Comparison of Classifier Performance in their Ability to Classify Respiratory Sounds

Shree Shyamalee V R, Vaishali Gupta, Vyshnnavi Parthasarathy, Pravin Kumar S, Mahesh V
Department of Biomedical Engineering, SSN College of Engineering, OMR, Chennai-603110, India

ABSTRACT
Concepts of machine learning are potentially useful tools in reducing human effort and time. In rural India, there is a dearth in accessibility and affordability of excellent and sound medical diagnostic facilities. Implementing machine learning concepts to predict the presence of an illness as a part of an automated diagnostic system can go a long way in bridging the gap. As a part of developing one such system to diagnose respiratory illness using respiratory sound, an attempt has been made to analyze the performance of a set of six classifiers that include the nearest neighbor, the parzen window, the support vector machine, the relaxation batch margin, the relaxation single sample margin and the least square classifier with respect to their ability to classify the healthy and the non-healthy subjects. Four sets of features have been used, namely the statistical feature set, feature set based on the Gray Level Co-occurrence Matrix (GLCM) obtained from the spectrogram of the sound, the Mel- Frequency Cepstral Coefficients (MFCC) and Wavelet Packet Decomposition Coefficients. These features have been employed individually and in combinations to train and test the classifier performance. The performance has been interpreted by obtaining the confusion matrix and parameters such as sensitivity, specificity, precision, negative predictive value and accuracy and also by plotting the Receiver Operating Characteristic (ROC) curve for each classifier. Based on sensitivity that measures the ability of a classifier to identify the true class correctly and the accuracy that measures the correctness of the predicted class, it is inferred that the wavelet packet decomposition coefficients and MFCC are good features in characterizing respiratory sound. Further, in terms of the classifier, the relaxation classifiers, both batch margin and single sample margin and the support vector machine that classifies using a hyper plane yielded appreciable results with a maximum accuracy of 0.83 in clear contrast to that of Parzen and Nearest Neighbor. Further the result demonstrates that the classifiers used in this work will assist the physician in diagnosing the abnormal nature of respiratory sound and the system can be used as a mass screening tool.

General Terms
Machine learning, Classifiers, Features

Keywords
Classifier; ROC curve; Support vector machine; Respiration;Diagnostic tool

1. INTRODUCTION
Respiratory sounds, breath sounds or lung sounds refer to the specific sounds generated by the movement of air through the respiratory system. Breath sounds originate in the large airways where air velocity and turbulence induce vibrations in the airway walls. These vibrations are then transmitted through the lung tissue and thoracic wall to the surface where they may be heard readily with the aid of a stethoscope. Respiratory sounds can be classified into normal and abnormal. The abnormal sounds are found to accompany different respiratory illnesses like asthma, bronchitis, and sleep apnea amongst others. Though abnormal sounds include a broad range of sounds, the most predominant ones include crackles and wheezes. Crackles are discontinuous, explosive, 'popping' sounds that originate within the airways occurring in the frequency range of 100-2000Hz [1]. They are heard when an obstructed airway suddenly opens and the pressures on either side of the obstruction suddenly equilibrates resulting in transient, distinct vibrations in the airway wall. Wheezes are continuous musical tones that are most commonly heard at end inspiration or early expiration. They result as a collapsed airway lumen gradually opens during inspiration or gradually closes during expiration. Wheezes can be classified as either high pitched or low pitched wheezes. High pitch wheezes are associated with disease of the small airways whereas low pitch wheezes are associated with disease of larger airways. Wheezes may be monophonic or polyphonic and occur anywhere between 60 to 1200 Hz [1, 2].

2. MATERIALS AND METHODS
2.1 Samples
Respiratory sound samples were obtained from the R.A.L.E repository [3] and the Littman Stethoscope lung sound collection [4] for the purpose of analysis and classification. Samples from a set of nonsmoking and healthy adult subjects were obtained using a suitable acquisition circuit.

Training Data Set - The training data set included a set of 11 sound samples out of which 8 were abnormal and 3 were normal. The 8 abnormal samples and 1 normal sample used were obtained from the R.A.L.E and Littman repositories. The other 2 normal samples used were recorded using the acquisition circuit developed earlier.

The classifier performance was measured by testing with a set of 2 abnormal and 4 normal samples obtained from the repository and via the hardware developed. All the sounds utilized during classification that were obtained via the acquisition hardware were recorded from non-smoking subjects who had no known respiratory illness. The samples were played before a physician so as to validate the classifier performance.

2.2 Features
Combinations of four feature sets were used in the training of the classifier. These features were obtained for both the set normal and abnormal sound samples. Statistical features, Gray Level Co-occurrence Matrix (GLCM), Mel Frequency Cepstral Coefficients (MFCC), Wavelet Packet Decomposition Coefficients (WPDC) were used singularly and in combinations to test the efficiency of the classifiers. Standard statistical features; mean, mode, median, deviations, entropy,

kurtoses, skewness, RMS, MSE were used. GLCM was calculated for the gray scale spectrogram of the sound to obtain the GLCM feature set made of 23 features [5]. MFCC were calculated for the sampled power spectra [6]. Wavelet coefficients were calculated for decomposed signal samples using the Debauchies wavelet (level 3) and a Hanning window of length 512 with the window increment value set at 32 [7].

The feature values found were found to have larger values for abnormal samples in comparison to the normal samples used for training. Among the abnormal samples, there is an observed variation between different samples. This observed difference was found between low & high-pitched sound samples, as well as between fine & coarse samples.

2.3 Classifiers

A series of classifiers – the nearest neighbor, the Parzen window, Support Vector Machine (SVM), Batch Margin and Single Sample Margin (Relaxation – BM and Relaxation – SSM) and Least Square (LS) classifiers were used. The nearest neighbor (k=1) and the parzen window classifiers both being non parametric density estimator work by calculating the posterior probability density estimates by placing a cell of volume V around x and capturing k samples of which k_i samples belong to class w_i. If one were to fix the volume as a function of n ($V_n=1/\sqrt{n}$), it is the parzen window estimator. If Vn is expanded till a specified number of samples are captures (k is a function of n), it is the case of k-nearest neighbor classifier. The support vector machine works by calculating a separating hyper plane that maximizes the margin or the minimum distance of a feature vector of a particular class from the hyper plane.

The relaxation classifiers work on updating weights based on misclassified patterns. The weights are updated till the set of misclassified patterns becomes zero. In case of least square classifier, the error vector is defined and the algorithm tries to minimize the sum of the squared error value.

3. RESULTS AND DISCUSSION

For assessing classifying performance the receiving operating characteristic curves were plotted and parameters such as the sensitivity, negative predictive value and accuracy were calculated from the respective confusion matrices [8].

The receiver operating characteristic curve (ROC) plot measures the tradeoff between sensitivity and specificity. The closer the curve follows the left-hand border and then the top-border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test. The classifier performance measured through the respective ROC curves and parameters such as the sensitivity, specificity, accuracy values gave interesting results, with different feature sets working differently with each classifier.

While plotting the ROC graph for K-NN classifier as shown in figure 1, the curve corresponding to the GLCM features, corresponds to the 45-degree diagonal thereby showing the least discrimination ability. The wavelet decomposition features shows ideal curve characteristics with high Area under Curve (AUC). Though MFCC features have below par performance individually, its combination with statistical feature values shows high AUC value. While using the Parzen Window as a classifier, statistical features and its combination with MFCC shows best Classification capability of parzen window classifier based on the ROC curve can be seen from figure 2. Though these have a comparatively high accuracy value; MFCC when used to train the classifier corresponds to the diagonal thereby having less ability to discriminate between the two sets of data.

SVM shows best results with MFCC and WPDC. This is explained by the highest AUC visible corresponding to the curves associated with these feature sets. MFCC proves to be the most efficient feature set as its combination with statistical feature also proves to have high discrimination efficiency. The corresponding ROC curves are shown in figure 3. Using the Relaxation – Single Sample Margin (R-SSM) classifier, though statistical features show high AUC values, its curve characteristics which does not represent good discrimination ability as ideally extremely high sensitivity values are required with even low values of specificity. WPDC shows best curve characteristics with high AUC and is so also in its combination of features as can be seen from figure 4.

The Batch Margin Relaxation (R-BM) classifier gives best results with the combination of MFCC & WPDC while just the MFCC curves corresponds to the diagonal corresponding to least discrimination ability as can be seen from figure 5. WPDC gives the best results with the LS classifier with the highest AUC followed by the MFCC feature set which is clear from figure 6. Combinations of different feature sets have low sensitivity in the classification of respiratory sound samples as does not serve as the preferred training feature set.

The Batch Margin Relaxation (R-BM) classifier gives best results with the combination of MFCC & WPDC while just the MFCC curves corresponds to the diagonal corresponding to least discrimination ability as can be seen from figure 5. WPDC gives the best results with the LS classifier with the highest AUC followed by the MFCC feature set which is clear from figure 6. Combinations of different feature sets have low sensitivity in the classification of respiratory sound samples as does not serve as the preferred training feature set.

The graphs for comparison of parameters Accuracy, sensitivity and negative predictive value calculated from the corresponding confusion matrices are given in Figures 7-9. Higher sensitivity is observed with SVM, Relaxation – SSM and LS classifiers when used with MFCC, GLCM and a combination of Statistical and WPDC respectively. Parzen window followed by Nearest Neighbor Classifier are worst performers in terms of sensitivity. In terms of accuracy, all feature sets except GLCM give high accuracy (0.83) with at least one of the classifiers. Wavelet packet decomposition coefficients yield high accuracy with half the classifiers followed by the combination of WPDC with MFCC.

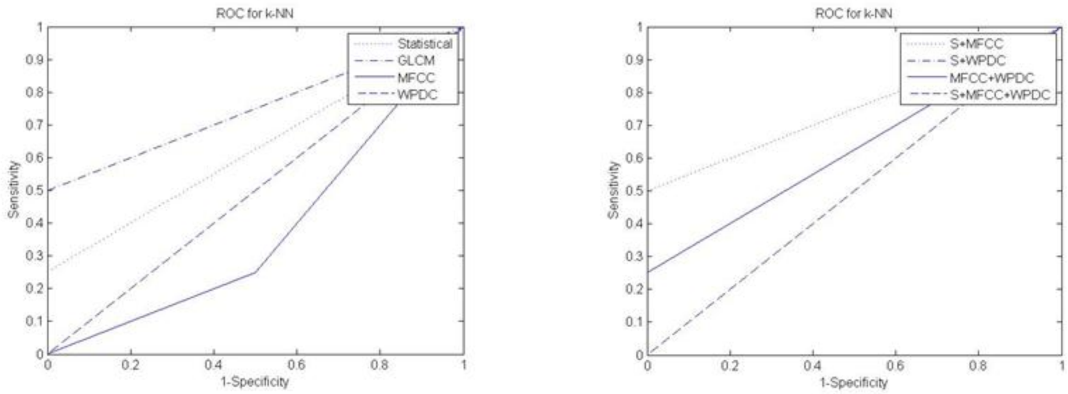

Fig 1: ROC Curves for K-NN

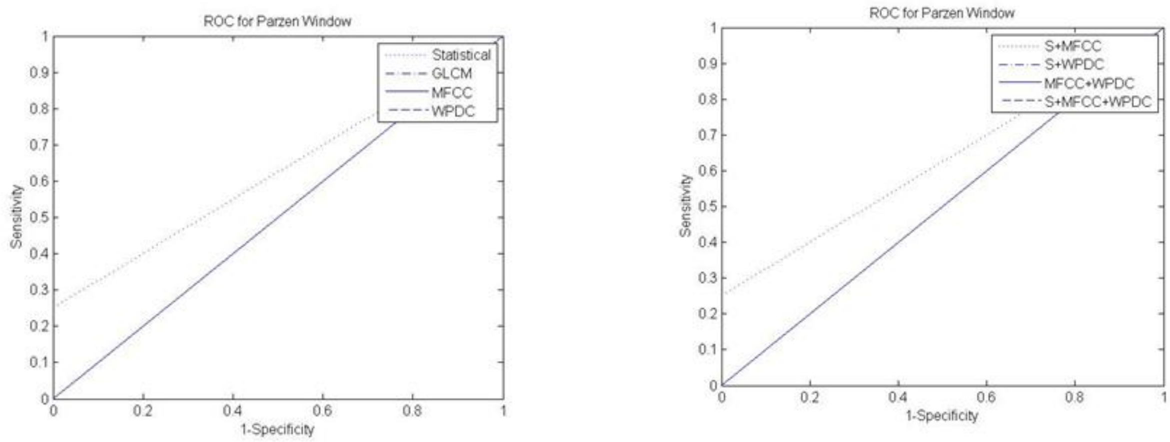

Fig 2: ROC Curves for Parzen Window Classifier

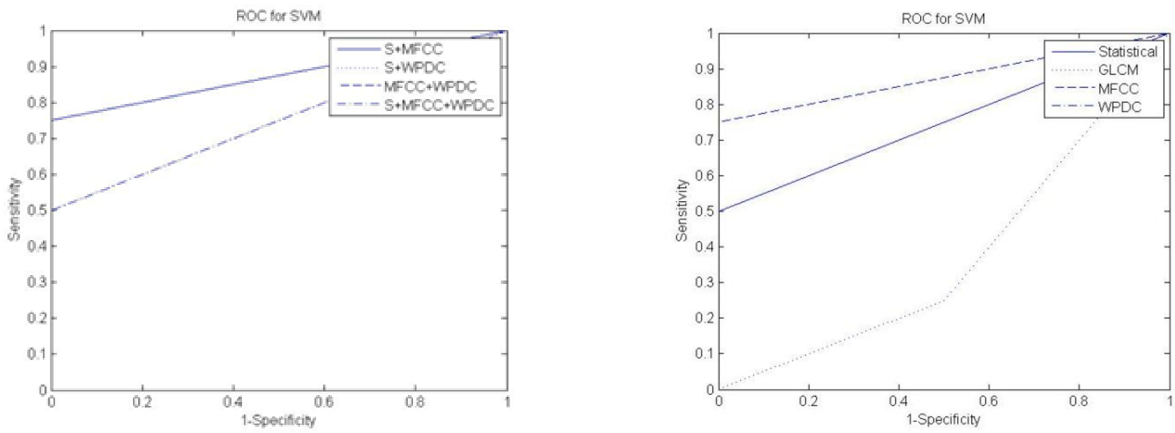

Fig 3: ROC Curves for SVM Classifier

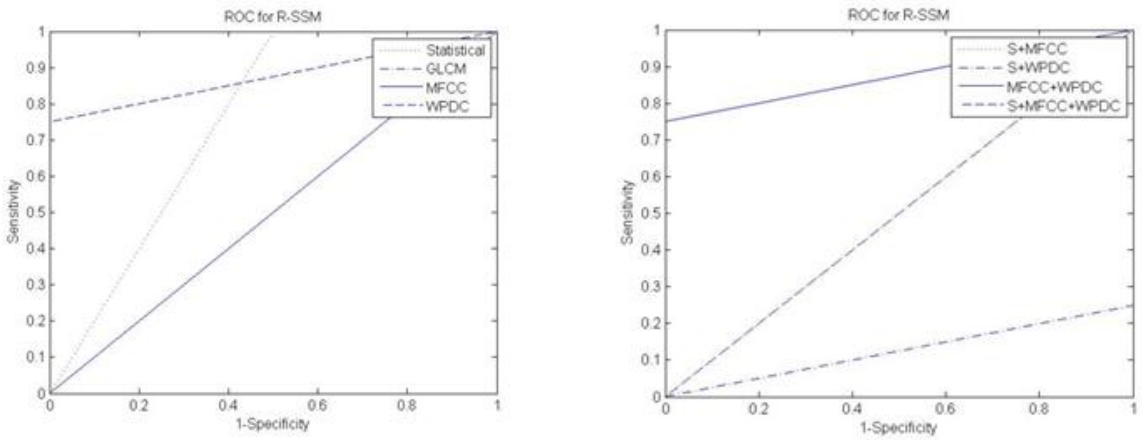

Fig 4: ROC Curves for R-SSM Classifier

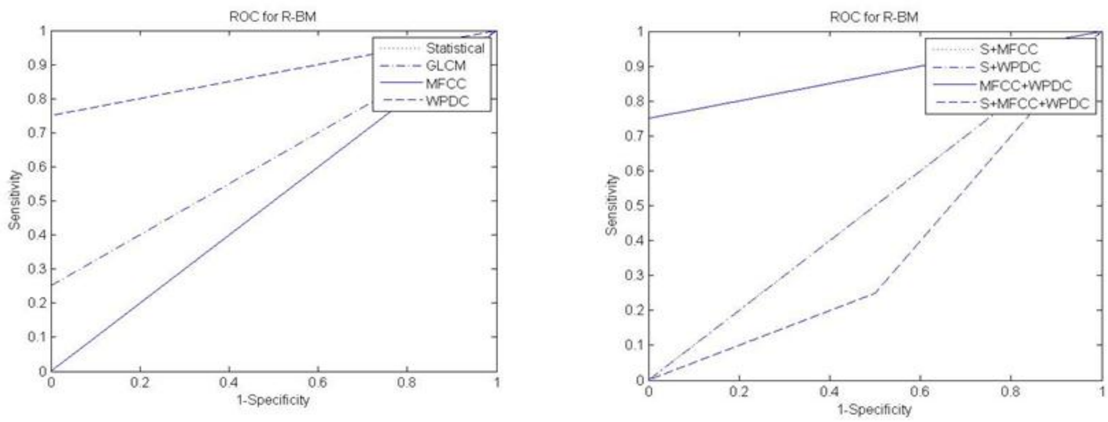

Fig 5: ROC Curves for R-BM Classifier

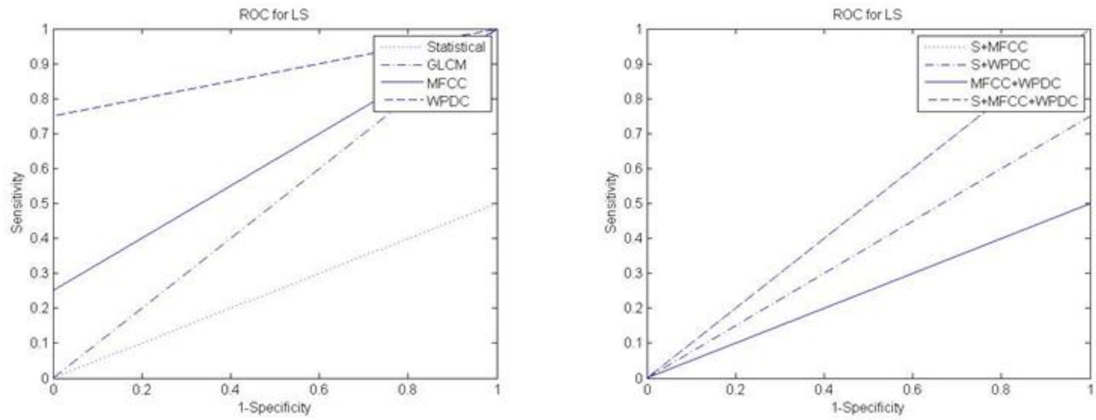

Fig 6: ROC Curves for LS Classifier

Since sensitivity measures the ability of a classifier to identify the true class correctly and the accuracy measures the correctness of the predicted class, one may infer that wavelet packet decomposition coefficients and MFCC are good features in their ability to characterize respiratory sound. Further, in terms of the classifier, the relaxation classifiers, both BM and SSM which operate based on updating weights till misclassification is zero and the support vector machine that classifies using a hyper plane yielded appreciable results in clear contrast to that of Parzen and Nearest Neighbor whose maximum accuracy values are 0.67 and 0.5 respectively. This could be the direct result of the way in which the corresponding algorithms function in order to classify samples. In case of SVM, higher the distance of particular feature from the plane, more accurate is the classification and clearly MFCC features are such that these can be clearly placed on either side of the plane and hence more accurate will be the predicted class. Hence, to develop a classification system to classify respiratory sounds, the best features to be used would be the MFCC and WPDC and the classifiers could be SVM, R-SSM, R-BM or LS.

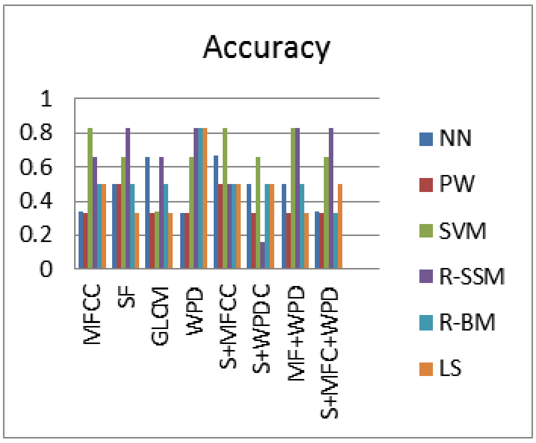

Fig 7: Graph for Accuracy

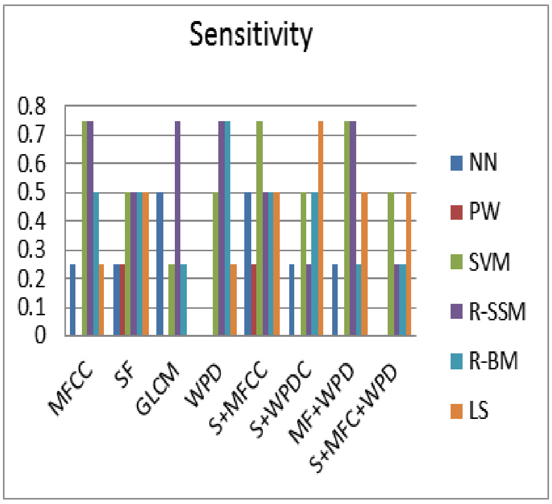

Fig 8: Graph for Sensitivity

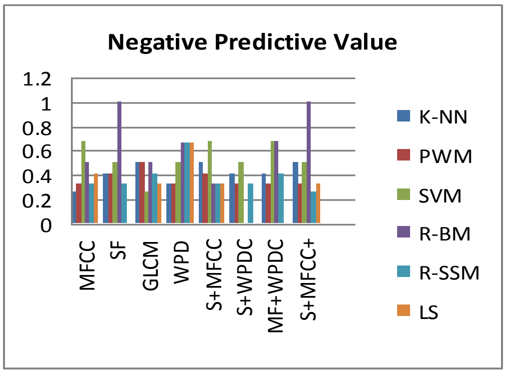

Fig 9: Graph for Negative Predictive Value

4. CONCLUSION

In this study, the performance of different classifier – feature sets was observed. Crackles and wheezes have been taken as training data for the abnormal case. The study can be expanded by increasing the number of training samples and at the same time making it a multi-class problem rather than binary to make a more specific diagnosis. Incorporating a highly accurate (100%) classification system as a part of an automated diagnostic system potentially means that only the sound samples of patients will be enough to predict any illness, which not only saves time but also eliminates the necessity of the physician. Hence, further exploration and standardization using a particular classifier and or two classifiers to confirm the outcome can result in development of a respiratory illness monitor with a good scope of commercialization.

5. REFERENCES

[1] Sovijarvi ARA, Maimberg LP, Charbonneau G, Vanderschoot J, Dalmasso F, Sacco C, et al. 2000. Characteristics of breath sounds and the adventitious respiratory sounds. European Respiratory Review, 10: 591–596.

[2] Forgacs P. 1967. Crackles and wheezes. Lancet, 2: 203–205.

[3] H. Pasterkamp. 1990. "RALE: A computer-assisted instructional package," Respiratory Care, 35: 1006.

[4] Respiratory sounds from www.littmann.com/wps/portal/3M/en_US/3M-Littmann/stethoscope/littmann-learning-institute/heart-lung-sounds/

[5] Pasterkamp H, Fenton R, Tal A, Chernick V. 1984. Tracheal vs lung sounds in acute asthma. Am Rev Respir Dis, 129: 256A.

[6] R Palaniappan and K Sundaraj. 2013. 'Respiratory Sound Classification using Cepstral Features and Support Vector machine'. Proceedings of 2nd Conference on Recent advances in Intelligent Computational Systems, IEEE; Kerala, India.

[7] Kandaswamy, A and Kumar, CS and Ramanathan, RP and Jayaraman, S and Malmurugan, N. 2004. 'Neural classification of lung sounds using wavelet coefficients'. Computers in Biology and Medicine, 34(6): pp.523-37.

[8] Kahya, YP and Bayatli, E and Yeginer , M and Ciftci, K and Kilinc, G. 2003. 'Comparison of different feature sets for respiratory sound classifiers'. Proceedings of the 25th Annual International Conference of the IEEE EMBS; Cancun, Mexico: pp.28532856.