Digital Signal Processing 23 (2013) 1012–1021

Contents lists available at SciVerse ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Pulmonary crackle detection using time–frequency and time–scale analysis

Gorkem Serbes [a], C. Okan Sakar [b], Yasemin P. Kahya [c], Nizamettin Aydin [d,*]

[a] Department of Biomedical Engineering, Bahcesehir University, Istanbul, Turkey
[b] Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey
[c] Department of Electrical and Electronics Engineering, Bogazici University, Istanbul, Turkey
[d] Department of Computer Engineering, Yildiz Technical University, Istanbul, Turkey

ARTICLE INFO

Article history:
Available online 20 December 2012

Keywords:
Lung sounds
Crackle detection
Time–frequency and time–scale analysis
Dual-tree complex wavelet transform
Denoising
Ensemble methods
Support vector machines

ABSTRACT

Pulmonary crackles are used as indicators for the diagnosis of different pulmonary disorders in auscultation. Crackles are very common adventitious transient sounds. From the characteristics of crackles such as timing and number of occurrences, the type and the severity of the pulmonary diseases may be assessed. In this study, a method is proposed for crackle detection. In this method, various feature sets are extracted using time–frequency and time–scale analysis from pulmonary signals. In order to understand the effect of using different window and wavelet types in time–frequency and time–scale analysis in detecting crackles, different windows and wavelets are tested such as Gaussian, Blackman, Hanning, Hamming, Bartlett, Triangular and Rectangular windows for time–frequency analysis and Morlet, Mexican Hat and Paul wavelets for time–scale analysis. The extracted feature sets, both individually and as an ensemble of networks, are fed into three different machine learning algorithms: Support Vector Machines, *k*-Nearest Neighbor and Multilayer Perceptron. Moreover, in order to improve the success of the model, prior to the time–frequency/scale analysis, frequency bands containing no-crackle information are removed using dual-tree complex wavelet transform, which is a shift invariant transform with limited redundancy compared to the conventional discrete wavelet transform. The comparative results of individual feature sets and ensemble of sets, which are extracted using different window and wavelet types, for both pre-processed and non-pre-processed data with different machine learning algorithms, are extensively evaluated and compared.

1. Introduction

Chest auscultation of pulmonary sounds by using a stethoscope is a commonly used, economic and noninvasive method for the evaluation of the respiratory disorders. However, due its inherent subjectivity and limited frequency response (the stethoscope attenuates frequencies above 120 Hz), stethoscope is considered to be an inadequate diagnostic method for respiratory disorders. Over the last three decades, the analysis of pulmonary sound signals with computers has become an established research area with the improvements in digital acquisition systems and advanced digital signal processing techniques [1–4].

Although the exact mechanism is still a mystery, the pulmonary sounds are assumed to be produced in the lungs due to the air turbulence produced in their airways. Pulmonary sounds can be divided into two classes, vesicular sounds and adventitious sounds. Vesicular sounds are the normal respiratory sounds which can be heard over the chest wall. Vesicular sounds are synchronous with the air flow occurring in the airways. Adventitious sounds, on the other hand, are additional sounds which usually occur because of respiratory disorders [5].

Crackles are discontinuous, adventitious non-musical respiratory sounds which are attributed to sudden bursts of air within bronchioles. Their duration is less than 20 ms and their frequency range is between 150 to 2000 Hz. Crackles frequently occur in pathological conditions and are superimposed on vesicular sounds. Crackles' morphologic character is explosive and transient, and they occur frequently in respiratory diseases. The inherent properties of pulmonary crackles such as timing, epochs of occurrence, and pitch can be used in the diagnosis for various types of pulmonary diseases such as pneumonia, bronchiectasis, fibrosing alveolitis and asbestosis [6–9].

For an automatic computerized analysis of pulmonary diseases, proper detection of crackles is very important. In this study, a novel method is proposed for pulmonary crackle detection. For the analysis, a pulmonary dataset consisting of 3000 512-point crackle containing signals and 3000 512-point non-crackle

* Corresponding author. Fax: +90 (212) 383 57 32.
E-mail addresses: gorkem.serbes@bahcesehir.edu.tr (G. Serbes), okan.sakar@bahcesehir.edu.tr (C.O. Sakar), kahya@boun.edu.tr (Y.P. Kahya), naydin@yildiz.edu.tr (N. Aydin).

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2012.12.009

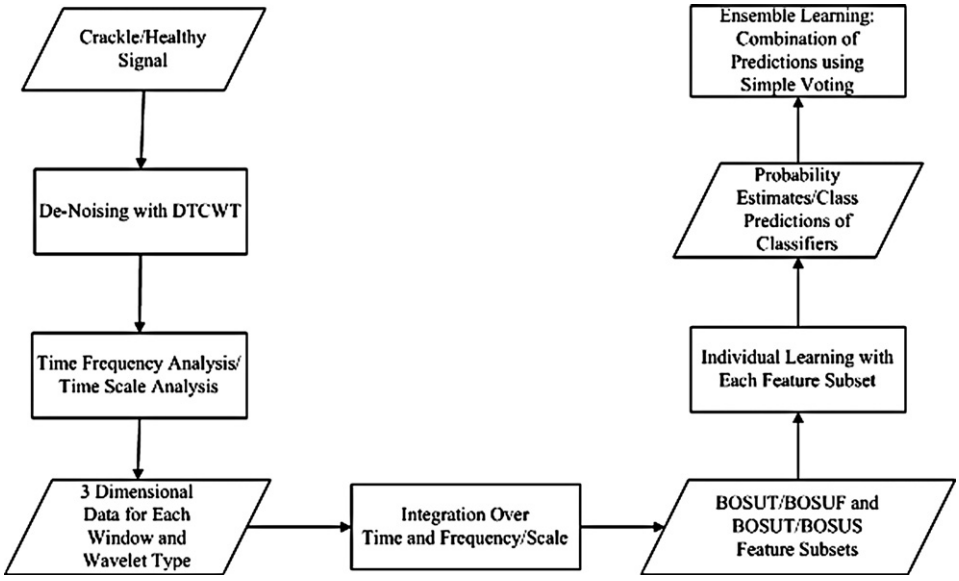

Fig. 1. Overall block diagram of proposed method for one machine learning algorithm. (BOSUT is the abbreviation of "behavior of signals upon time", BOSUF and BOSUS are the abbreviations for "behavior of signals upon frequency" and "behavior of signals upon scale" respectively. A more detailed explanation of the abbreviations can be found in Section 2.2.1.)

containing signals, are used. In order to facilitate detecting crackle signals, various feature sets, which contain frequency and scale information, are extracted by using time–frequency (TF) and time–scale (TS) analysis. With the aim of obtaining best crackle detection performance, different window and wavelet types such as Gaussian, Blackman, Hanning, Hamming, Bartlett, Triangular and Rectangular for TF analysis and Morlet, Mexican Hat and Paul for TS analysis are tested for these 6000 signals. Eventually, as an end-product of these trials 20 different feature subsets are obtained.

In order to improve the generalization and crackle detection capability of the model, frequency components of processed signals containing no-information (below 150 Hz and above 2400 Hz) are removed using dual-tree complex wavelet transform (DTCWT), which is an improved version of discrete wavelet transform (DWT) with better shift invariance property, as a pre-processing stage. Then, the extracted feature subsets are fed into Support Vector Machines (SVM), *k*-Nearest Neighbor (*k*-NN) and Multilayer Perceptron (MLP) classifiers as inputs both individually and as an ensemble of networks which are very commonly used machine learning tools. The comparative results of individual and ensemble feature sets with pre-processed and non-pre-processed data using SVM, *k*-NNs and MLPs are presented and analyzed. An overall block diagram of the proposed method can be seen in Fig. 1.

The remaining of the paper is organized as follows: Section 2 describes the materials and methods. Section 3 presents the experimental results on the pulmonary dataset. Section 4 provides the discussions and conclusions.

2. Materials and methods

2.1. Data acquisition system

In the data acquisition system fourteen air-coupled electret microphones (Sony-ECM 44) are placed on the posterior chest, and airflow is recorded using Fleisch-type flowmeter (Validyne CD379) to synchronize on the inspiration–expiration phases. A low-noise preamplifier, 8th order Butterworth low-pass filters with 4 kHz cut-off frequency and 6th order Bessel high-pass filters with 80 Hz cut-off frequency are used in order to minimize frictional noise and heart sound interference and for an anti-aliasing filter. The amplified signals are digitized by a 12-bit ADC Card (NIDAQ500) at a 9.6 kHz sampling rate and stored [10]. The details of the system are described in [11]. In Fig. 2, an example consisting of two 512-point crackle signals and two healthy signals is illustrated.

2.2. Feature extraction

The spectral characteristics of lung sounds show different behaviors according to the state and pathology of the lung. The pathological sounds appear in higher frequency bands, i.e. as crackles which are explosive and transient in time. For feature extraction, we used the frequency characteristics of crackles using TF and TS analysis for both the non-pre-processed and pre-processed signals. For pre-processing, we applied DTCWT with the aim of removing the frequency bands which do not contain crackle information.

2.2.1. Feature extraction using TF and TS analysis

The signals collected from 26 subjects (13 healthy, 13 pathological) are divided into 6000 samples, 3000 non-crackle, i.e. healthy, and 3000 crackle, each consisting of 512 points. In the preparation of crackle samples, 3000 crackles which were identified previously by physicians are randomly placed into 512-point signals. The signal size is chosen as 512 points because the duration of crackles is less than 20 ms and due to the sampling frequency of the data acquisition system, which is 9600 Hz, 512 points of the signal is equal to 53.34 ms which guarantees that one or more crackles are located in a signal. In the preparation of healthy signals, 3000 512-point signals were randomly created from healthy subjects data.

To underscore frequency characteristics of crackles, TF and TS analysis are applied to both crackle signals and healthy signals with different window types in TF analysis and with different wavelet types in TS analysis. In order to obtain optimum frequency/scale resolution in the TF and TS analyses, 64-point Fourier transform with Gaussian, Blackman, Hanning, Hamming, Bartlett, Triangular and Rectangular windows and 64 scales wavelet transform with Morlet, Paul and Mexican Hat wavelets are used, respectively.

The output of TF analysis gives information about behavior of analyzed signals with respect to both time and frequency for each type of window. In order to obtain the behavior of signals with respect to time and frequency separately, for each window type, the

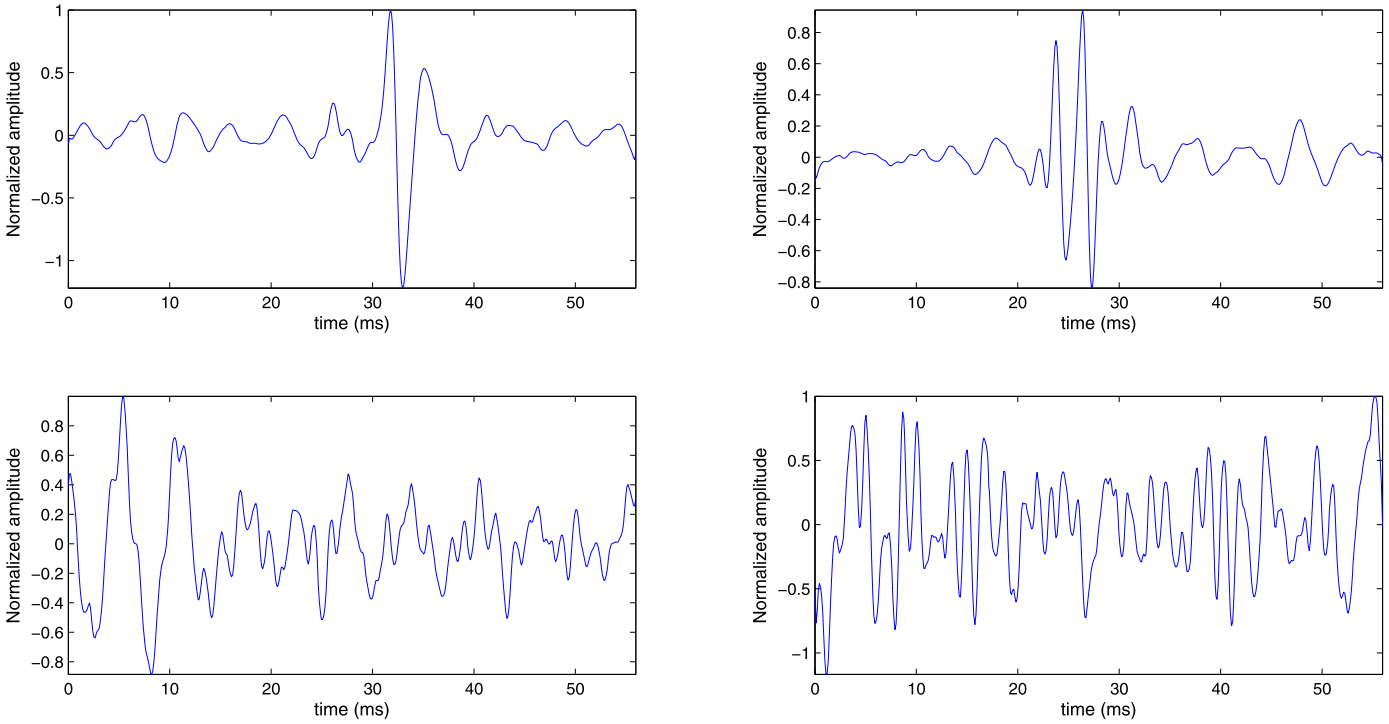

Fig. 2. Examples of 512-point signals containing crackles (upper two) and healthy signals (lower two).

outputs of TF analysis are integrated over frequency and the behavior of signals upon time (BOSUT) is obtained [12]. Similarly for each window type, the outputs of TF analysis are integrated over time, and the behavior of signals upon frequency (BOSUF) is obtained. This procedure was also carried out for TS analysis [13]. For forming feature sets with related wavelet type, the outputs of TS analysis are integrated over scale and time, and the behaviors of signal upon time and scale (BOSUS), respectively, are obtained. At the end of these integration operations, for each different TF analysis window and TS analysis wavelet type two new feature subsets are obtained (for TF analysis BOSUT and BOSUF, for TS analysis BOSUT and BOSUS) from an original crackle/healthy signal. As a result, from a single crackle or healthy signal analysis, using 7 TF windows and 3 TS wavelets, $2 * 10 = 20$ new feature subsets are obtained.

2.2.1.1. Time–frequency analysis The Fourier transform (FT) expands a time-domain signal into sines and cosines, which are completely unlocalized in time. That is, the spectrum gives us information on the frequencies contained in the signal as well as their amplitudes and phases, but does not give any information at which times these frequencies occur. Thus, FT processes signals assuming that they are stationary. In reality, however, most natural signals, including pulmonary signals, are non-stationary. In order to characterize a non-stationary signal properly, it is necessary to observe the changes in the signal both in time and in frequency. The windowed Fourier transform (WFT) has been used widely in this regard as it partially fulfills these requirements. The WFT introduces time dependency in the Fourier transform by pre-windowing the signal $s(t)$ around a particular time t, and calculating its fast Fourier transform (FFT), which is repeated for each time instant t. The WFT of $s(t)$ is given by

$$F_s(t, f) = \int_{-\infty}^{+\infty} s(\tau) g^*(\tau - t) e^{-j2\pi f \tau} \, d\tau \tag{1}$$

where $g(t)$ is a short time analysis window function. As multiplication by relatively short window suppresses the signal outside a neighborhood around analysis time point $\tau = t$, the WFT is a local spectrum of the signal $s(t)$ around a particular t.

An important FFT parameter which may influence the crackle detection performance is window type of WFT. Multiplying the signal with a suitable window function highlights the information near the middle of window and suppresses the information near the ends of window. Many window functions have been proposed in literature [14,15].

In this study for the TF analyses, 64-point WFT with Gaussian, Blackman, Hanning, Hamming, Bartlett, Triangular and Rectangular windows were used. In WFT, shifting the analysis window by less than the window length results in an overlapped FFT of analyzed signals [16]. Conventionally, the signals are processed sequentially by sliding the window less than the window size at each processing stage, most of the times sliding by 1 point. Consequently, overlapping FFT windows produces higher-dimensional WFTs. Some of the information in an overlapped WFT is redundant, and some of it is novel. In this study 15 points of window-sliding is used in the TF analyses for crackle detection which gives us a reduced computational complexity rather than conventional sliding factor 1 with approximately same detection performance. Then, the three-dimensional outputs of TF analyses are integrated over frequency and over time in order to obtain the behavior of signals upon time and frequency separately.

2.2.1.2. Time–scale analysis The wavelet transform (WT) decomposes a time dependent signal into time–scale (TS) space and allows exact localization of any abrupt change, or an exact time and duration to be attributed to a short signal, which may not be evidenced by conventional signal processing techniques. A complete WT analysis creates a two-dimensional decomposition of a one-dimensional signal, typically with the horizontal axis as time and vertical axis corresponding to the wavelet scale. The third dimension is the amplitude of the WT coefficients. It is performed by projecting a signal $s(t)$ onto a family of zero-mean functions

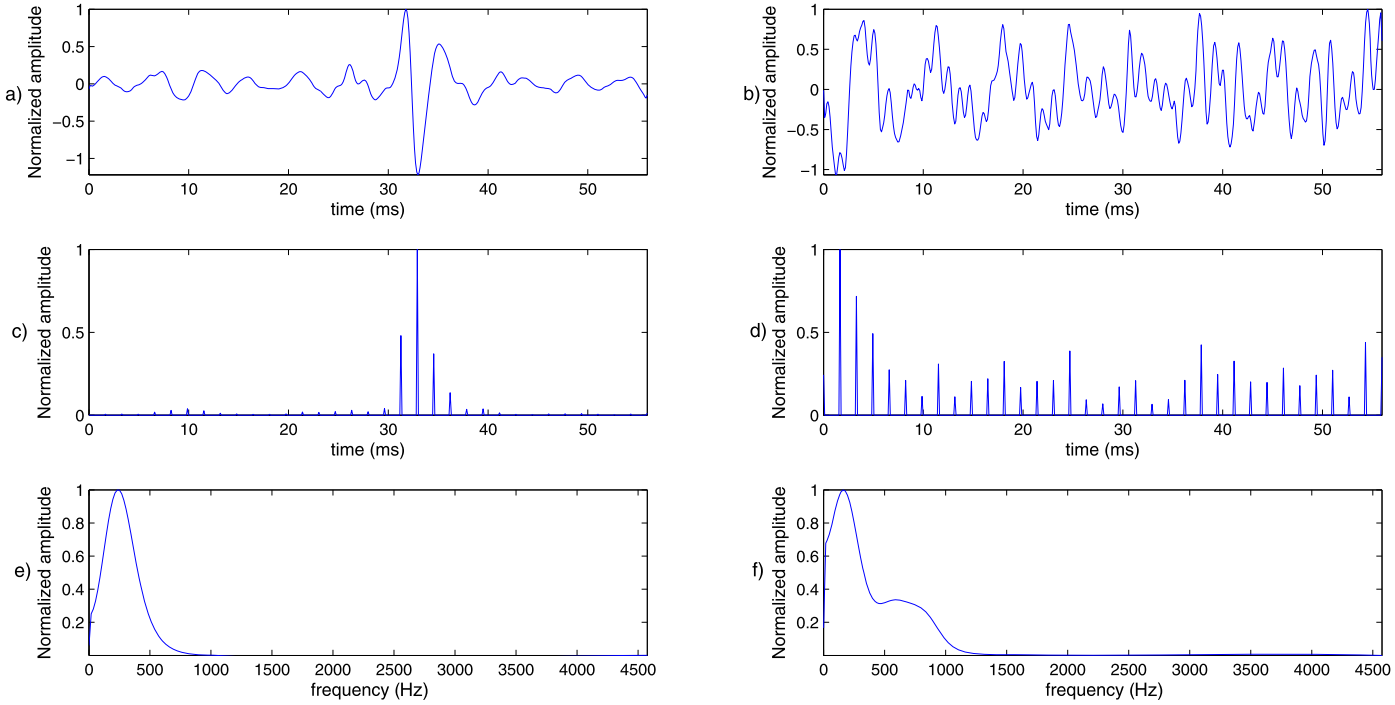

Fig. 3. For the WFT analysis using Gaussian window. (a) Original crackle signal, (b) original healthy signal, (c) BOSUT for crackle sample, (d) BOSUT for healthy sample, (e) BOSUF for crackle sample, (f) BOSUF for healthy sample.

deduced from an elementary function $\varphi(t)$ by translations and dilations, and given by

$$W_s(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} s(t)\varphi^*\left(\frac{t-b}{a}\right) dt \tag{2}$$

where $\varphi^*(t)$ is the analyzing wavelet. In the TS analysis, in which the wavelet can be defined as a complex function, the variable $a(\neq 0)$ controls the scale of the wavelet. The variable b is the time translation and controls the position of the wavelet. The basic difference between the WT and the WFT is that when the scale factor a can be changed, the duration and the bandwidth of the wavelet are both changed but its shape remains the same. The WT uses short windows at high frequencies and long windows at low frequencies in contrast to the FFT, which uses a single analysis window. This partially overcomes the TF resolution limitation of the WFT. In this study 64 scales wavelet transform with Morlet, Paul and Mexican Hat wavelets are used in order to obtain new feature sets from original crackle and healthy signals. Three-dimensional output of TS analysis for each wavelet type was integrated over scale and time, and the behaviors of signal upon time and scale, separately, were obtained. In Fig. 3, a crackle signal (a) and healthy signal (b) WFT analysis results with Gaussian window are depicted. The behavior of crackle upon time (c) and upon frequency (e) is distinctively different than the behavior of healthy signal upon time (d) and upon frequency (f).

2.2.2. Denoising using dual tree

Vesicular sounds mainly have frequency components between 0–200 Hz, rarely extending up to 600 Hz whereas the frequency components of crackles extend between 150–2000 Hz. In order to improve the performance of the proposed method, a preprocessing step which removes the frequency components with no-crackle information, is applied to the dataset before extracting the feature sets. For this task, a five-level DTCWT is applied on both crackle and healthy signals. The DTCWT is developed to overcome the lack of shift invariance property of ordinary discrete wavelet transform (DWT). Moreover it has limited redundancy (2^m:1 for m-dimensional signals, which is a very good ratio as compared with undecimated DWT). In the analysis of non-stationary crackles, which are transient signals, DTCWT removes undesirable signal components more successfully than DWT because of its shift invariance property [17,18]. With DTCWT, for each crackle and healthy sample in the dataset, the frequency bands below 150 Hz and above 2400 Hz are replaced with null vectors and then the processed signals are reconstructed. The details of DTCWT are given in [19,20], whereas block diagram of a two-level DTCWT which demonstrates both analysis and synthesis parts is depicted in Fig. 4.

In order to visualize the effect of feature extraction and denoising, we reduce the dimensions of the original signal feature set and extracted feature sets, which are obtained at the end of WFT analysis with Gaussian window, using Principal Component Analysis (PCA). The projections of original signal feature sets into two-dimensional space before and after denoising are shown in Fig. 5(a) and (b). Additionally, the projections of extracted feature sets, the output of time–frequency analysis upon time (TFAUT) and time–frequency analysis upon frequency (TFAUF), into two-dimensional space before denoising and after denoising are shown in Fig. 5(c)–(e) and (f). It is seen that extracting new features by WFT and denoising by DTCWT improves the discriminative capability of the crackle detection algorithm.

2.3. Individual learning with feature sets

In crackle detection method, the original crackle–healthy signals and extracted feature sets obtained by TF and TS analysis are fed individually into three classifiers: Support Vector Machines (SVM), Multilayer Perceptron (MLP), and k-Nearest Neighbor (k-NN). The aim is to measure and compare the predictive power of the feature sets with various classifiers. The effect of denoising with DTCWT on the classifiers is also tested. The aim of each of the classifiers is to build a predictive model capable of distinguishing between the crackle and non-crackle signals. For this purpose, we divided

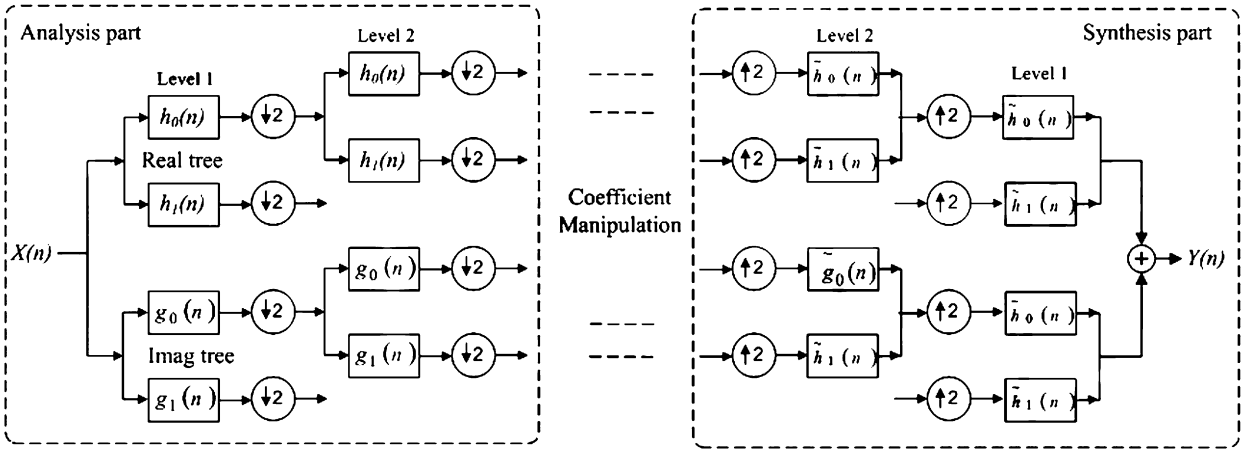

Fig. 4. Block diagram for a 2-level DTCWT which demonstrates both analysis and synthesis parts. $X(n)$ is input signal and $Y(n)$ is processed signal.

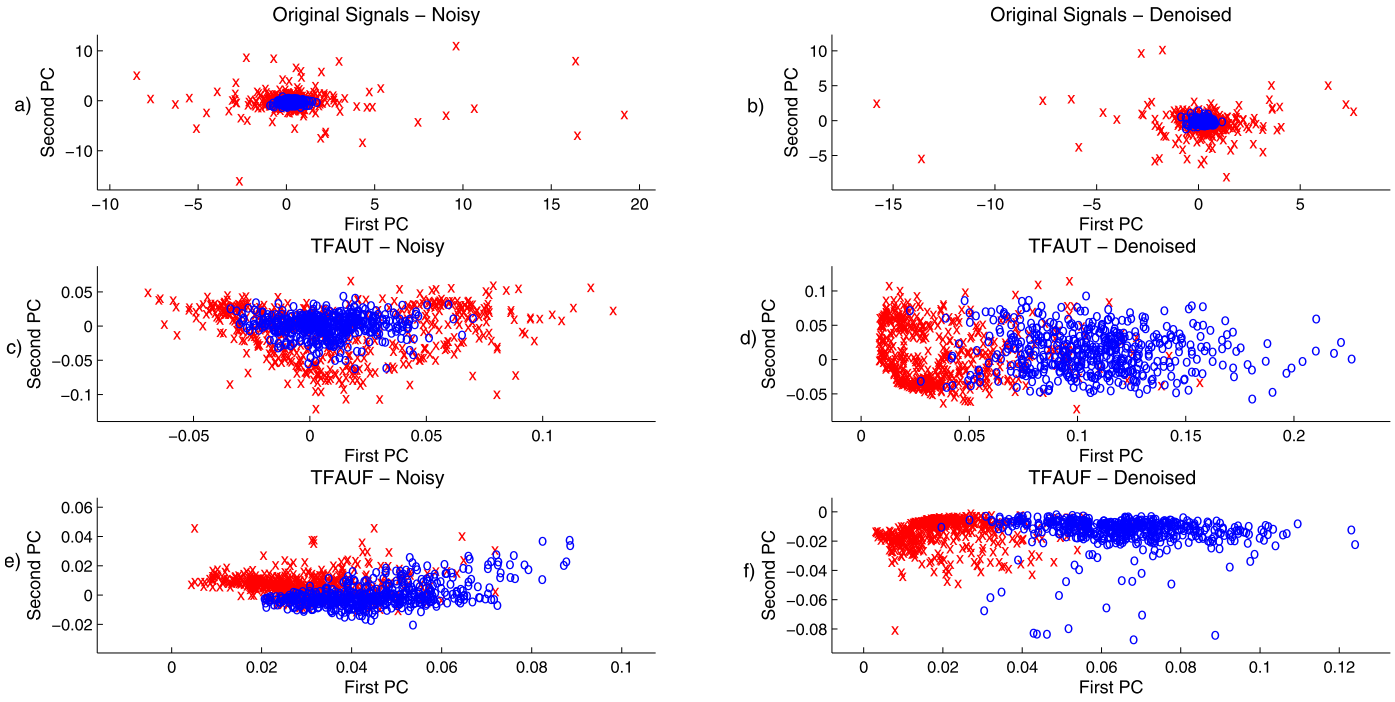

Fig. 5. Projections of the original feature sets before (a), and after denoising (b). Projections of the outputs of time–frequency analysis upon time before (c), and after denoising (d) with Gaussian window. Projections of the outputs of time–frequency analysis upon frequency before (e), and after denoising (f) with Gaussian window ('x' represents crackle, and 'o' represents healthy signals).

the dataset into three groups with equal number of samples: 2000 samples for the training, 2000 for the validation, and 2000 for the test. The distribution of the samples to the datasets has been done such that each set contains 1000 samples from each class type.

SVM is a very popular machine learning algorithm which aims to find the optimally placed hyperplanes to discriminate the classes from each other [21]. The closest samples to these hyperplanes are called support vectors, and the solution is defined in terms of this subset of samples which limits the complexity of the problem. Because the optimization problem has a unique solution, any iterative optimization procedure is not needed for convergence [22]. We used LIBSVM [23] implementation of SVM. We trained each feature subset using training set and tested on validation set in order to find the most suitable kernel type among linear, polynomial and Gaussian. The parameter values of the SVM, C (cost) and g (the spread parameter), are also optimized for each of the feature sets. The complexity of the solution is controlled by parameter C. Higher values of C may result in overfitting to the training set. After tuning the parameters on validation set, the optimized models are finally tested on the yet unseen test sets, and the unbiased success of each feature set is proposed.

We also used MLP to classify the pulmonary sounds which is the mostly used Artificial Neural Network (ANN) model for non-linear modeling. MLP is a feed-forward ANN model consisting of an input layer, an output layer, and at least one hidden layer. MLP is included in this study for comparison purposes because it is capable of modeling complex non-linear problems with many interactions among the input variables [24]. The elements of the hidden and output layers are called neurons. Each neuron is a processing element with a non-linear activation function. Our MLP model is composed of an input, an output, and a hidden layer. The same training, validation, and test sets of SVM model are used for MLP model. The parameters of the MLP model of each feature set such as number of neurons in the hidden layer (in the analysis hidden layer neuron number is chosen as *10*), number of iterations and learning rate are fine tuned on the validation set, and the most

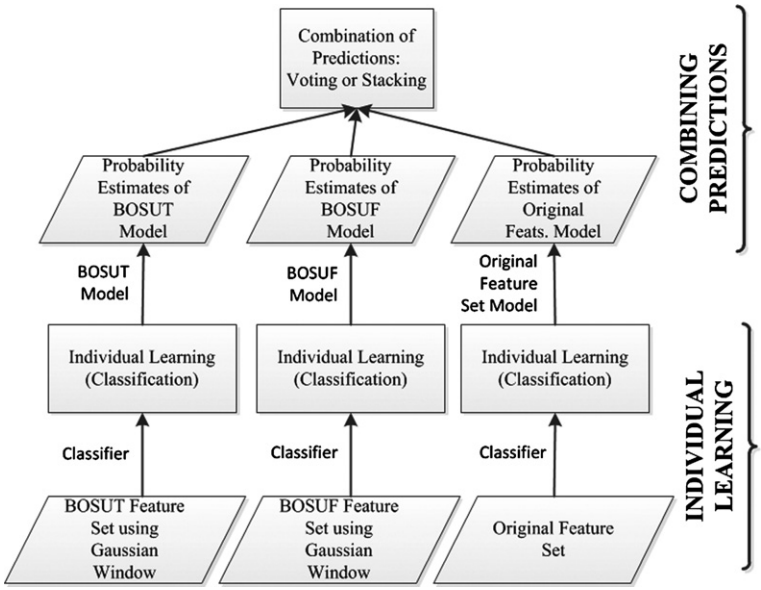

Fig. 6. SVM ensemble learning architecture for Gaussian window type.

suitable model is applied to the test set to propose the unbiased success of each feature set.

The third classifier we used for comparison purposes is k-Nearest Neighbor (k-NN) classifier, which is based on non-parametric density estimation. The k-NN method requires an integer k, a set of labeled examples and a measure of "closeness", and assigns the input to the class having most examples among the k neighbors of the input [22]. Small values of k result in more sensitive classifiers to the undenoised observations. The k parameter is optimized on the validation set (in the analysis k parameter is chosen as 3 or 5 depending on the feature subset) and the accuracy of the optimized model on the unseen test set for each feature set is proposed. Type of closeness measure has big impact on determining which set of learning examples is closest to the new example. In our study, we use City-block distance as the k-NN closeness metric which is a special case of the Minkowski metric,

$$d(x_s, x_t) = \sqrt[p]{\sum_{i=1}^{n} \left| x_s^i - x_t^i \right|^p} \tag{3}$$

where $p = 1$, $d(x_s, x_t)$ is the distance between samples s and t, and n is the dimension of the feature space.

2.4. Ensemble learning

We used the ensemble of feature subsets in order to improve the overall accuracy and generalization capability of the constructed model based on the proof of Hansen and Salamon [25]: if each member of the ensemble, i.e. feature subset, can get the right answer more than half the time, and if the responses of members are independent, the likelihood of an error by a majority voting strategy will monotonically decrease with the increasing number of members. Learning from multiple sets of features, called ensemble learning, is based on employing separate classifiers on each feature subset and combining the predictions of the views using techniques such as voting and stacking [22,26]. The final prediction, y, of an ensemble network is given by

$$y = \sum_{i=1}^{M} w_i d_i \tag{4}$$

satisfying

$$w_i \geqslant 0, \quad \forall i \quad \text{and} \quad \sum_{i=1}^{M} w_i = 1 \tag{5}$$

where w_i is the weight of the prediction of ith network member, d_i is the prediction of ith network member, and M is the total number of network members. The weight of the vote of each network member, i.e. w_i, is equal in the simple voting scheme ($w_i = 1/M$). The class with the maximum number of votes is the final prediction of the network. This voting strategy is called majority voting for two class classification problems. Besides, the information of how much confident the network member is for its prediction can be used to specify the final prediction. We use the class posterior probability estimates as the votes of the network members for SVM case.

We trained each of the 20 extracted feature sets and also the original signals individually on the training set. Then the individual predictions of the obtained models are used as a member of an ensemble network. Each ensemble network consists of three members: the two feature subsets which were obtained by the TF and TS analysis of each window/wavelet type resulting behavior of signals upon time/frequency and time/scale, respectively, and the original signals as themselves. For the SVM case, the class posterior probability estimates of the network members were combined using simple voting. For the MLP and k-NN case, the hard label predictions of each individual subset were combined using simple voting, and the ensemble learning results are obtained for each undenoised and denoised case. The SVM ensemble learning architecture with Gaussian window type is depicted in Fig. 6.

3. Experimental results

The overall accuracy error rates for SVM, MLP and k-NN learning algorithms of 20 extracted feature subsets and the original signal subset individually can be seen in Fig. 7. In this figure, the leftmost column set results belong to the original signal subset calculations. The following 14 column sets belong to the TF feature subset calculations. Final 6 column sets belong to the TS feature subset calculations. The feature set order is the same as the order in Table 1, column 1. From the figure it can be seen that MLP gives the worst performance for most of the feature subsets in both before denoising (16 of 21) and after denoising (18 of 21) cases. SVM

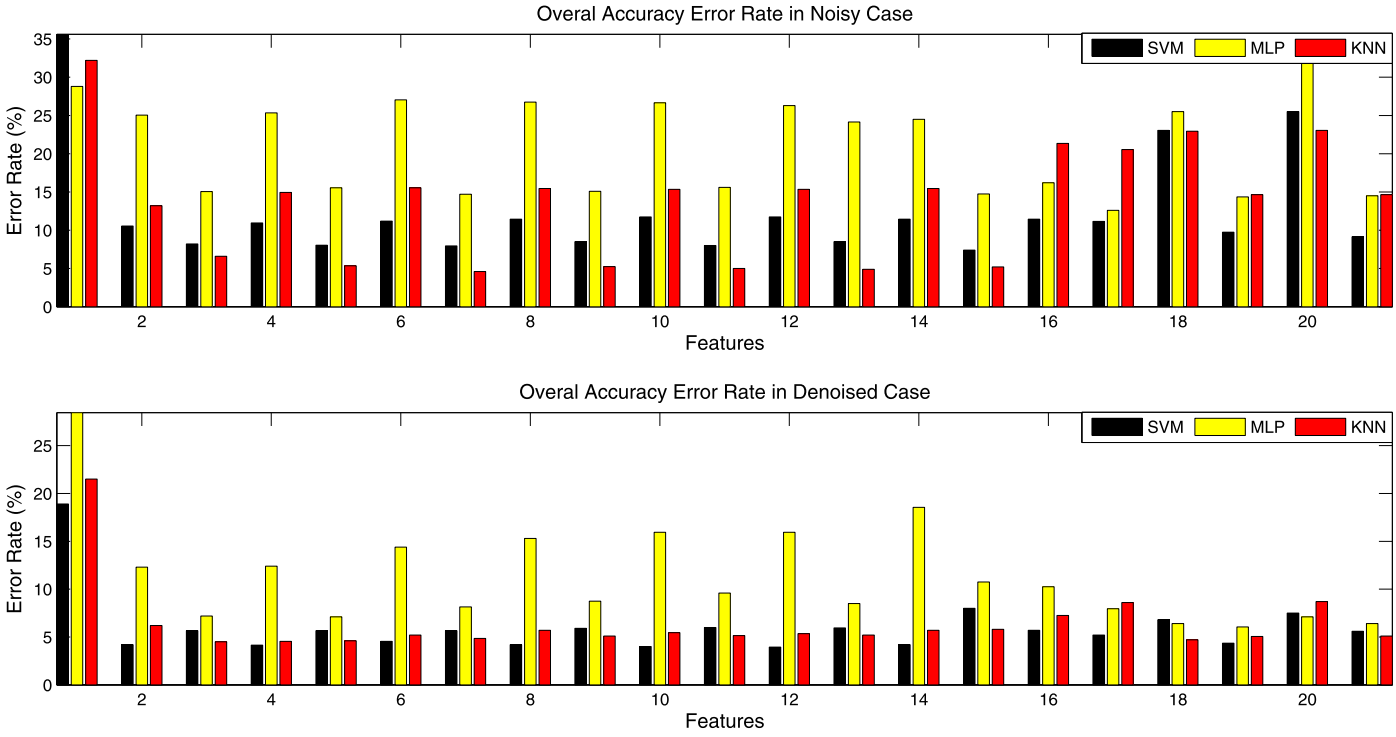

Fig. 7. Overall accuracy error rate of individual feature sets with undenoised and denoised case for three learning algorithms. (The leftmost column set results belong to the original signal subset calculations. The following 14 column sets belong to the TF feature subset calculations. Final 6 column sets belong to the TS feature subset calculations.)

Table 1
Overall accuracies, true positive (TP), and true negative (TN) rates of the individual feature sets and their ensembles for SVM case. (TFAUT and TFAUF indicate the feature subsets obtained by time–frequency analysis upon time and time–frequency analysis upon frequency respectively. Similarly, TSAUT and TSAUS indicate the feature subsets obtained by time–scale analysis upon time and by time–scale analysis upon scale respectively.)

	Denoised						Undenoised					
	Overall	TP	TN	Ensemble			Overall	TP	TN	Ensemble		
				Overall	TP	TN				Overall	TP	TN
Original signals	81.10	70.70	91.50				64.40	66.20	63.00			
TFAUT for Gaussian	95.80	96.30	95.30	**97.50**	**97.30**	97.70	89.45	80.80	98.10	91.55	84.00	99.10
TFAUF for Gaussian	94.35	95.10	93.60				91.80	88.60	95.00			
TFAUT for Blackman	95.85	96.00	95.70	97.40	97.10	97.70	89.05	80.10	98.00	91.50	83.90	99.10
TFAUF for Blackman	94.35	95.40	93.30				91.95	89.00	94.90			
TFAUT for Hanning	95.45	96.10	94.80	97.30	96.90	97.70	88.80	79.10	98.30	91.05	83.00	99.10
TFAUF for Hanning	94.35	94.90	93.80				92.05	89.60	94.50			
TFAUT for Hamming	95.80	96.10	95.50	97.20	96.80	97.60	88.55	79.40	97.70	90.70	82.20	99.20
TFAUF for Hamming	94.10	94.60	93.60				91.50	88.80	94.20			
TFAUT for Bartlett	96.00	96.30	95.70	97.15	96.80	97.50	88.25	78.70	97.80	91.05	82.90	99.20
TFAUF for Bartlett	94.00	94.30	93.70				92.00	89.60	94.40			
TFAUT for Triangular	96.05	96.30	95.80	97.15	96.80	97.50	88.25	78.70	97.80	90.70	82.30	99.10
TFAUF for Triangular	94.05	94.40	93.70				91.50	88.50	94.50			
TFAUT for Rectangular	95.80	96.10	95.50	97.05	96.10	98.00	88.55	79.70	97.70	90.55	82.40	98.70
TFAUF for Rectangular	92.00	92.80	91.20				92.60	89.40	95.80			
TSAUT for Morlet	94.30	93.20	95.40	96.70	95.50	97.90	88.55	81.00	96.10	90.90	86.10	95.70
TSAUS for Morlet	94.80	93.90	95.70				88.85	87.70	90.00			
TSAUT for Paul	93.20	92.40	94.00	96.25	94.80	97.70	76.95	70.90	83.00	87.35	81.20	93.50
TSAUS for Paul	95.65	94.90	96.40				90.25	90.00	90.50			
TSAUT for Mexican Hat	92.50	91.80	93.20	96.25	94.90	97.60	74.50	68.00	81.00	87.85	80.80	94.20
TSAUS for Mexican Hat	94.40	94.60	94.20				90.85	91.10	90.60			

gives the best performance for more than half of the feature subsets, both in undenoised (11 of 21) and after denoised (11 of 21) cases.

Fig. 8 depicts the true positive error rates for the SVM, MLP and k-NN learning algorithms of 20 extracted feature subsets and the original signal subset (feature set order is the same as Fig. 7) individually. SVM gives the best performance for more than half of the feature subsets, in both undenoised (13 of 21) and denoised (17 of 21) cases. MLP gives the worst performance for most of the feature subsets in both undenoised (15 of 21) and denoised (18 of 21) cases.

Fig. 9 shows how denoising with DTCWT affects TP error rate of all feature subsets for SVM, MLP and k-NN learning algorithms individually. It can be seen from the figure that for all three methods, the TP error rate is decreased by denoising with the DTCWT. This indicates that denoising improves the success of the proposed method in detecting the crackle signals.

All individual and ensemble learning results for SVM, MLP, and k-NN classifiers are listed in Tables 1, 2 and 3, respectively.

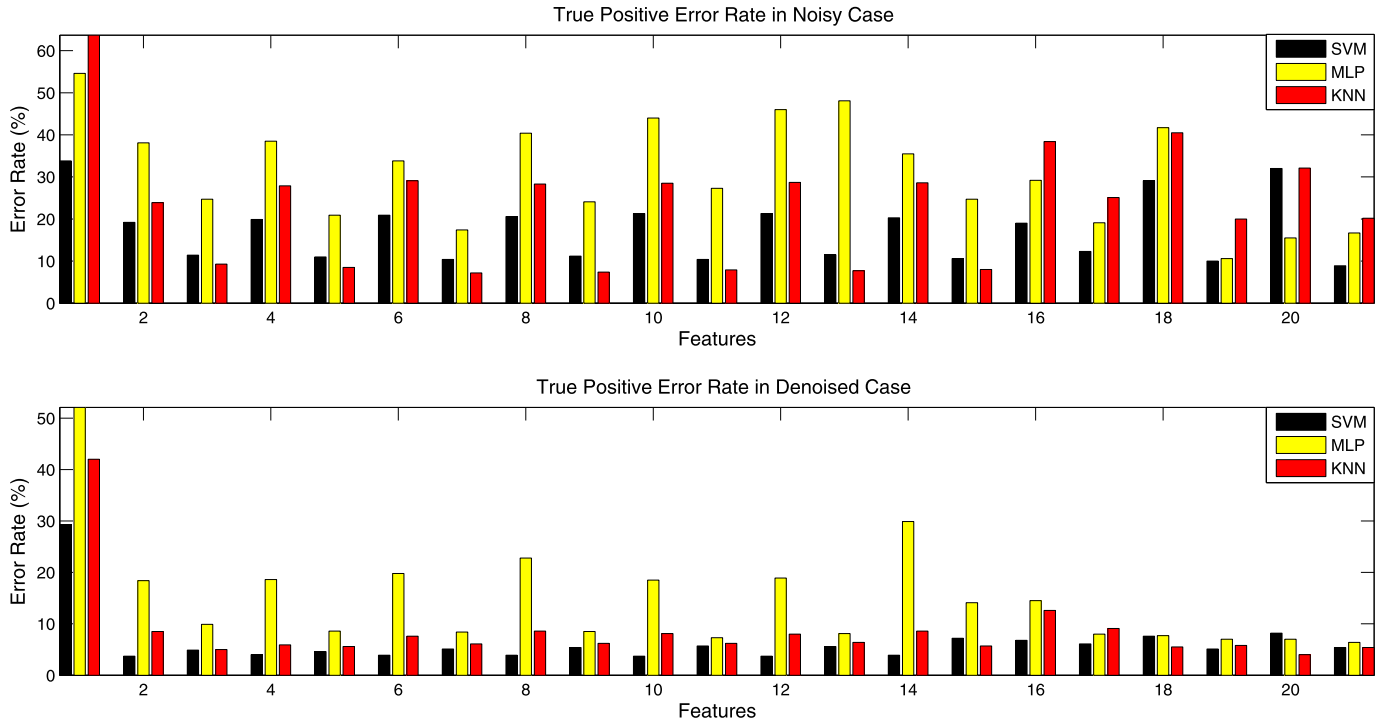

Fig. 8. True positive error rate of individual feature sets with undenoised and denoised case for three learning algorithms. (The leftmost column set results belong to the original signal subset calculations. The following 14 column sets belong to the TF feature subset calculations. Final 6 column sets belong to the TS feature subset calculations.)

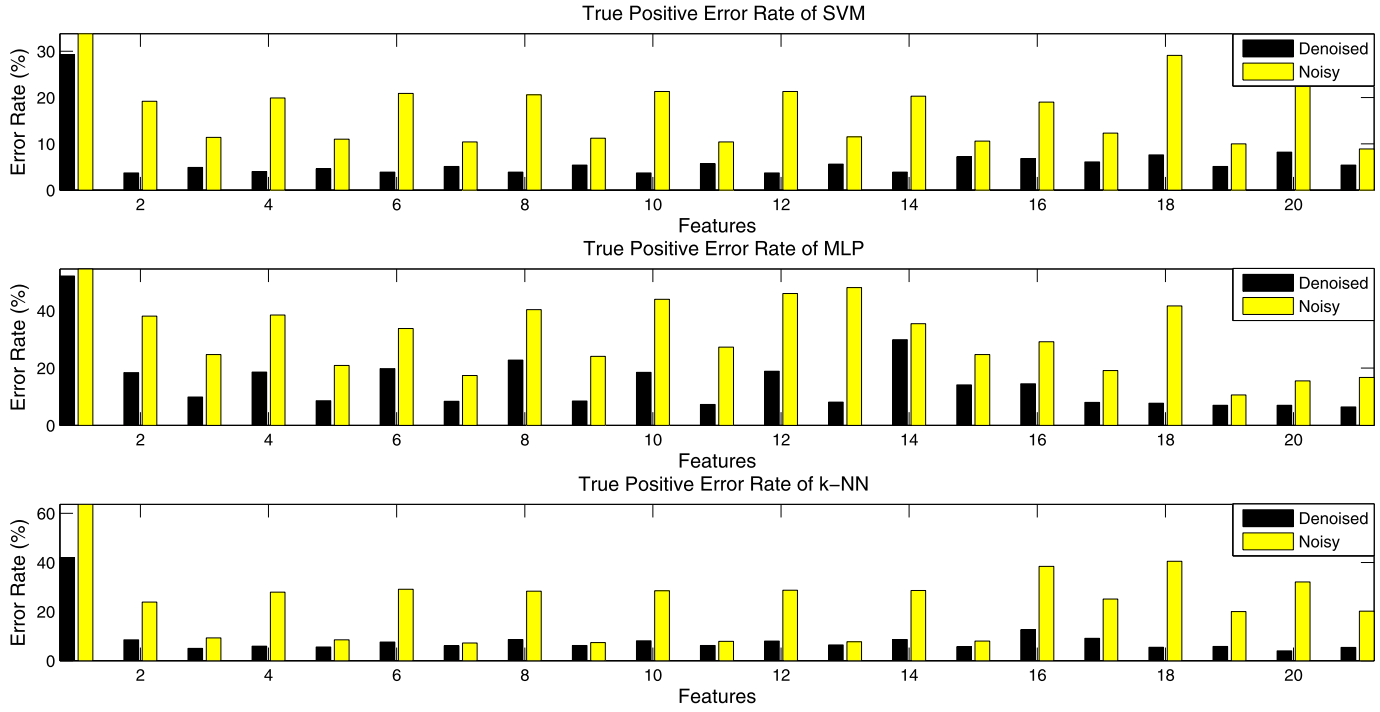

Fig. 9. True positive error rate of three learning methods sets for undenoised and denoised case. (The leftmost column set results belong to the original signal subset calculations. The following 14 column sets belong to the TF feature subset calculations. Final 6 column sets belong to the TS feature subset calculations.)

It is seen that the highest overall accuracy (97.50%) is obtained with our proposed method when a Gaussian window is used and DTCWT is applied as a pre-processing step, and the resulting feature sets are used as an ensemble of networks. Also the highest TP rate with 97.30% is obtained with the same procedure (with Gaussian window, DTCWT pre-processing and ensemble of networks).

4. Discussion and future works

The computerized analysis of pulmonary sound signals is a recent research area due to the improvements in digital recording systems and advanced digital signal processing techniques. In this study, a crackle detection method in which the DTCWT is used as a pre-processing step for removing the frequency bands containing

Table 2
Overall accuracies, true positive (TP), and true negative (TN) rates of the individual feature sets and their ensembles for MLP case. (TFAUT and TFAUF indicate the feature subsets obtained by time–frequency analysis upon time and time–frequency analysis upon frequency respectively. Similarly, TSAUT and TSAUS indicate the feature subsets obtained by time–scale analysis upon time and by time–scale analysis upon scale respectively.)

	Denoised						Undenoised					
	Overall	TP	TN	Ensemble			Overall	TP	TN	Ensemble		
				Overall	TP	TN				Overall	TP	TN
Original signals	71.55	47.90	95.20				71.20	45.40	97.00			
TFAUT for Gaussian	87.70	81.60	93.80	91.65	84.80	98.50	74.95	61.90	88.00	81.10	62.80	99.40
TFAUF for Gaussian	92.80	90.10	95.50				84.95	75.30	94.60			
TFAUT for Blackman	87.60	81.40	93.80	91.70	84.90	98.50	74.65	61.50	87.80	82.10	64.90	99.30
TFAUF for Blackman	92.90	91.40	94.40				84.45	79.10	89.80			
TFAUT for Hanning	85.60	80.20	91.00	91.50	84.90	98.10	72.95	66.20	79.70	83.95	68.80	99.10
TFAUF for Hanning	91.85	91.60	92.10				85.30	82.60	88.00			
TFAUT for Hamming	84.70	77.20	92.20	91.20	84.30	98.10	73.25	59.60	86.90	80.90	62.30	99.50
TFAUF for Hamming	91.25	91.50	91.00				84.90	75.90	93.90			
TFAUT for Bartlett	84.05	81.50	86.60	91.80	86.60	97.00	73.35	56.00	90.70	80.00	60.20	99.80
TFAUF for Bartlett	90.40	92.70	88.10				84.40	72.70	96.10			
TFAUT for Triangular	84.05	81.10	87.00	91.80	86.20	97.40	73.70	54.00	93.40	75.00	50.00	100
TFAUF for Triangular	91.50	91.90	91.10				73.85	51.90	99.80			
TFAUT for Rectangular	81.45	70.10	92.80	87.45	76.20	98.70	75.50	64.50	86.50	81.20	63.00	99.40
TFAUF for Rectangular	89.25	85.90	92.60				85.25	75.30	95.20			
TSAUT for Morlet	89.75	85.50	94.00	93.70	89.60	97.80	83.80	70.80	96.80	85.35	71.20	99.50
TSAUS for Morlet	92.05	92.00	92.10				87.40	80.90	93.90			
TSAUT for Paul	93.60	92.30	94.90	95.40	93.00	97.80	74.50	58.30	90.70	85.05	71.40	98.70
TSAUS for Paul	93.95	93.00	94.90				85.65	89.40	81.90			
TSAUT for Mexican Hat	92.90	93.00	92.80	95.20	93.20	97.20	67.85	84.50	51.20	86.70	77.70	95.70
TSAUS for Mexican Hat	93.60	93.60	93.60				85.50	83.30	87.70			

Table 3
Overall accuracies, true positive (TP), and true negative (TN) rates of the individual feature sets and their ensembles for *k*-NN case. (TFAUT and TFAUF indicate the feature subsets obtained by time–frequency analysis upon time and time–frequency analysis upon frequency respectively. Similarly, TSAUT and TSAUS indicate the feature subsets obtained by time–scale analysis upon time and by time–scale analysis upon scale respectively.)

	Denoised						Undenoised					
	Overall	TP	TN	Ensemble			Overall	TP	TN	Ensemble		
				Overall	TP	TN				Overall	TP	TN
Original signals	78.50	58.00	99.00				67.80	36.30	99.30			
TFAUT for Gaussian	93.80	91.50	96.10	95.70	93.10	98.30	86.80	76.10	97.50	87.85	87.85	99.60
TFAUF for Gaussian	95.50	95.00	96.00				93.40	90.70	96.10			
TFAUT for Blackman	95.45	94.10	96.80	96.10	94.00	98.20	85.05	72.10	98.00	86.15	86.15	99.90
TFAUF for Blackman	95.40	94.40	96.40				94.65	91.50	97.80			
TFAUT for Hanning	94.80	92.40	97.20	95.20	92.40	98.00	84.45	70.90	98.00	85.65	85.65	99.50
TFAUF for Hanning	95.15	93.90	96.40				95.40	92.80	98.00			
TFAUT for Hamming	94.30	91.40	97.20	95.05	91.60	98.50	84.55	71.70	97.70	85.80	85.80	99.60
TFAUF for Hamming	94.90	93.80	96.00				94.75	92.60	96.90			
TFAUT for Bartlett	94.55	91.90	97.20	95.00	91.70	98.30	84.65	71.50	97.80	85.65	85.65	99.60
TFAUF for Bartlett	94.85	93.80	95.90				95.00	92.10	97.90			
TFAUT for Triangular	94.65	92.00	97.30	95.00	91.60	98.40	84.65	71.30	98.00	85.55	85.55	99.60
TFAUF for Triangular	94.80	93.60	96.00				95.10	92.30	97.90			
TFAUT for Rectangular	94.30	91.40	97.20	94.95	91.60	98.30	84.55	71.40	97.70	85.80	85.80	99.60
TFAUF for Rectangular	94.20	94.30	94.10				94.80	92.00	97.60			
TSAUT for Morlet	92.75	87.40	98.10	93.50	88.50	98.50	78.65	61.60	95.70	79.40	79.40	98.20
TSAUS for Morlet	91.40	90.90	91.90				79.45	74.90	84.00			
TSAUT for Paul	95.30	94.50	96.10	96.15	94.50	97.80	77.05	59.50	94.60	79.55	79.55	98.40
TSAUS for Paul	94.95	94.20	95.70				85.35	80.00	90.70			
TSAUT for Mexican Hat	91.30	96.00	86.60	96.00	95.70	96.30	76.95	67.90	86.00	82.00	82.00	96.90
TSAUS for Mexican Hat	94.90	94.60	95.20				85.35	79.80	90.90			

no-crackle information, is proposed. In this method, various feature sets using TF and TS analysis are extracted and fed to the SVM, MLP, and *k*-NN classifiers.

It is observed that the overall accuracy performances of the SVM and the *k*-NN classifiers are very close to each other for both undenoised and denoised signals. However, the SVM appears to be superior over the *k*-NN in detecting the crackle signals on both undenoised and denoised data observations. It must be noted that SVM is successful at separating the class partitions when there are enough number of data observations as is the case with our dataset to determine the optimal value of cost parameter, C, which controls the complexity of the SVM model. One of the advantages of SVM is that the solution is defined in terms of support vectors which consist of a small subset of the data observations which limits the complexity of the problem. On the other hand, *k*-NN does not have an explicit model training, so the computational load of the testing step is very high in contrast to SVM since we need to compute the distance of the test instance to all training instances. Besides, *k*-NN is very sensitive to outliers and undenoised features because the algorithm does not reduce the effect of irrelevant features unless a feature selection as a pre-processing step is applied.

According to our results, it is concluded that ensemble of networks increases the overall and TP accuracy performances of SVM classifier for all 10 ensemble results. In contrast, for *k*-NN and MLP classifiers ensemble of networks is not as efficient as SVM case which originates from the fact that the individual learning accuracy of original signals with SVM (overall accuracy rate = 81.10%, TP rate = 70.70% in denoised case) is higher than MLP (overall

accuracy rate = 71.55%, TP rate = 47.90% in denoised case) and k-NN (overall accuracy rate = 78.50%, TP rate = 58.00% in denoised case) cases.

The MLP classifier provided the worst performance for most of the feature sets. There are two main reasons for this. First, neural networks are prone to overfitting unless some build-in mechanisms like weight decay are applied. We observed that in contrast to test set accuracies, best performance on training set is obtained with MLP. Second, neural networks require a very careful pre-processing step. The signal-to-noise ratio must be increased by eliminating the undenoised or irrelevant features in the dataset by applying feature selection methods.

As a conclusion we can say that using DTCWT as a pre-processing step, extracting features instead of using the original signals and combining the feature sets as an ensemble of networks improve the crackle detection capability of the proposed model. It should also be noted that combining the SVM models with different feature sets as an ensemble improves the overall accuracy and TP rate of the proposed method more than k-NN and MLP classifiers. In the future, it will be a challenge to implement the proposed method in real time as an online crackle detection system.

Acknowledgments

This work was supported by Bogazici University Research Fund under Project No. 06A202. The work of G. Serbes and C.O. Sakar is supported by the Ph.D. scholarship (2211) from Turkish Scientific Technical Research Council (TÜBITAK). G. Serbes is a Ph.D. student in the Biomedical Engineering Department at Bogazici University, Istanbul, Turkey. C.O. Sakar is a Ph.D. student in the Computer Engineering Department at Bogazici University, Istanbul, Turkey.

References

[1] N. Gavriely, Breath Sounds Methodology, CRC Press, Boca Raton, FL, 1995.
[2] H. Pasterkamp, S. Kraman, G. Wodicka, Respiratory sounds. Advances beyond the stethoscope, Am. J. Respir. Crit. Care Med. 156 (1997) 974–987.
[3] S. Lehrer, Understanding Lung Sounds, 3rd ed., WB Saunders Company, 2002.
[4] A. Sovijärvi, L. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, J. Earis, Characteristics of breath sounds and adventitious respiratory sounds, Eur. Respir. Rev. 77 (2000) 591–596.
[5] M. Yeğiner, Y.P. Kahya, Elimination of vesicular sounds from pulmonary crackle waveforms, Comput. Methods Programs Biomed. 89 (2008) 1–13.
[6] A. Nath, L. Capel, Inspiratory crackles: early and late, Thorax 29 (1974) 223–227.
[7] P. Piirilä, A.R. Sovijärvi, Crackles: recording, analysis and clinical significance, Eur. Respir. J. 8 (12) (1995) 2139–2148.
[8] A.R. Sovijärvi, P. Piirilä, R. Luukkonen, Separation of pulmonary disorders with two-dimensional discriminant analysis of crackles, Clin. Physiol. 16 (2) (1996) 171–181.
[9] A. Vyshedskiy, F. Bezares, R. Paciej, M. Ebril, J. Shane, R. Murphy, Transmission of crackles in patients with interstitial pulmonary fibrosis, congestive heart failure, and pneumonia, Chest 128 (3) (2005) 1468–1474.
[10] M. Yeğiner, Y.P. Kahya, Feature extraction for pulmonary crackle representation via wavelet networks, Comput. Biol. Med. 39 (2009) 713–721.
[11] I. Sen, Y.P. Kahya, A multi-channel device for respiratory sound data acquisition and transient detection, in: Proceedings of the International Conference of IEEE/EMBS, 2005, pp. 6658–6661.
[12] L. Cohen, Time–frequency distributions—A review, Proc. IEEE 7 (1989) 941–981.
[13] N. Aydin, H.S. Markus, Time–scale analysis of quadrature Doppler ultrasound signals, IEE Proc. Sci. Meas. Technol. 148 (1) (2001) 15–22.
[14] F.J. Harris, On the use of windows for harmonic analysis with discrete Fourier transform, Proc. IEEE 66 (1978) 51–83.
[15] S.M. Kay, Modern Spectral Estimation, Prentice Hall, Englewood Cliffs, NJ, 1988.
[16] N. Aydin, H.S. Markus, Optimisation of processing parameters for the analysis and detection of embolic signals, Eur. J. Ultrasound 12 (1) (2000) 69–79.
[17] G. Serbes, N. Aydin, Denoising embolic Doppler ultrasound signals using dual tree complex discrete wavelet transform, in: Proceedings of the International Conference of IEEE/EMBS, 2010, pp. 1840–1843.
[18] N.G. Kingsbury, Shift invariant properties of the dual-tree complex wavelet transform, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1999, pp. 1221–1224.
[19] N.G. Kingsbury, The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters, IEEE Digital Signal Processing Workshop, SP 98, Bryce Canyon, paper No. 86, 1998.
[20] I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Process. Mag. 22 (6) (2005) 123–151.
[21] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[22] E. Alpaydin, Introduction to Machine Learning, The MIT Press, 2010.
[23] C.W. Hsu, C.J. Lin, A comparison of methods for multi-class support vector machines, IEEE Trans. Neural Netw. 13 (2002) 415–425.
[24] B. Warner, M. Misra, Understanding neural networks as statistical tools, Amer. Statist. 50 (1996) 284–294.
[25] L. Hansen, P. Salamon, Neural network ensembles, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 993–1001.
[26] D. Wolpert, Stacked generalization, Neural Netw. 5 (2) (1992) 241–259.

Gorkem Serbes received B.Sc. degree in 2006 and M.Sc. degree in 2009 both in Electrical and Electronics Engineering from Bahcesehir University. He has currently completed his second year in Ph.D. in Institute of Biomedical Engineering of Bogazici University. He is a research assistant in Mechatronics Engineering and Biomedical Engineering at Bahcesehir University since 2006. His research interests are time–frequency and time–scale analysis, biomedical signal processing, pattern recognition applications in biomedical engineering and neural signal processing.

C. Okan Sakar is a research fellow at Bahcesehir University, and Ph.D. candidate at the Computer Engineering Department, Bogazici University. He received his M.Sc. degree from the Department of Computer Engineering of Bahcesehir University and B.Sc. degree from the Department of Mathematical Engineering of Yildiz Technical University. His research interests are in the field of pattern recognition.

Yasemin P. Kahya received the B.S. degrees in Electrical Engineering and in Physics from Bogazici University, Istanbul, Turkey, with highest honors in 1980, M.S. degree in Engineering and Applied Science from Yale University, Connecticut, USA, in 1981 and Ph.D. degree in Biomedical Engineering from Bogazici University in 1987. She is a professor in the Department of Electric and Electronic Engineering at Bogazici University, where has been teaching since 1988. Her current research interests are in the areas of biomedical instrumentation, respiratory acoustics and biomedical signal processing applications. She was the program co-chair of IEEE-EMBS 2001 Conference which was held in Istanbul and has served on the Adcom of IEE-EMBS.

Nizamettin Aydin received B.Sc. (1984) and M.Sc. (1987) degrees in Electronics and Communication Engineering at Yildiz Technical University, Turkey, and Ph.D. degree in Medical Physics (1994) at the University of Leicester, UK. Dr Aydin worked as an assistant professor in Electronics Engineering Department, Gebze Institute of Technology, Turkey, from 1995 to 1999. He worked in the Department of Clinical Neurosciences at Kings College London and the Division of Clinical Neuroscience at St George's Hospital Medical School as a Research Fellow between 1998 and 2001. He was a Senior Research Fellow in the Institute for Integrated Micro and Nano Systems, School of Engineering and Electronics at the University of Edinburgh from 2001 to 2004. Between 2004 and 2005 he was appointed as a head of computer Engineering Department and in 2006 he founded Software Engineering Department, both at Bahcesehir University, Istanbul. He is now with Computer Engineering and serves as the head of the department at Yildiz Technical University, Istanbul. He is a full member of IEEE since 1992. He was awarded the IEE Institute Premium Award for 2000/2001. His research interests include time–frequency and time–scale analysis, physiological measurements, Doppler ultrasound, digital signal processing, VLSI, system-on-chip design, bioinformatics, wireless communication, and scientific computing.