement of molecules of air. When a sound is made, air molecules move out from the source in waves. Sound types are voice, sound effects, music. The waves radiate out 360 degrees and in 3 dimensions from the source until it dissipates. In recent years many techniques for classifying bird species based on recorded vocalizations have been pro-posed and developed. The most successful techniques are based on manual inspection and labeling of bird sound spectrograph by experts, but this process is tedious and dependent upon the subjective judgment of the observer. The reliability of classification can be improved if a panel of experts is used, but this is expensive, time consuming and unsuitable for real time classification. Automatic recognition of birdsong, syllables from continuous recordings. Their method directly compared the spectrograms of input bird sounds with those of a set of predefined templates representative of categories chosen by the investigator. Audio classification systems typically begin by extracting acoustic features from audio signals. Such features often pertain to individual frames (i.e., very short segments of signal).

In rapidly changing world, the monitoring of birds communities is becoming increasingly important. Birds are able to produce a wide variety of sounds. The main objective of this research work is to develop and to implement an efficient sound classification system for supporting to classify (detect) bird or non-bird sound. To achieve this one has to identify relevant features from the given sound files. Extract relevant features from the given sound files Classify given sound into bird sound or non-bird sound using pattern classification techniques, such as, Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN).

The rest of this paper is organized into five sections: Section 2 provides the literature survey. Section 3 presents the proposed system with feature extraction. Section 4 shows the experimental results. Section 5 concludes the paper with obtained results.

II. LITERATURE SURVEY

The objective stated in [1], is to focus on the choice of audio features used as input data. It must work with high accuracy across large number of possible species, on noisy outdoor recordings and at big data scales. In this paper we study the interaction between dataset characteristics and choice of feature representation through further empirical analysis. 12 different features are represented, 4 different datasets are used in this paper. Each dataset consisting of long dawn-chorus recordings with a substantial amount of audio. The feature values are approximately de-correlated from each other, and they give a substantially dimension reduced summary of spectral data. For largest dataset, feature learning led to classification performance up to 85.4%.

In [2], theobjective is to peak tracking of spectral analysis data demonstrates the usefulness of the sum-of-sinusoids model for rapid automatic recognition of isolated bird syllables. The LPC and MFCC are used to extracting linear predictive code. Then the two different classifiers are used to classify the data. The classifiers such as, DTW and HMM are perform well for two specific birds in a low noise environment. In the proposed method use extract a variable

number of spectral peak track from one syllable of the desired bird sound. Totally 28 sounds are used. Overall performance of this project is 95%. Future research is continuing on method to classify such rapidly varying signals and also to identify broad band and noisy sounds.

The analysis of Bird and non-Bird sound classification is a challenging and an important area of research in speech processing. From the Internet the Birds and non-Birds sounds of wav files are collected. Further it presents a method for the identification and classification of Bird sound and Non-Bird sound using SVM. Mel-Frequency Cepstral Coefficients (MFCC) features extracted from sounds are used for this purpose [6, 7].

The artificial neural network (ANN's) is used to perform mapping of MFCC feature of source database (Inputs). The results of Sound classification evaluated using subjective and objective measures, confirm that an ANN based performance measure gives better result for classification [10].

The work at [12] presents a method for the identification and classification of Bird sound and Non-Bird sound using Artificial Neural Network. The Back Propagation Algorithm used for Bird and Non-Bird sound classification. Mel-Frequency Cepstral Coefficients (MFCC) features extracted from sounds are used for this purpose [12].

III. PROPOSED SYSTEM AND SOUND FEATURE EXTRACTION

The proposed Support Vector Machine (SVM) is used to model the spectral variability's in sound and it is used to perform classification. After the features are extracted, for each segment is categorized as bird sound or non-bird sound using Support Vector Machine (SVM) classifier. The Figure 1 shows the proposed system for classifying sounds in the form of bar diagram.

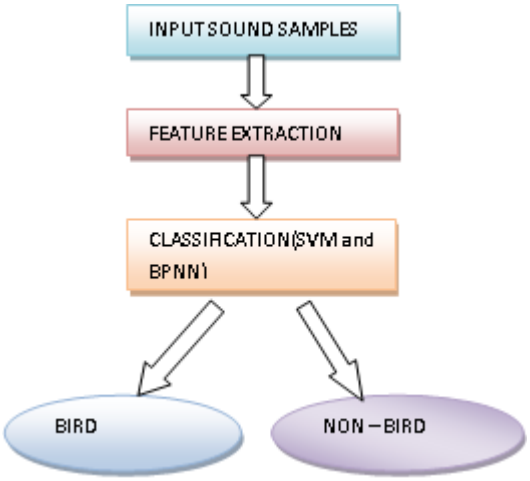


Figure 1 Block diagram of proposed system for classifying sounds.

It consists of four stages, namely,

(i) MFCC feature extraction.
(ii) Bird or non-bird sound classification using SVM.
(iii) Bird or non-bird sound classification using BPNN
(iv) Calculate average performance for sound classification.

3.1 Mel Frequency Cepstral Co-efficients

MFCC is the most widely used feature extraction technique. These coefficients represent audio based on perception and are derived from the Mel frequency Cepstral. Feature extraction using Mel Frequency Cepstral Co-efficient (MFCC), such as the Mel-Frequency Cepstral Co-efficient features which represent a manually designed summary of spectral information. The Figure 2 shows the various steps involved in convertin of input signal into MFCC.

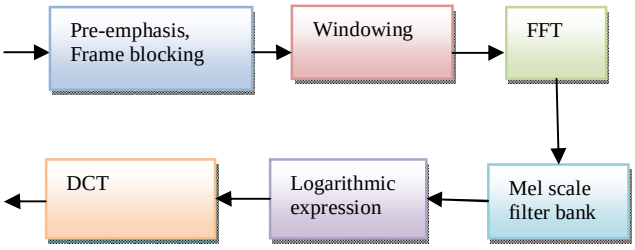


Figure 2 Block diagram to convert input signal into Mel Frequency Cepstrum Coefficient.

The aim of Pre-emphasis and Frame blocking stage is to boost the amount of energy in the high frequencies. This step processes the passing of signal through a filter which emphasizes higher frequencies. This process will increase the energy of signal at higher frequency. The signal is splitted into

several frames such that each frame in the short time can be analyzed instead of analyzing the entire signal, also on overlapping is applied to frames.A window function to smoothly attenuate both ends of the signal towards zero, this unwanted artifacts can be avoided. The window function is used to smooth the signal for the computation of the DFT.the Fourier Transform is to convert the convolution of the glottal pulse U[n] and the vocal tract impulse response H[n].Mel filter bank is used for two principal reasons: Smooth the magnitude spectrum such that the pitch of a speech signal is generally not presented in MFCC's.Logarithm compresses dynamic range of values. Makes frequency estimates less sensitive to slight variations in input.

3.2 Support Vector Machine

Support Vector Machine is used to classify the group of data as either bird or non-bird sounds depending on the feature values. Support vector machines (SVM's) are a set of related supervised learning methods used for classification and regression. SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors. If the data are linearly separated, SVM trains linear machines for an optimal hyper plane that separates the data without error and into the maximum distance. The original input space can always be mapped to some higher-dimensional feature space where the training set is separable

Figure 3 shows a support vector machine construct a hyper plane or set of hyper planes in a higher or infinite dimensional space, which can be used for classification, regression or other tasks.

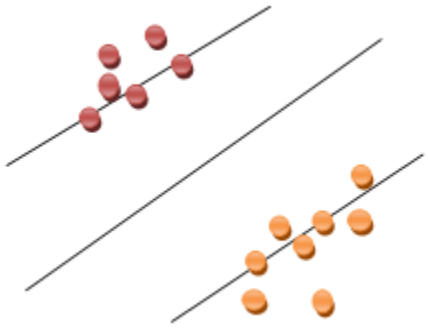
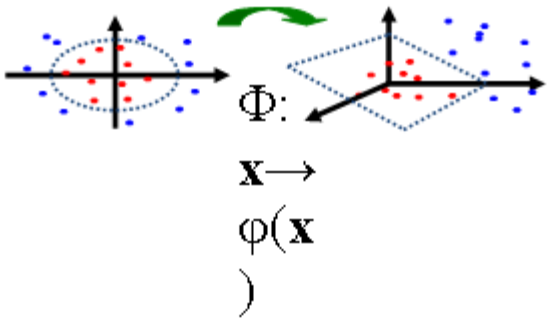


Figure 3 Optimal Hyper planes Maximizing Margin and Support Vectors Machine.

The support vectors are the (transformed) training patterns. The support vectors are (equally) close to hyper plan. The support vectors are training samples that define the optimal separating hyper plane and are the most difficult patterns to classify. Figure 4 shows a Data transformation of 2-Dimension to 3-Dimension by SVM.



$\Phi: \mathbf{x} \rightarrow \varphi(\mathbf{x})$

Figure 4. Data transformation by (2-D to 3-D) SVM.

3.3 Back Propagation Neural Network (BPNN)

An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. NN are constructed and implemented to model the human brain. It performs various tasks such as pattern-matching, classification, optimization function, approximation, vector quantization and data clustering. These tasks are difficult for traditional computers. Figure 5 shows three layer neural network.

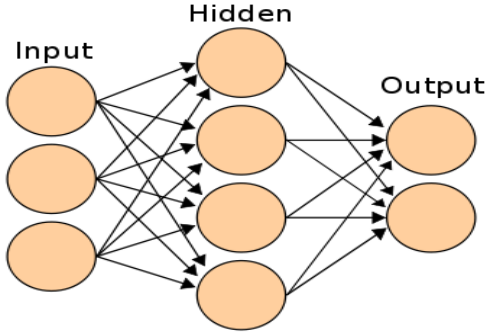


Figure 5 Three layer neural network**.**

Neural network: input / output transformation is given in Equation (1).

$$\mathbf{Y}_{\text{out}} = F(x, W) \quad (1)$$

where x = Input, W = Weights, Y out= Output.

The Back Propagation Neural Network is used for variable selection, formation of training, testing and validation sets, neural network architecture, evaluation criteria and neural network training.The back propagation algorithm (Rumelhart and McClelland, 1986) is used in layered feed-forward ANNs.

This means that the artificial neurons are organized in layers, and send their signals "forward", and then the errors are propagated backwards. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers. The back propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. The idea of the back propagation algorithm is to reduce this error, until the ANN learns the training data. The training begins with random weights, and the goal is to adjust them so that the error will be minimal. The McCullough-Pitts model spikes are interpreted as spike rates. Synaptic strength is translated as synaptic weights. Excitation means positive product between the incoming spike rate and the corresponding synaptic weight. Inhibition means negative product between the incoming spike rate and the corresponding synaptic weight.

IV. EXPERIMENTAL RESULTS

Source of Data:

In this thesis work, 40 wave files are pertaining to birds sound and 40 wave files pertaining to non-birds sound with a total of 80 wave files are collected from following websites. The bird and non-bird sounds data are archived (as wave files) at,
(i) http://www.grsites.com/archive/sounds and
(ii) http://soundbible.com/tags-bird.html

All sounds are downloaded in wave form. The 40 bird sounds in wave forms are consisting of (crow, peacock, parrot, cuckoo, owl, eagle …etc). And the 40 non-bird sounds in wave forms are consisting of (tiger, lion, bear, deer, dog, elephant, cow …etc).

Feature Extraction:

From the various literatures reviewed for classifies sound into bird or non-bird are can arrive at MFCC can be used as features important to classify the sounds. From these sounds, 12 Mel Frequency Cepstrum Co-efficient are extracted. The feature set is of size 12 (say 1, 2…, n).

N-Fold Cross Validation:

To prepare data for training and testing, an N-Fold Cross validation is done. The total sample data in divided into 4 folds namely, fold1 (f1), fold2 (f2), fold3 (f3), and fold4 (f4). Each fold fi (fi, i=1, 2, 3, 4) contains 20 samples (total sample (80)/no. of folds (4)). Training and test data are prepared by using the Table 5. 1. Out of these 80 sample wave files training data contains 60 wave files and testing data contains 20 wave files. The training and testing data by N-fold cross validation is given in Figure 6.

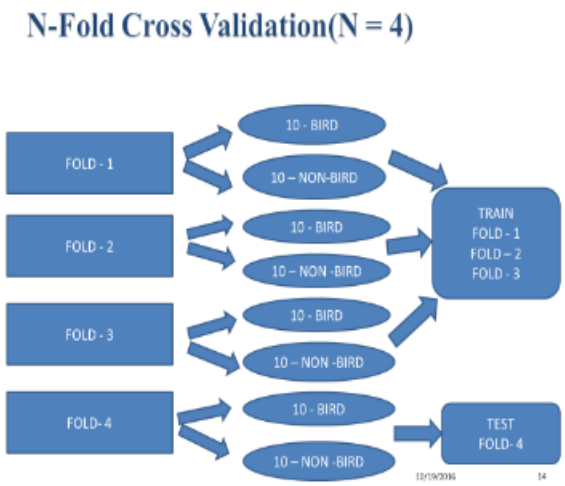


Figure 6. N- Fold Cross Validation.

Experimental Results for Bird Or Non-Bird Sound Classification Using SVM:

The support vector machine (SVM) Model is prepared by using the training and testing data which are given in table 1.

Table 1. Preparation of Training and Testing data for Four-fold cross validation

FOLD NO	***TRAINING DATA***	***TESTING DATA***
	No. of wave files	
	60	***20***
1	**{f1, f2, f3}**	**{f4}**
2	**{f1,f2,f4}**	**{f3}**
3	**{f1,f3,f4}**	**{f2}**
4	**{f2,f3,f4}**	**{f1}**

Table 2 shows experimental results given by SVM for different folders by correctly classified and misclassified in

birds and non-birds in percentage. From the Table 2 one can observe the fold 3 and fold 4 have highest percentage (80%) when compare the fold 1 and fold 2 by SVM. Overall percentage of classification of sounds by SVM is 70%.

Table 2. Percentage of classification of sound by SVM for different folder

Name of the folder	*Percentage of correct classification (%)*
FOLD 1	**65**
FOLD 2	**55**
FOLD 3	**80**
FOLD 4	**80**
OVERALL AVERAGE	**70**

Performance Measure for Classification Using BPNN:

In order to evaluate the performance of the classifier BPNN of ANN in classifying sounds, and to allow comparisons, several ratios have been taken into account. In order to analyze various bird sound classification of Performance measures using the confusion matrix.The diagonal cells show the number of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The network outputs are very accurate, The high numbers of correct responses in the green squares. The low numbers of incorrect responses in the red squares. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red).

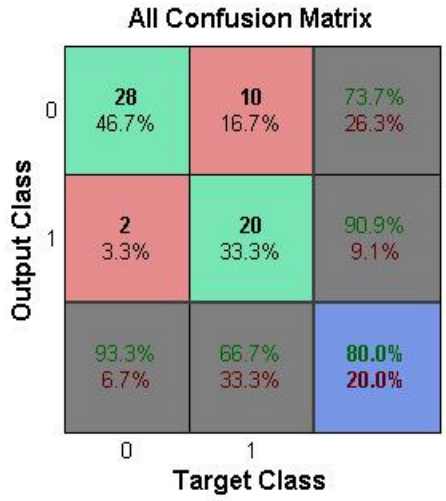


Figure 7. Overall confusion matrix used in BPNN.

The Figure 7 shows that the overall confusion matrix represents the overall performance of training, testing, and validation. It gives 80% of sounds are correctly predicted and remaining 20% of data is incorrectly predicted using BPNN. The Receiver Operating Characteristic (ROC) curve in the neural network pattern recognition tool. The receiver operating characteristic is a metric used to check the quality of classifiers. For each class of a classifier, roc applies threshold values across the interval [0,1] to outputs. For each threshold, two values are calculated, the True Positive Ratio (the number of outputs greater or equal to the threshold, divided by the number of one targets), and the False Positive Ratio (the number of outputs less than the threshold, divided by the number of zero targets).The ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (1 - specificity) as the threshold is varied. A perfect test would show points in the upper-left corner, with 100% sensitivity and 100% specificity.

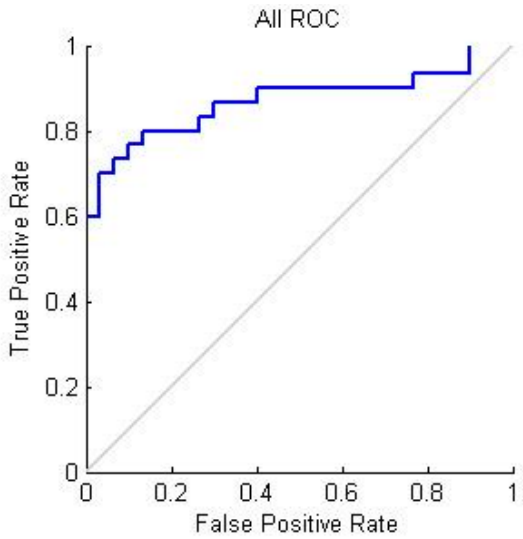


Figure 8. Overall performance of Receiver Operating Characteristic (ROC) for data used in BPNN.

The Figure 8 shows the overall performance of receiver operating characteristic using BPNN. The colored lines in each axis represent the accuracy rate of correctly predicted (80%) sounds. The Histogram block computes the frequency distribution of the elements in the input.

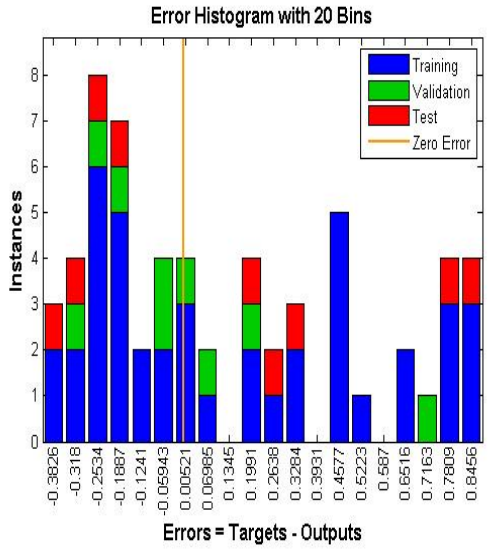


Figure 9. Error Histogram.

Error histogram represents the mean squared error is the average squared difference between outputs and targets. Lower values are better. Zero means no error. Figure 9 shows the ratio of error histogram. And Table 3 shows testing results of bird and non-bird sounds classification using BPNN. The block distributes the elements of the input into the number of discrete bins specified by the Number of bins parameter, n. The training data error ratio is representing in blue color bin. The testing data error ratio is representing in green color bin and the testing data errors in representing in red color bin. In (0.00521) represents the zero error in the input data. In point (0. 4577) represents no error occurs in both validation and testing. In (0. 7163) represents no error occurs in training and testing.

Table 3. BPNN Testing Results is sounds into bird or non-bird

FOLD	TRAINING DATA	TESTING DATA	PERFORMANCE (%)
FOLD 1	42	18	83
FOLD 2	42	18	88
FOLD 3	42	18	55
FOLD 4	42	18	80

The average performance of classifying sound into bird or non-bird using BPNN is 80%.

V. CONCLUSION

The classification of sounds is a challenging area of research in signal processing field. In this thesis sounds are classified as bird sound or non-bird sound and calls. Further, eighty (80), wave files of sounds are collected from the Internet. Out of these 80 wave files, 40 wave files are from bird sounds and remaining 40 wave files are from the non-bird sounds. The sound signal is then analyzed in order to extract the Mel Frequency Cepstral co-efficient (MFCC). The MFCC features are identified and extracted from input sounds. All input sounds are wav files. Training and test data prepared with four – fold cross validation. Then classification of sounds into bird or non-bird is implemented using Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN) classifiers. Finally performance in the classification of sounds is computed by finding confusion matrix, Receiver Operating Characteristic (ROC), Error histogram. This work demonstrates describes the implementation of Support Vector Machine and Artificial Neural Network (BPNN) for classifying Bird and Non-Bird sound using sound samples. From experimental results, Classification of Bird and Non- Bird sound using SVM gives 70% accuracy. But ANN (BPNN) is gives better performance 80% accuracy.

REFERENCES

[1] Dan Stowell and d Mark D. Plumbley, “Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning”, Center for Digital Music, Queen Mary University of London, 2014.

[2] Zhixin Chen and Robert C. Maher, “Semi-automatic classification of bird vocalization using spectral peak tracks” , Montana State University, 2006.

[3] Panu Somervuo, Aki Harma, and Seppo Fagerlund, “Parametric Representation of Bird Sounds for Automatic Species Recognition” , IEEE Transaction on Audio Speech, and language Processing, vol. 14, no. 6, 2006.

[4] Maria Sandsten, Marelie Grobe Ruse and Martin Jonson , “Robust feature representation for classification of bird song syllables” , Sandsten et al. EURASIP journal on Advances in Signal Processing, 2016.

[5] Miguel A. Acevedo, Carlos J. corrada-Bravo, Hector Corrada-Bravo, Luis J. Villanueva-Rivera, T. Mitchell Aide, “ Automated classification of bird and amphibian calls using machine learning: A comparison of methods” , University of Puerto Rico, 2009.

[6] Ilyas Potamitis, “ Unsupervised dictionary extraction of bird vocalization and new tools on assessing and

visualising bird activity" , Technological Education Institute of Crete, 2015.

[7] Babacar Diop, Dame Diongue, Ousmane Thiare, "Bird Sounds detection using Normalized Audio power Sequences" , 4th International Conference on Modeling and Simulation, 2015.

[8] Kun Qian, Zixing Zhang, Fabien Ringeval, bjorn Schuller, "Bird sounds classification by large scale aqoustic feartures and extreme learning machine" , IEEE Global conference on Signal and Information Processing, 2015.

[9] Ilyas Potamitis, Stavros Ntalampiras, Olaf Jann, Klaus Riede, "Automatic bird sound detection in long real-field recordings:Application and tools" , Technological Education Institute of Crete. 2014.

[10] Seppo Fagerlund, Unto K.Laine, "New parametric representation of bird sound for automatic classification" , IEEE International Conference on Acoustic, Speech and Signal Precessing, Aalto university, 2014.

[11] Sakurako Yazawa, Masatoshi Hamanaka, Takehito Utsuro, " A novel approach to separation of musical signal sources by NMF" , IEEE International conference , 2014.

[12] Miguel F. M. Lima, J. A. Tenreiro Machado, "Towards a classification scheme for musical sounds", IEEE international conference for signal processing algorithms, architecture, arrangement and application, Portugal, 2013.

[13] Sha -sha Chen, Ying Li, "Automatic Recognition of Bird Songs Using Time-frequency Texture" , 5th International Conference on Computational Intelligence and Communication Networks, 2013.

[14] Tim Fischer, Johannes Schneider, Wilhelm Stork, "Classification of breath and snore sounds using audio data recorded With smart phones in the home environment" , FZI Research Center for Information Technology, IEEE International conference, 2016.

[15] Fernando Rios-Gutierrez, Rocio Alba-Flores, Spencer Strunic, "Recognition and Classification of Cardiac Murmurs using ANN and Segmentation" , Georgia Southern University, IEEE conference paper, 2012.

[16] Vishal Kumar, S.N. Sharma, D.K. Shakya, "Detection heart sounds s1 and s2 using optimized s-transform and Back-propagation algorithm" , Samrat Ashok Technological Institute, IEEE international conference paper, 2015.