

A Computer-based Sound Recognition System for the Diagnosis of Pulmonary Disorders

A. E. El-Alfi

Computer Science Department
Mansoura University, Egypt

A. F. Elgamal

Computer Science Department
Mansoura University, Egypt

R. M. Ghoniem

Computer Science Department
Mansoura University, Egypt

ABSTRACT

This paper presents a computer-based sound recognition system for diagnosis of pulmonary disorders based on the interpretation of the lung sound signals (LSS). We propose a novel method of analysis of LSS using the Mel-frequency cepstral coefficients, the spectral and temporal parameters estimated from the frequency subbands of the discrete wavelet transform. A Linde Buzo Gray (LBG) clustering neural network model is developed for classifying the LSS to one of the six categories: normal, wheeze, crackle, squawk, stridor, or rhonchus. Experimental results demonstrate the effectiveness of the proposed system in detecting pulmonary disorders.

Keywords

Lung Sound Signals (LSS), Discrete Wavelet Transform (DWT), Mel-Frequency Cepstral Coefficients (MFCC), Spectral and temporal parameters, Linde Buzo Gray (LBG), Multi-Layer Perception (MLP) Network.

1. INTRODUCTION

Chest auscultation is an inexpensive and efficient way to evaluate pulmonary dysfunction. As the pathological changes of the lung produce characteristic sounds, auscultation gives direct information about the function of the lung. The conventional method of auscultation with a stethoscope has many limitations. It is a subjective process that depends on the physician's own hearing, experience, and ability to differentiate between different sound patterns. Moreover the stethoscope has a frequency response that attenuates frequency components of the lung sound signal above about 120 Hz and the human ear is not very sensitive to the lower frequency band that remains [1].

Lung sounds that are heard with the aid of a stethoscope can be classified into two categories: normal and adventitious sounds. The breathing-associated sound of a healthy person is called the normal lung sound. Normal lung sounds span in the frequency range 100–1000 Hz and are devoid of any discrete peaks. Adventitious breath sounds refer to extra sounds during a regular breathing cycle. The adventitious sounds that are commonly associated with respiratory disorders are wheezes, crackles, squawks, stridors, and rhonchi [2-4].

Wheezes are adventitious, continuous sounds having a musical character. Acoustically, they are characterized by periodic waveforms with a dominant frequency usually over 100 Hz and with a duration of ≥ 100 ms. Wheezing is a common sign of obstructive lung disease. Crackles are discontinuous adventitious lung sounds, explosive and transient in nature, and occur frequently in cardiorespiratory diseases. Their duration is less than 20 ms, and their frequency content typically is wide. Stridors are very loud

wheezes, which are the consequence of a morphologic or dynamic obstruction in larynx or trachea. Stridor is usually characterized by a prominent peak at about 1000 Hz in its frequency spectrum. Squawks are short inspiratory wheezes that occur primarily in restrictive lung diseases. They always occur along with crackles, and often begin with a crackle. Their duration rarely exceeds 400 ms. Rhonchi often have a low-pitched, rattling, rumbling or bubbling quality. They may even sound similar to wheezes on occasion, and therefore may be difficult to distinguish from them. They may have an even more liquid sound than either wheezes or crackles, but they could also sound dry. The dominant frequency of rhonchi is less than 200 Hz [3-5].

In the past two decades, computer-based technology has evolved to evaluate the acoustic properties of respiratory sounds and to provide objective measurements that may circumvent the shortcomings of clinical auscultation. A number of studies have investigated the reliability of acoustic analysis and practical application of this method in clinical diagnosis [6-7]. Because of the stochastic nature of the lung sound, many parametric modeling methods and power spectral methods have widely been applied to characterize information on respiratory sound signals for computer-aided lung sound recognition, such as Autoregressive Coefficients (AR) [8], Linear Prediction Cepstral Coefficients (LPCC) [9], Mel-Frequency Cepstral Coefficients (MFCC) [2,10], wavelet coefficients [3,11], spectrograms [12], and higher-order spectra [13]. Based on the specific feature representation, various pattern recognition algorithms, such as Artificial Neural Networks (ANN) [2,3,14], Gaussian mixture model [2,10], and 2D bilateral filter [15], have been developed in lung sound classification.

In this work, a computer-based sound recognition system has been proposed for the diagnosis of pulmonary disorders. Firstly, lung sounds are acquired with an electronic stethoscope plugged to a sound card on a portable computer. Secondly, Discrete Wavelet Transform (DWT) based denoising technique, namely wavelet shrinkage denoising, is applied. Then the task of signal analysis is performed using four techniques: (1) the discrete wavelet analysis; (2) the Mel-frequency cepstral analysis; (3) the spectral and temporal analysis; and (4) the proposed method. Finally, a Linde Buzo Gray clustering neural network model is developed for classifying Lung Sound Signals (LSS) and a comparison of classification efficiencies obtained using the different signal analysis methods was made.

2. METHODOLOGY

As shown in Fig. 1, the proposed system comprises the following functions: (1) signal acquisition; (2) signal pre-processing; (3) signal analysis; and (4) training and testing.

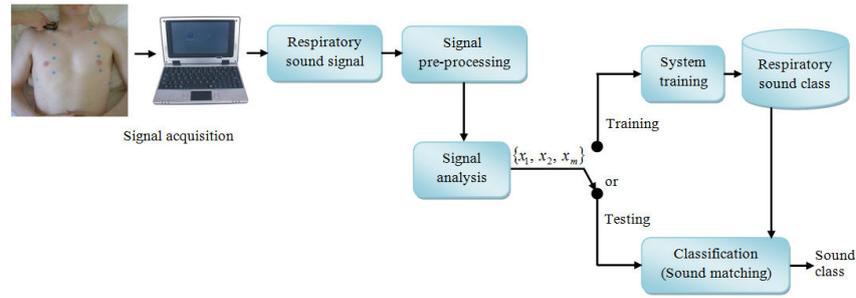


Fig 1: Block diagram of the proposed system

2.1 Signal Acquisition

During the signal acquisition stage, respiratory sounds are directly collected from patients and saved as waveforms for subsequent analysis. The lung sounds are captured on the chest wall of subject using an electronic stethoscope plugged to a sound card with 16-bit A/D conversion resolution on a portable computer. The classical stethoscope is modified by cutting the tube below the fork to the earpieces and inserting an electret microphone in the tube, which ensured accurate detection of lung sounds.

2.2 Signal Pre-processing

The noisy nature of the LSS is a serious impeding factor that prohibits further processing in order to identify useful diagnostic features. Hence denoising of the LSS is necessary for effective utilization for diagnosis. Since the frequency bands of these noises may overlap with the lung sounds. In this work, DWT based denoising technique, namely wavelet shrinkage denoising, was used. Wavelet shrinkage denoising consists of three steps:

- Obtain wavelet transform of the signal.
- Perform nonlinear shrinking of wavelet coefficients.
- Obtain inverse wavelet transform of the modified coefficients.

There are various types of wavelet shrinkage denoising techniques, classified according to the thresholding method used in nonlinear shrinking. SureShrink [16], which uses a hybrid of the universal threshold and the Stein's unbiased estimate of risk (SURE), was used for denoising in this work.

2.3 Signal Analysis

In this section, we will briefly describe the signal analysis methods which have been applied for classifying the LSS.

2.3.1 Discrete wavelet analysis

Wavelet transforms are rapidly surfacing in fields as diverse as telecommunications and biology. Because of their suitability for analyzing non-stationary signals, they have become a powerful alternative to Fourier methods in many medical applications, where such signals abound [2,3,11]. Fourier Transform (FT) does not provide enough information when used on non-stationary signals. FT determines only the frequency components of a signal, but not their location in time. The main advantage of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time-frequency resolution in all frequency ranges.

2.3.2 Mel-frequency cepstral analysis

The feature extraction process of MFCC [2,10,17] includes:

Step1: The LSS are blocked into short frames of N samples, with a predefined overlapping value.

Step2: Each individual frame of a signal X is windowed so as to minimize the signal discontinuities at the beginning and at the end of each frame and thus the spectral distortion is minimized. The window is defined as given below:

$$w(n); \text{ where } 0 \leq n \leq (N-1), \quad (1)$$

N is the number of samples in each frame. The result of windowing is the signal $y(n)$ and is defined as,

$$y(n) = x(n)w(n), \text{ where } 0 \leq n \leq (N-1), \quad (2)$$

the *hamming* window $y(n)$, used in this work is given by,

$$w(n) = 0.54 - 0.46 \cos \left[2\pi n / (N-1) \right], \quad 0 \leq n \leq (N-1), \quad (3)$$

the purpose of the window is to favour samples towards the centre of the window. This characteristic coupled with the overlapping attempts to smoothen the varying parameters.

Step3: Fast Fourier Transform (FFT) is applied to the windowed samples, which converts each frame of N samples from the time domain into the frequency domain, The FFT is defined on the set of N samples $\{X_n\}$ as:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi kn/N}, \text{ where } n=0, 1, 2 \dots N-1, \quad (4)$$

where X_n s are the complex numbers. The resulting sequence of X_n s is interpreted as given: (1) when $(n=0)$, it corresponds to zero frequency; (2) When $1 \leq n \leq (N/2-1)$, it corresponds to positive frequencies $(0 < f < F_S/2)$; and (3) When $N/2+1 \leq n \leq N-1$, it corresponds to negative frequencies $(-F_S/2 < f < 0)$. Here, F_S denotes the sampling frequencies. The obtained result is often referred to as 'spectrum' or 'periodogram'.

Step4: The Mel-frequency warping is implemented. Psychophysical studies have shown that human perception of the frequency contents of sounds does not follow a linear scale. MFCC are based on the known variation of the human ear's critical bandwidths with frequency. Thus the Mel-frequency scale is used which is the linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. The following approximate empirical relationship to compute the Mel frequencies (also called Mel's) for a given frequency f expressed in Hz is as given below:

$$\text{Mel}(f) = 2595 \times \log(1 + f/700), \quad (5)$$

in order to simulate the frequency warping process, we use a filter bank, one filter for each desired Mel-frequency component. That filter bank has a triangular band-pass frequency response, and the spacing as well as the bandwidth is determined by a constant Mel-frequency interval.

Step5: The log Mel-frequency spectrum is converted back to time domain using Discrete Cosine Transform (DCT). The resultant is called the MFCC and are calculated using:

$$c_n = \sum_{k=1}^K (\log S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{k} \right]. \quad (6)$$

2.3.3 Spectral and temporal analysis

The spectral analysis is involved in order to extract parameters that describe the frequency characteristics of the LSS (stationary analysis). The spectral parameters are based on the Short Time Fourier Transform (STFT) and are calculated for every short-time frame of a signal. The parameters selected for our analysis are: (1) the spectral centroid, (2) the spectral rolloff, and (3) the spectral flux. The temporal parameters that are directly extracted from the time domain, i.e., by the signal samples, are usually simple representations of the signal energy changes. Therefore, they can be used for signal discrimination based on energy differentiations.

2.3.4 Proposed method

LSS are non-stationary even when observed in a perfectly healthy normal subject. This non-stationarity is severe in case of abnormal subjects. Thus the mutually exclusive time and frequency domain representations are not highly successful in the diagnostic classification of LSS because their frequencies evolve with time. Hence, we need to represent LSS in two dimensions with time and frequency as co-ordinates.

Although wavelet Transform is a suitable technique for obtaining the time-frequency distribution of signals, the wavelet coefficients are often unsuitable to be used as features due to their number and their low-level meaning. In most cases, statistical features are derived from these coefficients to discriminate the different subbands of the transform. This approach is insufficient to capture the important time and frequency features of the signal. Thus a novel method is used to represent the frequency subbands of the DWT using the MFCC along with the spectral and temporal parameters. The MFCC obtained from the wavelet channels have the advantage that they can represent signals in an efficient way because of the Mel-frequency warping property which is an important property of a human ear. Hence, the MFCC mimic the perception of the physician's ear and its ability to differentiate between different sound patterns. The spectral and temporal parameters also correlate closely with auditory perception used in auscultation. The steps of our proposed methods are illustrated as follows:

Step1: the DWT is applied for the time-frequency analysis of the signal, yielding approximation and detail frequency subbands at different levels. The procedure of multi-resolution decomposition of a lung sound signal $x[n]$ is schematically shown in Fig. 2. Each stage of this scheme consists of two digital filters and two downsamplers by 2. The first filter, $h[n]$ is the discrete mother wavelet, high-pass in nature, and the second, $g[n]$ is its mirror version, low pass in nature. The downsampled outputs of first high-pass and low-pass filters provide the detail, D_1 and the approximation, A_1 , respectively. The first approximation A_1 is further decomposed and this process is continued to 7 levels. This computing is implemented using Biorthogonal Spline wavelet of order 1.5 (bior 1.5). The number of levels of decomposition was chosen based on the dominant frequency components of the LSS. Since the LSS do not have any useful frequency

components below 50 Hz, the number of levels was chosen to be 7. Thus the signal was decomposed into the details $D_1 - D_7$ and one final approximation, A_7 . The LSS were further characterized by the five subbands, $D_3 - D_7$, which correspond to the frequency band 43.07–1378.13 Hz. This is because the lung sound frequency spectrum ranges from 50 to 1000 Hz.

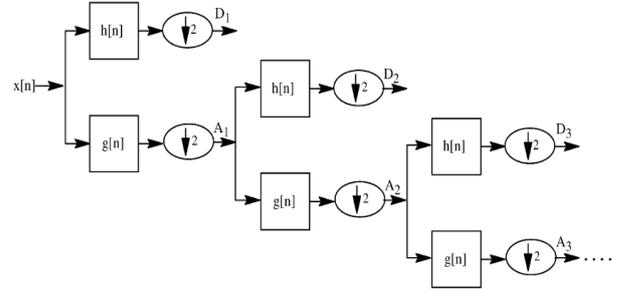


Fig 2: Subband decomposition of DWT implementation

Step2: Each one of the subband signals obtained from the DWT is split into a sequence of short-term frames of 23.22 ms (1024 samples per frame) with 50% overlapping. Each individual frame is windowed with a *hamming* analysis window.

Step3: The first five Mel-frequency cepstral coefficients, the spectral and temporal parameters shown in Table 1, are respectively estimated from each enframed subband signal, and two acoustic matrices are constructed and concatenated to form a final matrix of dimension $5 \times n$.

Step4: The constructed matrix at each level is reduced to a lower dimensional feature space using the LBG algorithm presented in the next section, and two centroids of clusters are estimated. Finally, the resulting centroids of clusters at each level are concatenated one-by-one to form the final vector [$2 \text{ centroids} \times 5 \text{ features} \times 5 \text{ levels} = 50 \text{ components}$] which is applied as input to the MLP network.

2.4 Training and testing

ANNs can learn complex functions from the input data and are relatively easy to implement in any application. On the other hand, a significant disadvantage of their usage is they usually high training time-need, which scales with the structural parameters of the networks and the quantity of input data. However, this can be done offline; the training has a non-negligible cost. To increase the speed of the training of the MLP network used for classification, we have proposed a new training procedure: instead of directly using the training data in the training phase, the data is first clustered using the LBG algorithm and the MLP network is trained by using only the centroids of the obtained clusters.

2.4.1 Linde-Buzo-Gray Clustering

LBG clustering or vector quantization (VQ) is a process of mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a cluster and can be represented by its centroid called a codeword. The collection of all codewords is called a codebook [2]. The block of samples which is subjected to quantization using the LBG algorithm may be:

- A vector of samples (one dimensional), typically for speech and audio coding.
- A block or matrix of samples, typically for image coding.

Table 1. The spectral and temporal parameters used in our work

Spectral and temporal parameters	
Spectral centroid: Is the center of gravity of the magnitude spectrum of the STFT:	
	$C_t = \frac{\sum_{n=1}^N M_t[n] * n}{\sum_{n=1}^N M_t[n]} \quad (7)$
where $M_t[n]$ is the magnitude of the Fourier transform at frame t and frequency bin n . The centroid is a measure of spectral shape and higher centroid values correspond to ‘brighter’ textures with more high frequencies.	
Spectral rolloff: Is the frequency R_t below which 85% of the magnitude distribution is concentrated:	
	$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^N M_t[n] \quad (8)$
The rolloff is another measure of spectral shape and yields higher values for high frequencies.	
Spectral flux: Is the squared difference between the normalized magnitudes of successive spectral distributions:	
	$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2 \quad (9)$
where $N_t[n]$ and $N_{t-1}[n]$ are the normalized magnitude of the Fourier transform at the current frame t , and the previous frame $t-1$, respectively. The spectral flux is a measure of the amount of local spectral change.	
Zero crossings rate: Is the rate of sign-changes of a signal, i.e., the number of times the signal changes from positive to negative or back, per time unit. It is defined as presented in the following equation.	
	$Z_t = \frac{1}{2} \sum_{n=1}^N \text{sign}(x[n]) - \text{sign}(x[n-1]) \quad (10)$
where the sign function is 1 for positive arguments and 0 for negative arguments and $x[n]$ is the time domain signal for frame t . Time domain zero crossings provide a measure of the noisiness of the signal.	
Energy entropy: Is a measure of abrupt changes in the energy level of the signal. It is computed by further dividing each-frame into K sub-frames of fixed duration. For each sub-frame, j the normalized energy e_j^2 is calculated, i.e., the sub-frame's energy, divided by the whole frame's energy. Therefore e_j^2 is a sequence of normalized sub-frame energy values, and it is computed for each frame. Afterwards, the entropy of this sequence is computed using:	
	$H_t = - \sum_{j=1}^K e_j^2 \cdot \log_2(e_j^2) \quad (11)$
where the value of the energy entropy is small for frames with large changes in energy level.	

The LBG clustering is implemented by the following recursive procedure [17]:

Step1: Design a 1-vector codebook; this is the centroid of the entire set of training vectors.

Step2: Double the size of the codebook by splitting each current codebook according to the rule:

$$y_n^- = y_n (1 - \varepsilon) \quad (12)$$

$$y_n^+ = y_n (1 + \varepsilon) \quad (13)$$

where n varies from 1 to the current size of the codebook, and ε is a splitting parameter ($\varepsilon = 0.01$).

Step3: Perform nearest-neighbor search. For each training vector, find the codeword in the current codebook that is closest in terms of VQ-distortion, and assign that vector to the corresponding cluster associated with the closest codeword.

Step4: Update the codeword in each cluster using the centroid of the training acoustic vectors assigned to that cluster.

Step5: Repeat steps 3 and 4 until the VQ distortion falls below a preset threshold.

Step6: Repeat steps 2, 3 and 4 until a codebook size of M is designed.

2.4.2 Linde-Buzo-Gray Clustering Multi-Layer Perception Neural Network (LBG-MLP)

One of the major problems of the classification performed by the LBG algorithm is the time taken to search for the closest codebook during the testing phase. In the training phase, an acoustical model is constructed for each class of sound and the models are stored in a database. For a given sound class, the resulting codewords constitute the codebook of this class

$\zeta = \{c_1, c_2, \dots, c_N\}$. Thus, there are K codebooks

$\zeta_1, \zeta_2, \dots, \zeta_K$ generated for K reference sounds (K is the number of sound classes). In the testing phase, the distortion between a set of testing feature vectors

$X = \{x_1, x_2, \dots, x_T\}$ and each trained codebook is computed, then an average quantization distortion Q_k to the

k th codebook (ζ_k) is performed [2,17], according to

$$Q_k = \frac{1}{T} \sum_{i=1}^T \min_{1 \leq j \leq N} d(x_i, c_j) \quad (14)$$

where $d(x_i, c_j)$ is the distortion measure between the input vector x_i and a centroid c_j of the k th codebook (ζ_k).

The unknown sound is then identified as the reference sound with the minimum average distortion measure

$$\hat{k} = \arg \min_{1 \leq k \leq K} \{Q_k\} \quad (15)$$

The fundamental advantage of our approach is that it introduces a clustering technique (LBG) into a discriminative system (MLP network). The capability of learning from examples, the ability to reproduce arbitrary nonlinear functions of input, and the highly parallel and regular structure of ANN make it especially suitable for respiratory sound classification task. Unlike the LBG based classification that requires one model per class, the MLP network needs only one model with a number of outputs equal to the number of the sound classes.

The MLP network that we have used for classification is characterized by D inputs, one hidden of N nodes, and K outputs. Each node j , in the hidden layer, receives the output of each node i from the input layer through a connection of weight $w_{j,i}^h$ and then produce a corresponding response z_j which is forwarded to the output layer. In fact, each node j performs a weighted sum which is transferred by a nonlinear function f_h according to

$$z_j = f_h \left(\sum_{i=0}^D w_{j,i}^h x_i \right), \quad j = 1, \dots, N \quad (16)$$

where $x_0 = 1$ is the bias input of the hidden layer.

In the same manner, the output of each node j , in hidden layer, is given by

$$y_j = f_o \left(\sum_{i=0}^N w_{j,i}^o z_i \right), \quad j = 1, \dots, K \quad (17)$$

where f_o is the transfer function, $w_{j,i}^o$ is the connection weight, and $z_0 = 1$ is the bias input of output layer. In this work, we use hyperbolic tangent sigmoid function $f_h(x) = (1 - e^{-2x}) / (1 + e^{-2x})$ for hidden layer and logistic sigmoid function $f_o(t) = (1 + e^{-x})^{-1}$ for output layer.

The connection weights $w = \{w_{j,i}^h, w_{j,i}^o\}$ are determined in the training phase using a set of inputs for which the desired outputs are known $\{(x_1, d_1), (x_2, d_2), \dots, (x_p, d_p)\}$. To accomplish this, the backpropagation algorithm is commonly used. When the i th pattern (x_i, d_i) is presented, the error of training is defined as [18]

$$E(W) = \frac{1}{2} \sum_{j=1}^K (d_{i,j} - y_{i,j})^2 \quad (18)$$

where $d_{i,j}$ and $y_{i,j}$ are the desired response and the actual response value of the output neuron j . Error is reduced by updating the weights using the gradient descent method:

$$\Delta w_{j,i} = -\eta \frac{\partial E}{\partial w_{j,i}} \quad (19)$$

where $0 < \eta < 1$ is the learning rate. A small value of η can guarantee convergence but involves a slow learning. On the other hand, a large value of this parameter involves a rapid learning but can lead to oscillation or even divergence. To overcome this limitation, we use the Resilient backpropagation which is one of the faster algorithms [19]. The basic principle of this algorithm is to eliminate harmful influences of the derivatives' magnitude on the weight update. The direction of the weight update is specified by using the sign of the derivative, while the size of the weight change is determined by a separate update value [19]. The backpropagation algorithm can be implemented in two different ways: sequential mode and batch mode [2]. In sequential mode, the weights are updated after each training example (pattern) is applied to the network. In batch mode, all the training examples, that constitute an epoch, are applied to the network before the weights are updated. For a given epoch, a cost function is defined as the average squared error of Eq. (18)

$$E_{av}(W) = \frac{1}{2P} \sum_{i=1}^P \sum_{j=1}^K (d_{i,j} - y_{i,j})^2 \quad (20)$$

where P is the number of the training examples (patterns) per epoch and K is the number of the outputs. The synoptic weight $w_{j,i}$, connecting neuron i to neuron j , is updated using

$$\Delta w_{j,i} = -\eta \frac{\partial E_{av}}{\partial w_{j,i}} \quad (21)$$

2.4.3 Formation of MLP datasets

Once the clustering process is completed, a set of K input/output pairs $D = \{X_k, Y_k \mid k = 1, 2, \dots, K\}$ is available. We have evaluated the performance of the classifier with four different datasets obtained using: (1) DWT features [2 centroids of clusters \times 5 levels = 10 components]; (2) MFCC features [4 centroids of clusters \times 5 MFCC = 20 components]; (3) spectral and temporal parameters [4 centroids of clusters \times 5 parameters = 20 components]; and (4) the proposed method. Each dataset is divided into two subsets, training set and test set.

- $D_{train} = \{X_k, Y_k \mid k = 1, 2, \dots, K_{train}\}$ is used to perform the MLP network training which consists of the determination of the MLP running parameters, i.e. the MLP network connection weights and biases.
- $D_{test} = \{X_k, Y_k \mid k = 1, 2, \dots, K_{test}\}$ is used to validate off-line classification ability and quality of the MLP network once the training has been completed.

3. EXPERIMENTAL RESULTS

Our database is constructed from 96 real respiratory sounds obtained from: (1) sounds recorded on healthy and pathologically affected patients, all recordings were made using our proposed system under the supervision of a physician specialized in pulmonology; and (2) the R.A.L.E. database-CD [20]. The collected database comprises six categories including: normal, crackle, wheeze, squawk, stridor, and rhonchus, each containing 16 sounds.

A computer-based sound recognition system has been developed for the diagnosis of pulmonary disorders. The system can record, save, and replay lung sounds and analyze them in time and frequency domains. The graphical user interface (GUI) of signal acquisition is shown in Fig. 3. The duration of the recorded breath sound can be determined and controlled through the slider shown on the GUI. All the acquired sounds can easily be stored on or retrieved from the hard disk. Zooming functions are provided for lung sound plots. The GUI of signal pre-processing is shown in Fig. 4, where the denoising result of a crackle lung sound signal is illustrated. The GUI of the discrete wavelet analysis is shown in Fig. 5. Fig. 6 shows the graphical representation for the detail coefficients of a DWT decomposed pulmonary crackle. The Mel-frequency cepstral analysis is shown in Fig. 7. The GUI of spectral and temporal analysis is shown in Fig. 8. Fig. 9 represents the results of the spectral and temporal analysis of a pulmonary crackle. The GUI of acoustic analysis using our proposed method is shown in Fig. 10.

In order to assess the generalization capabilities of the system, a 6-fold cross validation was performed. Optimum number of neurons in the hidden layer was investigated for ascertaining how changes in the number of neurons in hidden layer contribute to the performance of the classification system. The training algorithm, the parameters of the training algorithm, and the activation functions of the hidden and output layers were also determined by repeated simulation. The architecture and parameters of the applied MLP network have been elucidated in Table 2.

The classification efficiency was defined as the percentage ratio of the number of lung sounds correctly classified to the total number of lung sounds considered for classification [3]. The MLP classifier was evaluated using the four obtained datasets. The results are shown in Fig. 11. From the figure, we can see that the MLP classifier performance obtained by the proposed signal analysis method outperforms those obtained by the other methods. Therefore, the proposed signal analysis method combined to the LBG-MLP model is well adopted to classify LSS. Finally, experimental results show that the designed system is effective in detecting pulmonary disorders.

4. CONCLUSIONS

Auscultation of the chest via a stethoscope provides useful information to the physician for the diagnoses of pulmonary disorders. However, due to the subjectivity in auditory perception among physicians, and variability in their verbal descriptions of sound characteristics, fuzzy and qualitative nature of the diagnosis has become the major problem for this rewarding method. This paper presented a computer-based sound recognition system for the diagnosis of pulmonary disorders. The system presents the following features: (1) it is able to digitally record the lung sounds which are captured with an electronic stethoscope plugged to a sound card on a portable computer; (2) display the lung sound waveform for auscultation sites, (3) denoise and store the acquired sounds; (4) acoustically analyze them for quantitative and off-line

consultation that would aid the physicians to build a consensus view, leading to more contemplative examination; and (5) classify the LSS to one of the six categories: normal, wheeze, crackle, squawk, stridor, or rhonchus, using a proposed Linde Buzo Gray clustering neural network model. Experimental results show that classification efficiency obtained by the proposed signal method combined to the LBG-MLP outperforms those obtained by the other methods.

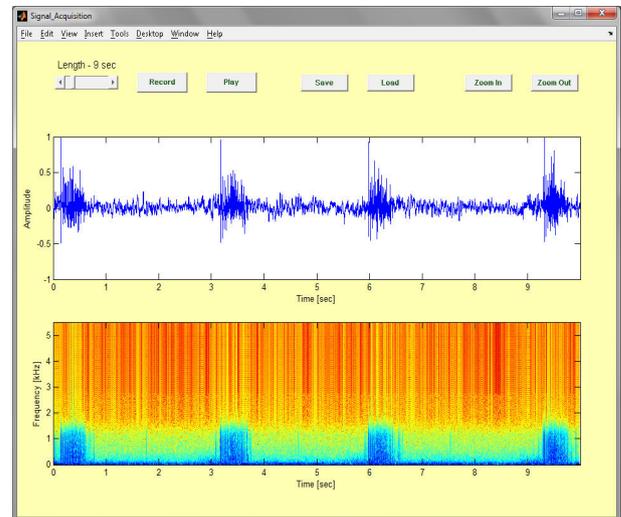


Fig 3: Signal acquisition using our proposed system

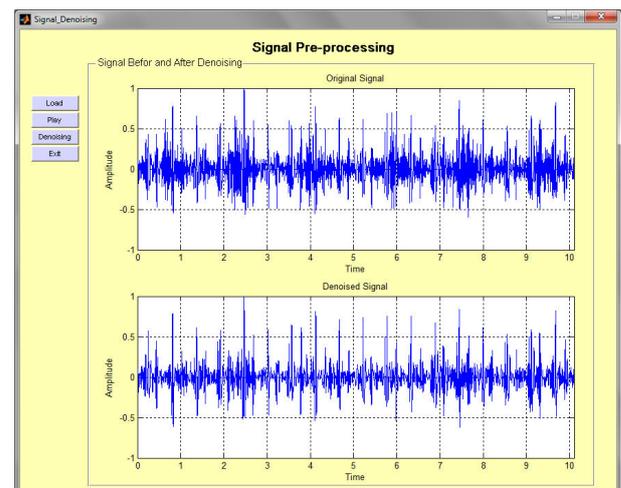


Fig 4: Signal pre-processing of a pulmonary crackle

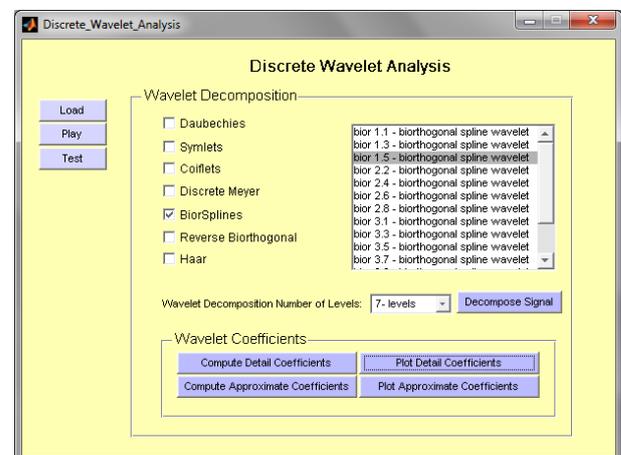


Fig 5: GUI of off-line discrete wavelet analysis

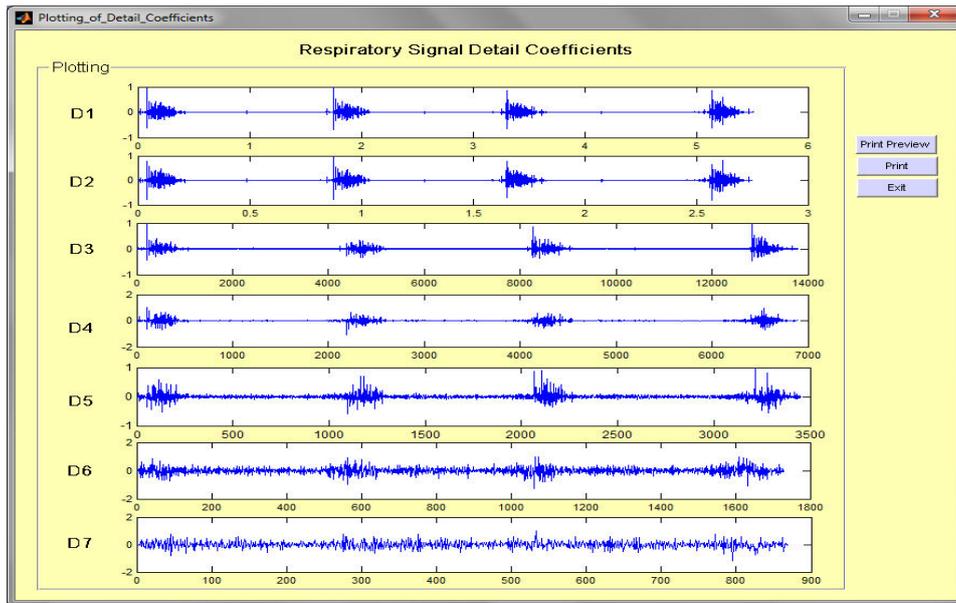


Fig 6: Detail coefficients of a DWT decomposed crackle sound

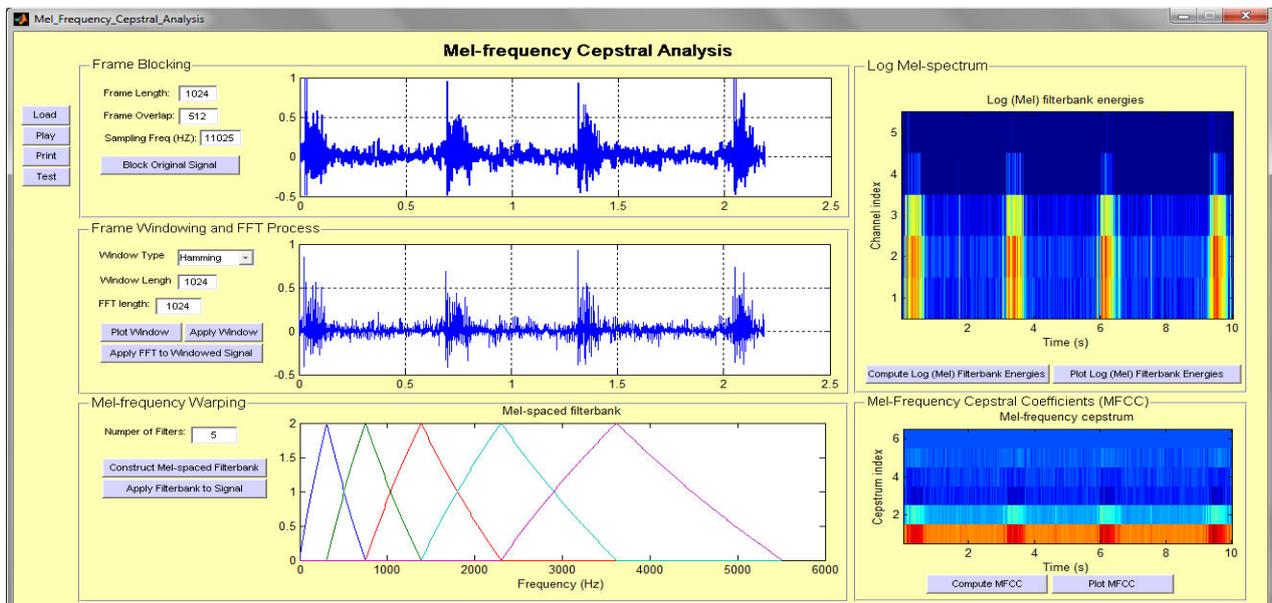


Fig 7: GUI of off-line Mel-frequency cepstral analysis

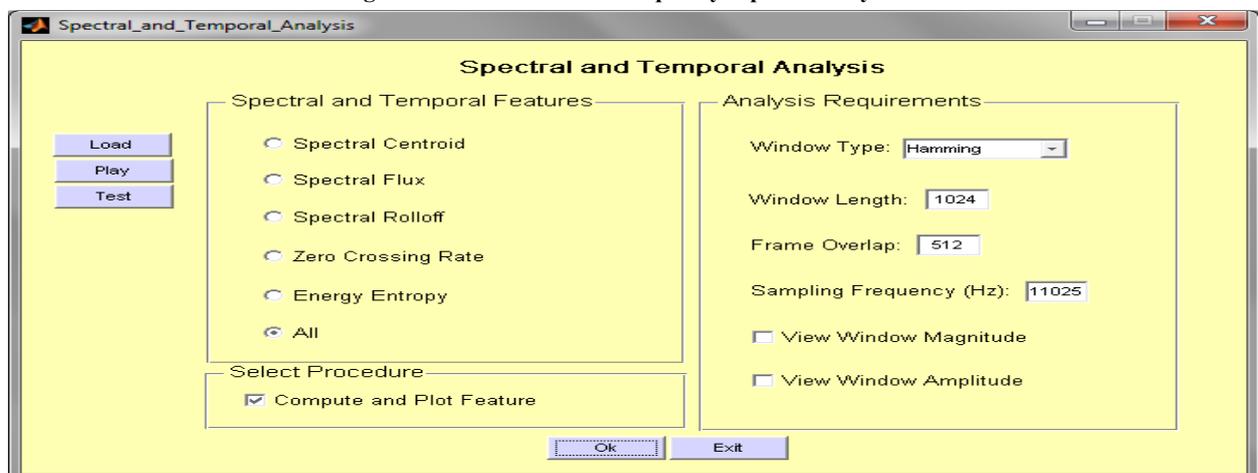


Fig 8: GUI of off-line spectral and temporal analysis

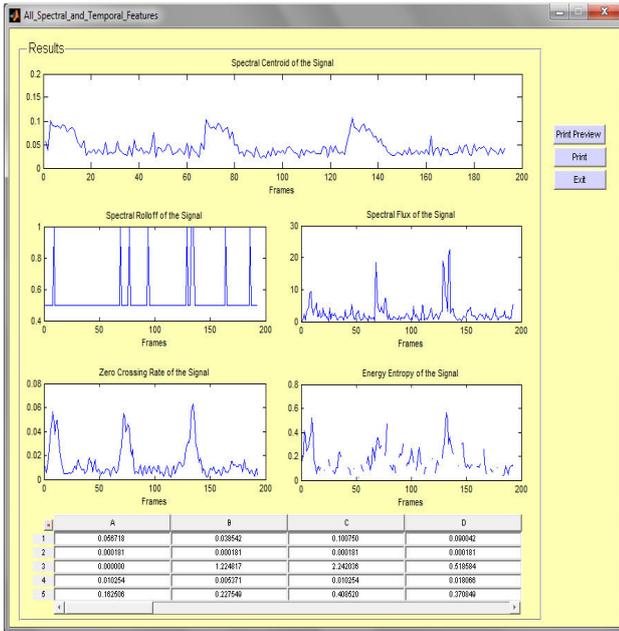


Fig 9: The graphs and results of the off-line spectral and temporal analysis of a pulmonary crackle

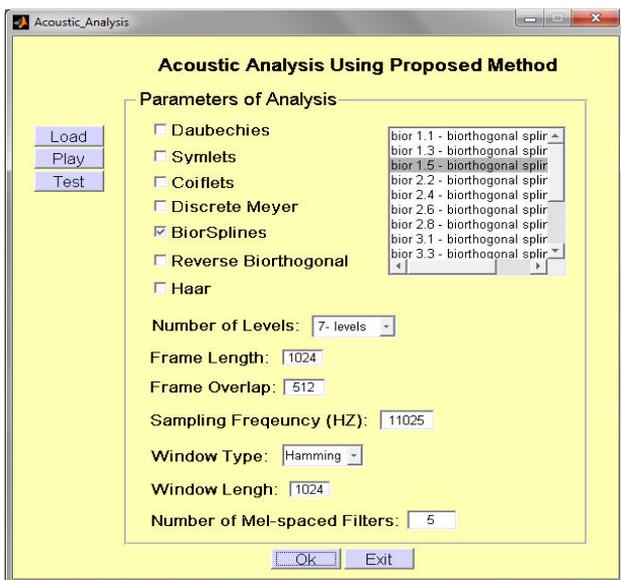


Fig 10: GUI of acoustic analysis using proposed method

Table 2. Network architecture and training parameters

MLP architecture	
Number of layers	3
Number of neuron on the layers	Input: vector of resulting centroids Hidden: 48 Output: 6
Activation functions	Hidden: hyperbolic tangent sigmoid Output: logistic sigmoid
MLP parameters	
Learning rule	Backpropagation
Learning rate	0.01
Epoch	5000
Average squared error	0.001

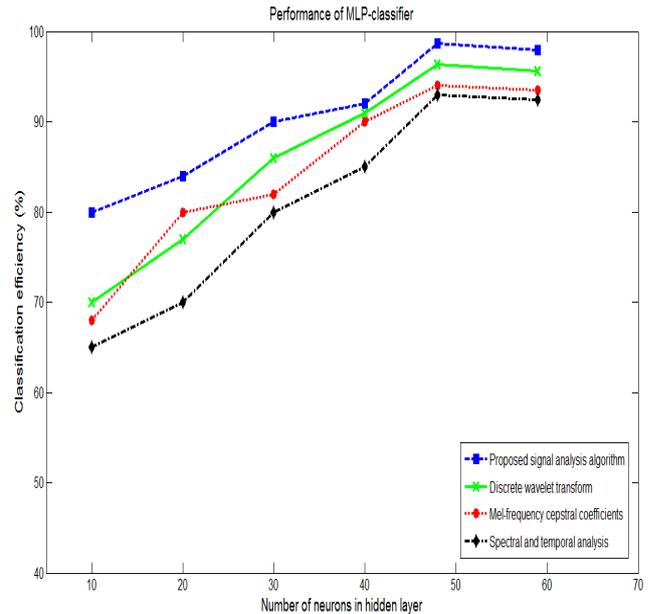


Fig 11: Classification efficiency comparisons based on the number of neurons in hidden layer

5. REFERENCES

- [1] Sovijarvi, A.R.A., Vanderschoot, J., and Earis, J.E. 2000. Standardization of computerised respiratory sound analysis. *European Respiratory Review*, 10 (77), 585.
- [2] Bahoura, M. 2009. Pattern recognition methods applied to respiratory sounds classification to normal and wheeze classes. *Computers in Biology and Medicine*, 39 (9), 824–843.
- [3] Kandaswamy, A., Kumar C. S., Ramanathan, Rm.Pl., Jayaraman, S., and Malmurugan, N. 2004. Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34 (6), 523–537.
- [4] Sovijarvi, A.R.A., Dalmaso, F., Vanderschoot, J., Malmberg, L.P., Righini, G., and Stoneman, S.A.T. 2000. Definition of terms for applications of respiratory sounds. *Eur Respir Rev*, 10 (77), 597–610.
- [5] Sovijarvi, A.R.A., Malmberg, L.P., Charbonneau, G., and Vanderschoot, J. 2000. Characteristics of breath sounds and adventitious respiratory sounds. *European Respiratory Review*, 10 (77), 591–596.
- [6] Mehta, A.C., Gat, M., Mann, S., and Madison, J.M. 2010. Accuracy of gray-scale coding in lung sound mapping. *Computerized Medical Imaging and Graphics*, 34 (5), 362–369.
- [7] Dalmy, F., Antonini, M.T., Marquet, P., and Menier, R. 1995. Acoustic properties of the normal chest. *European Respiratory Journal*, 8 (10), 1761–1769.
- [8] Sankur, B., Kahya, Y.P., Guler, E.C., and Engin, T. 1994. Comparison of AR-based algorithms for respiratory sound classification. *Computers in Biology and Medicine*, 24 (1), 67–76.
- [9] Guler, E., Sankur, B., Kahya, Y., and Raudys, S. 2005. Two-stage classification of respiratory sound patterns. *Computers in Biology and Medicine*, 35 (1), 67–83.

- [10] Bahoura, M., and Pelletier, C. 2004. Respiratory sounds classification using cepstral analysis and Gaussian mixture models. In 26th Annual Conference of the IEEE EMBS, San Francisco, CA, September 1–5, pp. 9–12.
- [11] Taplidou, S.A., Hadjileontiadis, L.J., Kittsas, I.K., and Panoulas, K.I. 2004. On applying continuous wavelet transform in wheeze analysis. In Proceedings of the 26th International Conference of the IEEE EMBS, San Francisco, CA, USA, pp. 3832–3835.
- [12] Polat, H., and Guler, I. 2004. A simple computer-based measurement and analysis system of pulmonary auscultation sounds. *J. Med. Syst.* 28 (6), 665–672.
- [13] Taplidou, S.A., and Hadjileontiadis, L.J. 2007. Nonlinear analysis of wheezes using wavelet bicoherence. *Computers in Biology and Medicine*, 37 (4), 563–570.
- [14] Waitman, L.R., Clarkson, K.P., Barwise, J.A., and King, P.H. 2000. Representation and classification of breath sounds recorded in an intensive care setting using neural networks. *Journal of Clinical Monitoring and Computing*, 16 (2), 95–105.
- [15] Lin, B.S., Wu, H.D., Chong, F.C., and Chen, S.J. 2006. Wheeze recognition based on 2D bilateral filtering of spectrogram. *Biomedical Engineering Applications, Basis & Communications*, 18 (3), 128–137.
- [16] Donoho, D.L., and Johnstone, I.M. 1995. Adapting to unknown smoothness via wavelet shrinkage, *J. Am. Stat. Assoc.*, 90 (432), 1200–1224.
- [17] Calvo de Lara, J.R. 2005. A method of automatic speaker recognition using cepstral features and vectorial quantization. Springer-Verlag Berlin Heidelberg, 3773, 146-153.
- [18] Jung, S., and Ghaboussi, J. 2006. Characterizing rate-dependent material behaviors in self-learning simulation. *Computer Methods in Applied Mechanics and Engineering*, 196 (1–3), 608–619.
- [19] Riedmiller, M., and Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA.
- [20] The R.A.L.E. lung sound 3.0. www.rale.ca, 2003.