Celal Bayar University Journal of Science

Adventitious and Normal Respiratory Sound Analysis with Machine Learning Methods

Burcu Acar Demirci[1]*, Yücel Koçyiğit[2], Deniz Kızılırmak[3], Yavuz Havlucu[4]

[1,2] Manisa Celal Bayar University, Electrical and Electronics Engineering Department, Manisa, Türkiye
[3,4] Manisa Celal Bayar University, Hafsa Sultan Hospital, Chest Disease Department, Manisa, Türkiye
* burcu.acar@cbu.edu.tr
* Orcid No: 0000-0002-7328-1267

Received: 1 October 2021
Accepted: 27 May 2022
DOI: 10.18466/cbayarfbe.1002917

Abstract

Computerized respiratory sound analysis systems provide vital information regarding the current condition of the lung. These systems, used by physicians for the diagnosis of various respiratory diseases, help to classify respiratory sounds. Since physicians have differing degrees of knowledge and experience, this can cause differences in diagnosis and therefore treatment. Well-calibrated machine learning tools can help physicians make more informed decisions. For this purpose, different machine learning classifiers and feature extraction models have been developed to classify respiratory sounds from healthy individuals and patients. In this study, the combinations of Empirical Mode Decomposition, Mel Frequency Cepstral Coefficients, and Wavelet Transform methods are used for feature extraction, and *k*-Nearest Neighbor, Artificial Neural Networks, and Support Vector Machines are used for classification. The highest accuracy has been achieved as 98.8% when Mel Frequency Cepstral Coefficient and *k*-Nearest Neighbor methods are used in combination.

Keywords: Respiratory sound, mel-frequency cepstral coefficient, empirical mode decomposition, k nearest neighborhood method.

1. Introduction

Respiratory diseases cause millions of premature deaths in the world [1]. Therefore, early detection of respiratory diseases is a crucial medical research area. Computed Tomography (CT), pulmonary function test, chest x-rays, and lung auscultation are effective methods for diagnosing respiratory diseases [2]. Auscultation is the most commonly used method for the capturing the sounds that occur in the internal organs such as circulatory and respiratory systems, examining the current status of systems, and diagnosing diseases of these systems. Pulmonary auscultation, a commonly used method for capturing the sounds that occur in the internal organs, is the most straightforward and cheapest method used in the diagnosis of respiratory diseases [3]. The method, carried out through classical stethoscopes, provides critical information to physicians to diagnose respiratory diseases. Despite these properties, it has many limitations. It is a subjective process that depends on the physician's hearing ability, experience, and skill to distinguish between various sound patterns [2]. Another limitation of the method is that classical stethoscopes have a frequency response that reduces frequency components of lung sound signals above 120 Hz and that the human auditory system is not sensitive to the remaining low-frequency band. In recent years, Computerized Respiratory Sound Analysis systems (CORSA) have been used to overcome these limitations, and classical stethoscopes have been replaced by electronic stethoscopes that reduce noise, increase the volume and allow recording. Electronic stethoscope auscultation devices have evolved from analog to digital, and it has enabled storage, analysis, and visualization in computer systems. Nevertheless, digital auscultation is not yet a mature and complete computational procedure [4].

Many studies aim to leverage digital auscultation from the point of data and algorithms. The sounds recorded in the memory of the stethoscope are analyzed with CORSA systems. The availability of CORSA systems has led to increased research in the field of lung sounds. Adventitious respiratory sounds associated with specific disease were compared with normal respiratory sounds in many studies [5-13]. For example, Corbera *et al.* [5] focused on wheeze respiratory sound because of widespread asthma and aimed to detect significant differences between wheeze and healthy sounds. In this regard, the study focused on identifying wheezing attacks from spectrograms by applying the temporal and spectral continuity criteria to the previously detected peaks. Sezgin *et al.* [6] used two types of respiratory sounds: normal and patient. In this study, the decision process comprises three stages: the normalization process, the feature extraction process, and the classification process. In the feature extraction process, the features are determined with wavelet analysis; afterward, the optimum features are selected by dynamic programming. The classification process was made with Artificial Neural Networks (ANN). Maruf *et al.* [7] have developed CORSA systems consisting of four modules to detect crackle sounds automatically. A 100-2500 Hz bandpass filter was applied to respiratory sounds in the pre-processing module and in the feature extraction module preferred Wavelet Transform (WT). They identified the best four features in the feature selection module and classified them using Gaussian Mixture Model (GMM), Support Vector Machines (SVM) and ANN methods in the classification module. Lozano *et al.* [8] have suggested a different method for automatically diagnosing normal and wheezing sounds. Empirical mode decomposition-based methods have used in the study. The proposed methods have eliminated the mode mixing problem of the Empirical Mode Decomposition (EMD) method and have offered high energy concentrations, high time, and high-frequency resolutions. The tests applied showed that the proposed Ensemble Empirical Mode Decomposition-Kay based Hilbert spectrum better results in wheezing sound detection .

The success of cepstral features in respiratory sound classification has inspired many studies [9], [10-12]. Palaniappan *et al.* [9] have used parenchymal pathology, obstructive pathology, and normal respiratory sounds. The Mel-Frequency Cepstral Coefficient (MFCC) method, a highly efficient feature extraction algorithm used in the processing of audio signals, was used in this study. The 13 cepstral coefficients obtained by MFCC have classified using the k- Nearest Neighbors Algorithm (k-NN) and SVM. Sengupta *et al.* [10] suggested a new set of cepstral features to classify normal, wheeze, and crackle sounds by considering the achievement of cepstral features in the classification of speech sounds.

Liu *et al.* [11] have proposed examining the normal, wheeze, and crackle respiratory sounds in time, frequency, and cepstral domains. They extracted 46 features, and 6 crucial features were selected from the obtained features. GMM to classify these three respiratory sounds is proposed. Sunil and Ganesan [12] have proposed an efficient method to classify normal and abnormal respiratory sounds. These sounds are analyzed by the MFCC and classified by the Adaptive Neuro-Fuzzy Inference System. Haider *et al.* [13] have suggested using auscultation and pulmonary function tests together in the study. A total of 39 features of respiratory sounds and 3 spirometry features were used. Various parametric and nonparametric tests have been conducted to determine the similarity level of the extracted features. Logistic Regression (LR) , Decision Tree (DT), Discriminant Analysis (DA) , SVM and k-NN have been used to classify normal and Chronic Obstructive Pulmonary Disease (COPD) respiratory sounds.

This study aims to diagnose normal and abnormal respiratory sounds similar to abovementioned studies. However, we used methods such as Empirical Mode Decomposition, and combined it with established classification methods to potentially identify the most accurate method. The success of the proposed method has been compared to frequency analysis and cepstral analysis. The cepstral and frequency analyses have been done by using MFCC and WT methods, respectively. Comparing the performances of different classifiers on the problem of classification of respiratory sounds in the literature is a common practice. It is also compared in classifiers such as ANN, k-NN, and SVM within this study's scope. The rest of the article's organization is as follows: Section 2 briefly describes the material and methods used in our study. It involves Respiratory Sound Acquisition, Preprocessing, Feature Extraction, and Classification Methods. Results and Discussion are presented in Section 3, and the Conclusions are presented in Section 4.

2. Materials and Methods

The rate, time, and sounds of respiratory are essential in diagnosing respiratory system diseases. Inspiration and expiration stages of respiratory sounds contain important information about the respiratory system. CORSA systems provide a new perspective on the detection and treatment of many diseases. Generally, the systems consist of two steps. Firstly, the crucial features of respiratory sounds are extracted. Secondly, these crucial features are used for detecting or classifying adventitious respiratory sounds [14]. Commonly preferred methods for feature extraction in the literature are MFCCs [9], spectral features [15], Fourier [16], The Autoregressive Model [17], and Wavelet coefficients [18]. For classification, algorithms such as ANN [19], SVM [20], GMM [11], k-NN [9], and LR models [13] are used. Figure 1 shows the block diagram for all system stages recommended in this study. The system is divided into four stages: the acquisition of respiratory sound, pre-processing, feature extraction, and classification.

Celal Bayar University Journal of Science
Volume 18, Issue 2, 2022, p 169-180
Doi: 10.18466/cbayarfbe.1002917

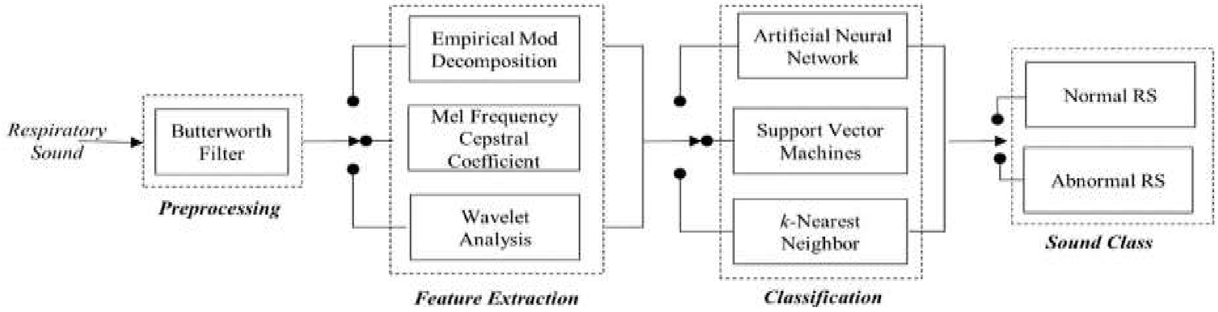

Figure 1. Block Diagram of the recommended system.

2.1 Respiratory sound acquisition

During the inspiration and expiration stage of respiration, vibrations occur as a result of rapid changes in gas pressure in the airways. Respiratory sounds occur when the vibrations pass through the lung tissue and reach the chest wall. Changes in vibration create sounds with a certain amplitude and frequency [21]. Respiratory sounds are divided into normal and adventitious (abnormal). Normal respiratory sounds are those are heard when there is no pathological airflow in the airways. Adventitious respiratory sounds are caused by pathological effects in the lungs or respiratory tract. A data set consisting of wheeze and rhonchi adventitious sounds and normal sounds was used in this study. The normal lung sounds are both louder and larger amplitude sounds during the inspiration stage than during the expiration stage. The signal frequency band of the sounds is between 150-1000 Hz [22]. Wheeze and rhonchus sounds are determined airway obstruction pathology. They are common signs of obstructive lung diseases like asthma or COPD. Wheeze respiratory sounds are musical, continuous, and coarse sounds commonly heard during the expiration stage as a result of high-speed airflow through narrowed airways [18]. Some parts of the respiratory tree must be narrowed or obstructed for the wheezing adventitious sound to occur. [23]. Rhonchus respiratory sounds are low-pitched and continuous sounds that result from obstruction or secretions in larger airways heard during the inspiration and expiration stage. According to the American Thoracic Society (ATS), wheezes have a dominant frequency of 400 Hz or more, while rhonchus has a dominant frequency of about 200 Hz or less, and the event is longer than 250 ms [1].

In this study, the Respiratory Sounds (RS) were recorded by specialist physicians in the Hafsa Sultan Hospital, Manisa Celal Bayar University. All records were obtained using Littmann 3200 Electronic stethoscope from 25 healthy and 25 patient volunteers treated in the clinic of respiratory medicine of the hospital. The study population was picked among the patients who have different demographic attributes and lack previous comorbidities of the study population. Normal respiratory sounds were recorded by selecting 7 female and 18 male volunteers among volunteers who had never smoked or used tobacco products. Wheeze respiratory sounds were recorded from 12 volunteers, 4 females and 8 males, with asthma or COPD. Rhonchus breath sounds were recorded from 13 volunteers, 7 females and 6 males, with Pneumonia and Chronic bronchitis. Each volunteer was asked to breathe in and out of the mouth four times, and the recording was made to include four full breaths. Thus, 100 normal, 52 rhoncus, and 48 wheeze RS were obtained. All sounds sampled at a frequency of 11025 Hz were recorded by the auscultation protocol determined by specialist physicians. According to this protocol, sounds were recorded in a calm environment, with the patient sitting and loosening his/her posture muscles. The records are obtained from the most appropriate places for the maximum collection of data about patients' pathologies, as determined by the CORSA standard [24].

100 abnormal and 100 normal respiratory sounds were used in the study. The Hotelling T-squared statistical method [25] was used to evaluate the adequacy of the data numbers. Hotelling's T-squared statistical method is used to determine whether there is a significant difference between the two groups for multivariate samples. The F distribution can be a good approximation of the T-squared statistic distribution when the dimension of the data is less than the size of the samples. As shown in the table, the results of Hotelling's T-squared statistics revealed a significant difference in all coefficients for normal and abnormal respiratory sounds

Table 1. Hotelling's T- squared Test For Normal and Abnormal Respiratory Sounds.

	T^2	*F*	*p Value*
MFCC	551.779	F(78,121)=7.878	0.0000000000
EMD	141.140	F(30,169)=4.016	0.0000000038
WT	118.729	F(30,169)=3.378	0.0000002985

2.2 Pre-processing

Pre-processing aims to reduce background noise and improve the quality of recorded respiratory sounds [7]. When respiratory sounds are recorded, they are affected by low-frequency sounds such as muscles and heart sounds and high-frequency noises due to sudden movements. To avoid these effects, pre-filtering should be performed considering the dominant frequency range of the respiratory sounds. Various filters such as Butterworth, Chebyshev, and Elliptical filters of varying degrees have been tested for the study. The accuracy and success of the filtering have been inspected by experts listening to the recorded respiratory sounds, and the filter has been selected. In this study, are used a fourth-order bandpass Butterworth filter. The frequency range of this filter is 100-1800 Hz [26].

2.3 Feature extraction

The feature extraction process enables converting high-dimensional vectors to lower-dimensional vectors [16]. The properties of signals formed after pre-processing the respiratory sounds are analyzed simultaneously in the time, frequency, or time-frequency domain by feature extraction methods. In this study, some feature extraction methods such as EMD, MFCC, and WT are used. The coefficients of respiratory sounds are determined by using these methods. The coefficients obtained by feature extraction methods are not used directly. The feature vectors constructed with statistical parameters such as the standard deviation, variance, mean power, entropy, mode, and energy values of the coefficients are used instead of the coefficients. The novelty of the paper is the combination of the EMD method and statistical analysis methods for feature extraction.

2.3.1. Empirical mode decomposition

Many biomedical signals such as EEG signals, EMG signals, and respiratory sounds have a nonlinear and non-stationary structure. Wigner-Ville Transform, Short Time Fourier Transform, and WT are widely used to analyze these types of signals. New methods are being investigated due to different restrictions on each. EMD has been proposed as an alternative and appropriate tool for analyzing multicomponent nonlinear and non-stationary signals [27]. EMD is an adaptive and direct decomposition method. Unlike Fourier and WT analysis, it is unnecessary to have prior knowledge of the signal properties to select parameters in this analysis [20]. EMD allows the target signal is separated into Intrinsic Mode Functions (IMFs) listed from the high-frequency components to low-frequency components [28]. A sifting process based on the estimated upper and lower envelopes of the input signals is used to obtain the IMF. There are two conditions for the statements obtained as a result of the sifting process to be IMF. First, the number of extrema and zero-crossings must be equal or differ by one at most in the whole data set. Second, the upper and lower envelopes' local average should be zero at any point [29].

The IMF components of x(t) signal is obtained by the following algorithm: [28, 30]

1. Find the local minima and local maxima of $x(t)$
2. The maximum envelope $e_{\max}(t)$ is calculated using local maxima points with cubic spline interpolation. Similarly, minimum envelop $e_{\min}(t)$ is calculated using local minima.
3. The local average envelope is found by taking the average of the maximum and minimum envelope:

$$m_1(t) = [e_{\min}(t) + e_{\max}(t)]/2 \quad \textbf{(1)}$$

4. The local average envelope is removed from the original signal:

$$h_1(t) = x(t) - m_1(t) \quad \textbf{(2)}$$

5. Whether $h_1(t)$ to become an IMF is checked. If the conditions are met, it is considered $IMF_1(t) = h_1(t)$. If $h_1(t)$ is not an IMF, $h_1(t)$ is considered a new signal. The loop is repeated using $h_1(t)$ to create $h_2(t)$. If $h_2(t)$ is not an IMF, the stop criterion is calculated to end the elimination process. The formula for the stop criterion, SD, is set out below.

$$\mathrm{SD}(i) = \sum_{t=0}^{N} \frac{|h_{i-1}(t) - h_i(t)|\text{^}2}{h^2{}_{i-1}(t)} \quad \textbf{(3)}$$

   If $h_2(t)$ meets SD, $IMF_1(t) = h_2(t)$. If it does not meet the stop criteria, $h_2(t)$ is treated as a new signal. During steps 1-6, its operations on $h_2(t)$ are repeated to form $h_i(t)$ until $h_i(t)$ meets the requirements of the IMF or SD. It is calculated $IMF_1(t) = h_i(t)$.

6. $IMF_1(t)$ is formed. By subtracting $IMF_1(t)$ from $x(t)$ signal, a residual signal $r_1(t)$ is obtained. To find the next residual signal, $r_1(t)$ is considered the original signal, and steps 1-6 are repeated. When the process is completed, the original signal is composed of several IMF components and residual $r_n(t)$. And is expressed as follows;

$$\mathrm{x}(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \quad \textbf{(4)}$$

IMF coefficients vary according to the state and frequency distribution of respiratory sounds. In the study, it has been observed that the amplitude of the IMF coefficients at Level 5 and beyond is very low and it has been determined that it contains redundant information. For this reason, 4 IMF coefficients are used. IMF coefficients of respiratory sounds obtained by the EMD method are given in Figure 2, Figure 3, and Figure 4.

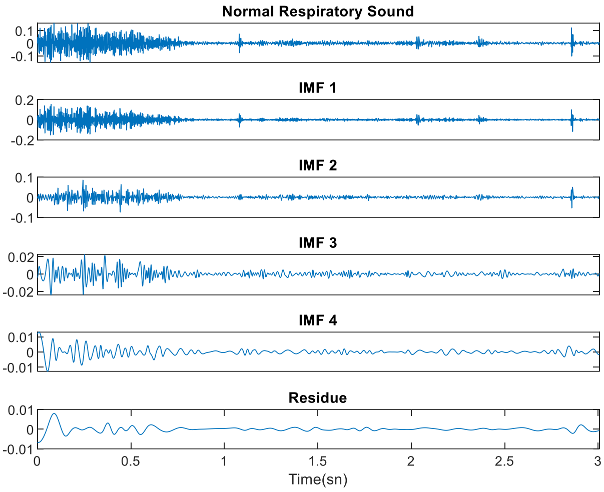

Figure 2. First 4 IMF coefficients and residue signal for Normal RS.

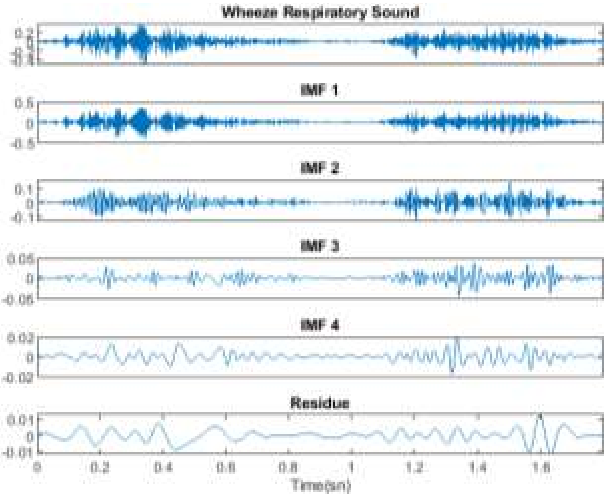

Figure 3. First 4 IMF coefficients and residue signal for Wheeze RS.

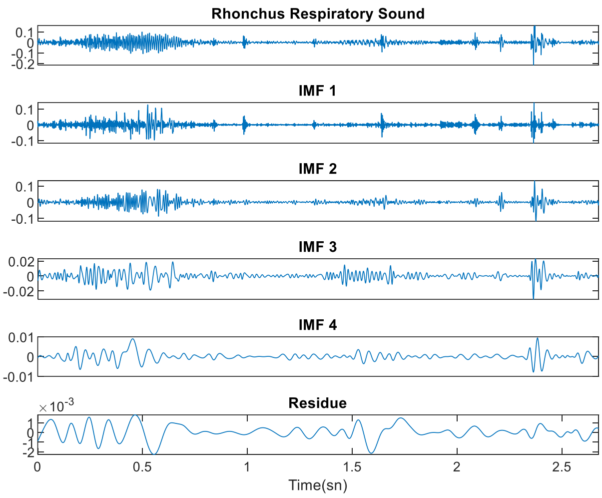

Figure 4. First 4 IMF coefficients and residue signal for Rhonchus RS.

2.3.2 Mel-frequency cepstral coefficient

MFCC is a highly effective feature extraction algorithm commonly used for automatic speech and speaker recognition [12]. This is because MFCC can distinguish speakers with high accuracy by imitating the frequency selectivity of the human ear. In addition, MFCCs are often preferred because they are much less affected by changes and sound wave structure.

In recent years, the MFCC method has been used in CORSA systems in many studies, and promising results have been obtained [31, 32]. The MFCC performs a nonlinear scaling, assuming that the audio signal's low-frequency components carry more critical information than the high-frequency components. MFCC analysis is similar to cepstral analysis, apart from frequency wrapping. In MFCC analysis, the frequency is wrapped according to the Mel-Scale [31]. There are several methods in the literature for calculating MFCC. Fast Fourier Transform (FFT) based method is one of the most commonly used methods among them. The block diagram of this method used to calculate the MFCC features is shown in Figure 5.

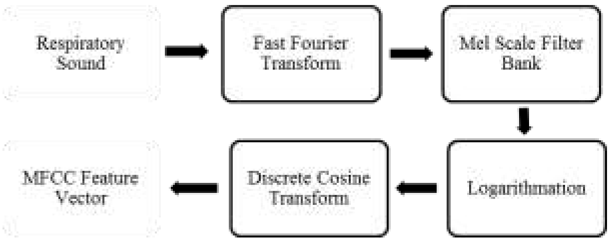

Figure 5. Block diagram used to calculate MFCC Features Vector.

In this method, firstly, FFT is applied to the windowed signal. A triangle bandpass filter bank known as a Mel-scale filter bank is applied to the obtained FFT spectrum. The Mel-Scale, designed based on the human hearing system, is based on mapping the actual frequency and the perceived pitch. This scale consists of linear ranges up to 1 kHz and logarithmic ranges after 1 kHz. The mapping of linear frequency to Mel-frequency is done by applying equation (5).

$$f_{mel} = 2595 * \log\left(1 + \frac{f_{lineer}}{700}\right) \quad \textbf{(5)}$$

A logarithm process is applied to the signal filtered using a Mel-Scale filter bank. With this process, the sensitivity of the feature vectors to changes is reduced. Discrete cosine transform is applied to the logarithmic scale applied signal finally, and the signal is converted back to the time domain. Discrete cosine transform is applied to the logarithmic scale applied signal finally, and the signal is converted back to the time domain. Thus, MFCCs with the amplitude of the spectrum are obtained. In this study, 13 MFCCs are used to classify RS. The obtained results are presented in Figure 6, Figure 7, and Figure 8.

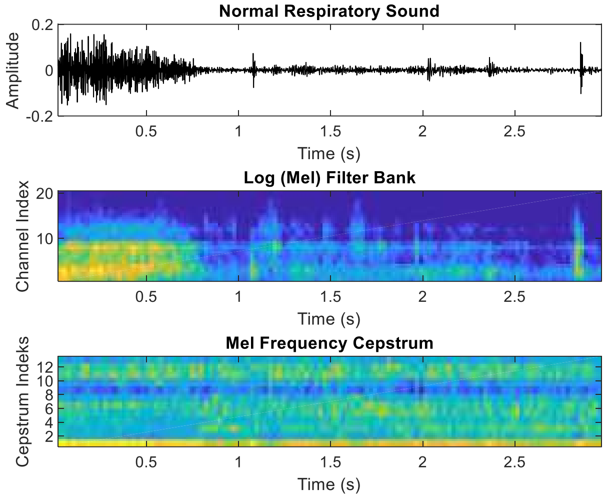

Figure 6. MFCCs of Normal RS.

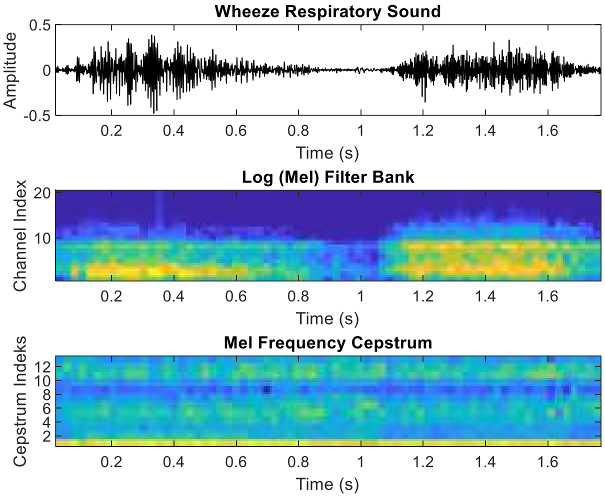

Figure 7. MFCCs of Wheeze RS.

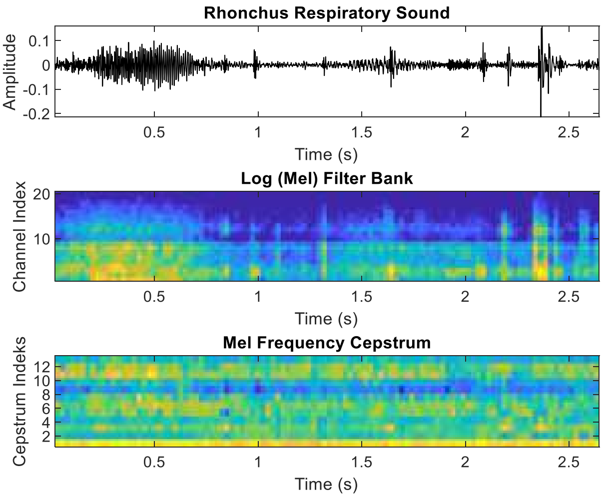

Figure 8. MFCCs of Rhonchus RS.

2.3.3 Wavelet transform

WT is a signal processing method used as an alternative to Fourier Transform (FT) [33]. FT is an analysis method used to analyze stationary signals defined in the time domain, providing frequency information by examining the signal in the frequency domain. However, only frequency analysis is not sufficient for dynamic and non-periodic signals. WT is a commonly used method for non-stationary, nonlinear, and non-periodic signal analysis, such as lung sounds. With WT, the signal is defined in both the time domain and frequency domain, thus providing information on how the signal's frequency components vary with time. WT uses short window size when high-frequency information is essential, while long window size uses when low-frequency information is important [34].

In Discrete Wavelet Transform (DWT), the signal is separated into sub bands by passing through the high-pass filter (h) and low-pass filters (g) according to the determined level. Equation conditions that these filters must meet are as follows; [35]

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 1 \quad \textbf{(6)}$$

$$H(z) = zG(-z^{-1}) \quad \textbf{(7)}$$

A sequence of filters with increasing length (indexed by i) can be obtained;

$$G_{i+1}(z) = G\left(z^{2^i}\right)G_i(z) \quad \textbf{(8)}$$

$$H_{i+1}(z) = H\left(z^{2^i}\right)G_i(z) \qquad \text{i=0, 1……, i-1} \quad \textbf{(9)}$$

With the initial condition $\mathbf{G_0}(z) = 1$. It is expressed as a two-scale relation in the time domain

$$g_{i+1}(k) = [g]_{\uparrow 2^i} g_i(k), h_{i+1}(k) = [h]_{\uparrow 2^i} g_i(k) \quad \textbf{(10)}$$

where the subscript $[.]\uparrow_m$ indicates the up-sampling by a factor of **m** and **k** is the equally sampled discrete-time. The normalized wavelet and scale basis functions $\varphi_{i,l}(k), \psi_{i,l}(k)$ can be defined as

$$\varphi_{i,l}(k) = 2^{\frac{i}{2}}\ g_i(k - 2^i l) \quad \textbf{(11)}$$

$$\psi_{i,l}(k) = 2^{\frac{i}{2}}\ h_i(k - 2^i l) \quad \textbf{(12)}$$

where the factor $\mathbf{2^{i/2}}$= inner product normalization, **i**= scale parameter and **l**= the translation parameter. The DWT decomposition can be described as

$$a_i(l) = x(k) * \varphi_{i,l}(k) \quad \textbf{(13)}$$

$$d_i(l) = x(k) * \psi_{i,l}(k) \quad \textbf{(14)}$$

$\boldsymbol{a_i}$ is approximation coefficient and $\boldsymbol{d_i}$ is detail coefficient. The approximation coefficients represent the lower frequency band, and the detail coefficients represent the higher frequency band.

These coefficients are used for the classification process in many signal processing applications. In the analysis of signals using DWT, the selection of the appropriate main wavelet function and the determination of the appropriate decomposition level are very important. The main wavelet function, which is one of the most important parameters of the wavelet transform, takes on the task of the window function in the Fourier transform. There are many main wavelet functions with different properties and uses. In previous studies on the application of DWT in respiratory sound analysis, Daubechies 8 (db8) main wavelet function was used and found to give good results. Therefore, db8 is also preferred in this study [18]. Another important parameter is the number of decomposition levels, determined according to the dominant frequency components of the signal. In the study, the number of decomposition levels is chosen to be 7. Thus, respiratory sounds are decomposed into detail coefficients D1-D7 and approximation coefficient A7. Since the frequency range of D3-D7 sub-bands carries important information, these sub-bands are preferred.

2.4 Classification

In CORSA systems, the classification stage comes after the feature extraction stage. For classification, are used *k*-NN, SVM, and ANN in this study. SVM and ANN classifiers have two stages training and testing. During the training stage, data from each RS class is introduced to the system as training data, and the system makes a distinction by class. Unknown sounds are analyzed during the test stage, and the most appropriate class is selected [19].

In the kNN classifier, which is instance-based learning, no training phase is required [36]. Samples divide into training and test samples. Training samples are multidimensional vectors, each with a class label. In the classification phase, unlabeled test vectors are labeled by taking into account the closest k training examples.

In classifier algorithms, the effects of model parameters on performance and the effects of these parameters on classifier capacity and complexity were observed, and the most suitable model parameters were determined. The parameters selected for each method are given in the relevant section. The classification performance of a classifier in medical tests depends on the ability to detect patients and healthy people. In this study, standard parameters such as sensitivity, specificity, and accuracy were used for performance evaluation. Sensitivity is the ratio of the number of correctly classified patients to the total number of patients, while specificity is the ratio of correctly classified healthy people to the total number of healthy people. Accuracy is the ratio of the number of sick and healthy people correctly classified to the total number of people [19].

2.4.1 Support Vector Machines

SVM method, based on statistical learning theory, was developed by Vladimir Vapnik in 1992. SVM, a supervised learning algorithm, is used to solve classification, regression analysis, and nonlinear function approach problems [37]. Provides high classification and high generalization performance in solving bioinformatics problems, text, voice, object, and image recognition problems [38].

SVM applies a useful learning algorithm to identify difficult-to-analyze patterns in complex data sets. In SVM, the objective is to create an n-dimensional hyperplane that optimally divides the data into different classes. It is also used to obtain the optimal limit of two data sets on a vector space, independent of the probability distributions of training vectors in sets. Like artificial neural networks, SVM models have a two-layer, feed-forward network structure that uses a sigmoid kernel function. Some of the commonly used kernel functions are RBF, linear, quadratic, and polynomial kernel. In this paper, second-order polynomial kernel functions are used.

2.4.2 *k* Nearest Neighbors Algorithm

k-NN is a supervised and nonparametric classification method that classifies data based on the proximity of training samples in the data set. This classification method finds the *k* nearest neighbors of unknown data between the dataset according to a distance equation. Then, it uses the majority vote approach to estimate the data label [39]. Distance equations such as Manhattan, Hamming, Euclidean, and Minkowski are used for distance calculation. In this study, the Euclidean distance equation was used to locate the nearest neighbor.

The basic steps to be applied for classification with the *k*-NN algorithm are as follows:

1. The number *k* is determined.
2. The new data is evaluated individually with all the data in the training data set, and the distances between them are calculated by distance functions.
3. The *k* data closest to the new data is selected.
4. The class to which most of the selected data belongs is determined, and the new data is assigned to this class.

In this study, the results were obtained for *k*=1 and *k*=3.

2.4.3 Artificial Neural Network

ANN is a topological structure created for a specific purpose inspired by the neuron functioning of the brain. The structure consisting of interconnected artificial neurons is widely used in various recognition, prediction, and modeling fields, as well as the recognition and classification of biological signals. In ANN, learning is carried out with special training algorithms that imitate the learning mechanisms of biological systems [35].

In this study, various model trials were conducted while determining the ANN classifier model. In models with three hidden layers, the excess number of layers and the number of neurons in the layers increases the processing and learning ability of the artificial neural network but reduces the generalization ability of the network and causes overfitting. In models with a single hidden layer, the pattern in the data cannot be learned sufficiently and underfitting occurs. For this reason, a model with two hidden layers is preferred. Tests were made for the number of neurons in the network (45:45), (45:30), (30:15), (15:15) and (10:10). The most successful result (15:15) was obtained with the number of neurons and the hyperbolic tangent activation function. Backpropagation (BP) algorithm, which is the most frequently used training algorithm for multilayer feedforward networks, and Mean Error Squares the most frequently used performance function, are also preferred in this study. Levenberg–Marquardt learning algorithm, which creates a balanced system structure in the network structure and reduces the processing load, is preferred.

3. Results and Discussion

In this study, while classifying with the ANN method, 80% of the data was used for training and 20% for testing. While using SVM and k-NN methods for classification, training and test groups were determined by applying 10 cross-validations to the data. Besides, we have iterated the whole classification method 10 times, and average performance values have been calculated.

The sensitivity, specificity, and accuracy parameters obtained with the proposed system are recorded in Tables 2, 3, and 4. Table 2 includes a comparison of MFCC-based features, Table 3 EMD-based features, and Table 4 WT-based features with different classifiers. Each classifier has its advantages and disadvantages. The *k*-NN algorithm has advantages such as no training required, being easy to perform, being analytically tractable, adaptable to local information, and resistant to noisy training data. There are also disadvantages, such as the need for a high amount of memory space, and the processing load and cost increase significantly as the data set and attribute size increase [40]. This method is crucial to selecting the *k* value from the optimal value; as the k value decreases, more sensitive results were obtained. The major advantages of ANN over traditional statistical techniques are that it requires fewer assumptions and can model nonlinear relationships depending on the choice of activation functions. Neural network models can learn to complex nonlinear relationships between independent and dependent variables, and they can make logical decisions in the face of similar events. The information is stored throughout the network, and some of the artificial nerve cells do not function, causing the loss of information. However, it has disadvantages, such as requiring excess computational overhead and having limited ability to identify possible causal relationships. Using trial and error in determining parameters such as the selection of activation function, the number of hidden layers,and neurons is one of the most significant disadvantages [41]. SVM models provide good scaling for high-dimensional data and can be used for both linear and nonlinear applications similar to artificial neural networks. Also, there is less risk of overfitting. But choosing the appropriate kernel function is not straightforward [42].

This study used ANN and SVM classifier models, determining the optimum selection of system parameters by trial-and-error methods creates a disadvantage. Furthermore, the fact that the data size is not too big and no training is required has provided an advantage for k-NN. The highest accuracy rate obtained using the MFCC and WT features were respectively 98.8% and 82.7% with the *k*-NN classifier. The highest accuracy rate achieved using the EMD features is 88.9% with the ANN classifier. Additionally, we observed that MFCCs provide the best results among all the feature extraction methods examined. MFCC analysis has been more successful in evaluating respiratory sounds because it is closer to the response of the human auditory system. MFCC can distinguish speakers with high accuracy by imitating the frequency selectivity of the human ear. The auditory perception-specific information captured by MFCCs reveals differences in respiratory sounds, and the use of these differences for diagnostic purposes has increased success. Providing a more successful analysis than FT and WT, the EMD method has been proven to have a mode mixing effect when applied to some respiratory sounds, as noted in earlier studies [20]. The mode mixing effect causes some frequency components of abnormal RS to occur in different IMFs, which has a negative effect on the success of the method.

The lack of large databases publicly available to develop algorithms and compare results is one of the field's most significant problems [4]. It is quite difficult to compare the performance of the studies due to different classifier models, different respiratory sounds classified, and different feature vectors used in the classification. Table 5 shows the results of the standard parameters of various studies in this field.

Table 2. Classification Result of MFCC Features.

	MFCC Features		
Classifier	***Sensitivity***	***Specificity***	***Accuracy***
ANN	100	97 ± 3.3	98.7± 1.6
SVM	97.7 ± 1.1	93.3± 1.1	95.7 ± 0.9
k-NN (k=3)	99.8 ± 0.6	93± 0.8	96.3 ± 0.6
k-NN (k=1)	99.5 ± 0.9	97.5 ± 0.8	**98.8 ± 0.6**

Table 3. Classification Result of EMD Features.

	EMD Features		
Classifier	***Sensitivity***	***Specificity***	***Accuracy***
ANN	93.5 ± 3.9	84 ± 6.3	**88.9 ± 3.2**
SVM	78.2 ± 2.1	77.1 ± 1.8	77.9 ± 1.6
k-NN (k=3)	81 ± 1.4	76.7 ± 1.8	79.1 ± 1.6
k-NN (k=1)	77.8 ± 1.8	74.3 ± 0.9	76.3 ± 1.2

Table 4. Classification Result of Wavelet Features.

	WAVELET Features		
Classifier	***Sensitivity***	***Specificity***	***Accuracy***
ANN	83 ± 7.4	74.5 ± 5.2	79 ± 3.7
SVM	77.9 ± 1.2	76.3 ±1.6	77.3± 1.3
k-NN (k=3)	82.1 ± 1.5	80.5 ± 1.9	81.6± 1.6
k-NN (k=1)	83.2 ± 1	82 ± 1.2	**82.7± 0.9**

4. Conclusion

The CORSA systems provide vital information about the current state of the lung. The proposed system diagnoses the disease by separating the respiratory sounds as normal and abnormal. While EMD, MFCC, and WT feature extraction methods are used for disease diagnosis, ANN, SVM, and k-NN classifiers are used for classification. Since respiratory sounds are not stationary and linear, classical frequency analysis methods are not adequate in the analysis of these sounds. For extraction of good acoustic characteristics from respiratory sounds, it is required to examine these sounds over sufficiently short periods of time. WT and MFCC methods are preferred in this study because of enabling short-time analysis. MFCC is a representation of the short-time power spectrum of the sound signal. Signal cepstrum is obtained by inverse transforming of the logarithm of the signal spectral representation. The frequency bands in mel-frequency cepstral representation of the power cepstrum are equally spaced on the mel-scale approximating the human auditory perception. EMD analyses are non-linear and non-stationary data without the assumption of linearity or short-time stationarity. EMD process does not involve a fixed basis but rather has a signal-specific approach to decompose the signal. The application of the EMD method for respiratory sounds is a quite new method. In the previous studies, IMF coefficients are obtained by using only the inspiration stage of respiration. In this study, both inspiration and expiration stages of respiration are used to obtain IMF coefficients. It is aimed to compare the EMD method with traditional feature extraction methods for respiratory sounds. In order to evaluate the performance of the coefficients obtained by different feature extraction methods, classification has been made with the statistical parameters. EMD method gives more successful results than the WT method. EMD provides lower success compared to the MFCC method because the MFCC analysis is more successful in assessing respiratory sounds as it is closer to the response of the human auditory system.

This comparative study shows that MFCC features are more successful in diagnosing respiratory sounds compared to the other features. The highest accuracy rate is obtained for the *k*-NN classifier with 98.8%.

Author's Contributions

Burcu Acar Demirci: Methodology, Software, Validation, Formal analysis, Data Curation, Writing-Original Draft, Visualition
Yücel Koçyiğit: Conceptualization, Funding aquisition, Writing-Review&Editing Draft, Supervision, Project administration
Deniz Kızılırmak: Investigation, Resources, Data Curation
Yavuz Havlucu: Investigation, Resources, Supervision

Acknowledgments & Ethics

This study is supported by Manisa Celal Bayar University Scientific Research Project Coordination Unit [Project No. 2017-191]. Within this study's scope, the recording of respiratory data was performed with the approval of the ethics committee number 20.478.486, dated 20.10.2017

Celal Bayar University Journal of Science
Volume 18, Issue 2, 2022, p 169-180
Doi: 10.18466/cbayarfbe.1002917

Table 5. Some Results obtained with CORSA Systems

Ref.	Analyzed: Sound/ Disorder	Dataset	Feature Extraction Method	Classification Method	Best Result
[5]	Normal, Wheeze	N:15 W:16	The Local Adaptive Wheezes Detection Algorithm (LAWDA)		Sensitivities: 71% - 100%
[6]	Normal, Abnormal	N:10 A:10	Wavelet Analysis	Artificial Neural Network	Accuracy: 98 %
[7]	Normal, Crackle	N:27 C:14	Wavelet Analysis	Artificial Neural Network Support Vector Machine Gauss Mixture Model	GMM: Accuracy: 97.56 % Sensitivity:92.85 % Specificity:100 %
[8]	Normal Abnormal	N:385 A:485	Ensemble Empirical Mode Decomposition	Support Vector Machine	Accuracy: 94.6 % Sensitivity:94.2 % Specificity:95 %
[9]	Normal, Obstruction, Parenchymal	N:17 O:26 P:25	Mel-Frequency Cepstral Coefficient	Support Vector Machine *k*-Nearest Neighbors	*k*-NN: Accuracy: 98.26 %
[10]	Normal, Wheeze, Crackle,	Total:30	Statistical Properties Of Cepstral Coefficients Based New Feature Set	Artificial Neural Network	Accuracy: 97.2 % Sensitivity:97.41 % Specificity:95.5 %
[11]	Normal, Wheeze, Crackle,	N: 20 W:20 C:20	Mel-Frequency Cepstral Coefficient	Gaussian Mixture Model	Accuracy: 98.4 %
[12]	Normal Abnormal	N:17 A:51	Mel-Frequency Cepstral Coefficient	Adaptive Neuro-Fuzzy Inference System	Accuracy: 97.25 % Sensitivity:99.37 % Specificity:95.3 %
[13]	Normal, COPD patient	N:25 P:30	Temporal, Spectral, and Spectra-temporal features (MFCC LPC, etc.)	Support Vector Machine k-Nearest Neighbors Logistic Regression Decision Tree and Discriminant Analysis	SVM and LR: Accuracy: 100 % Sensitivity:100 % Specificity:100 %
This Study	Normal Abnormal	N:100 A:100	Mel-Frequency Cepstral Coefficient Empirical Mode Decomposition Wavelet Analysis	Artificial Neural Network Support Vector Machine k-Nearest Neighbors	MFCC + k-NN: Accuracy: 98.8 % Sensitivity:99.5 % Specificity:97.5 % EMD+ANN: Accuracy: 88.9 % Sensitivity:93.5 % Specificity:84 %

References

[1]. World Health Organization. The Top 10 Causes of Death n.d. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed March 20, 2020).

[2]. Güler I, Polat H, Ergün U. 2005. Combining neural network and genetic algorithm for prediction of lung sounds. *Journal of Medical Systems*; 29:217–31. doi: 10.1007/s10916-005-5182-9.

[3]. Polat H, Güler I. 2004. A simple computer-based measurement and analysis system of pulmonary auscultation sounds. *Journal of Medical Systems*; 28: 665–72. doi: 10.1023/B:JOMS.0000044968.45013.ce.

[4]. Rocha BM, Filos D, Mendes L, Serbes G, Ulukaya S, Kahya YP, et al. 2019. An open access database for the evaluation of respiratory sound classification algorithms. *Physiological Measurement*; 40. doi:10.1088/1361-6579/ab03ea.

[5]. Homs-Corbera A, Fiz JA, Morera J, Jané R. 2004. Time-Frequency Detection and Analysis of Wheezes During Forced Exhalation. *IEEE Transactions on Biomedical Engineering*; 51: 182–186. doi: 10.1109/TBME.2003.820359.

[6]. Sezgin MC, Dokur Z, Olmez T, Korurek M. Classification of respiratory sounds by using an artificial neural network. *2001*

Proceedings of the 23rd Annual EMBS International Conference, Istanbul, Turkey, 2001, pp. 697-699. doi:10.1109/IEMBS.2001.1019035.

[7]. Maruf SO, Azhar MU, Khawaja SG, Akram MU. Crackle separation and classification from normal respiratory sounds using Gaussian Mixture Model. *2015 IEEE 10th International Conference on Industrial and Information Systems,* Sri Lanka, 2015, pp. 267-271 doi: 10.1109/ICIINFS.2015.7399022

[8]. Lozano M, Fiz JA, Jané R. 2016. Performance evaluation of the Hilbert-Huang transform for respiratory sound analysis and its application to continuous adventitious sound characterization. *Signal Processing*; 120: 99–116. doi:10.1016/j.sigpro.2015.09.005.

[9]. Palaniappan R, Sundaraj K, Sundaraj S. 2014. A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. *BMC Bioinformatics*; 15:223. doi:10.1186/1471-2105-15-223.

[10]. Sengupta N, Sahidullah M, Saha G. 2016. Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*; 75: 118–29. doi: 10.1016/j.compbiomed.2016.05.013.

[11]. Liu Y, Lin Y, Zhang X, Wang Z, Gao Y, Chen G, et al. Classifying respiratory sounds using electronic stethoscope. *2017 IEEE SmartWorld, Ubiquitous Intelligence& Computing, Advanced& Trusted Computed, Scalable Computing& Communications, Cloud& Big Data Computing, Internet of People and Smart City Innovation*, San Francisco, CA, USA, 2017. doi: 10.1109/UIC-ATC.2017.8397496

[12]. Sunil NKB, Ganesan R. 2015. Adaptive neuro-fuzzy inference system for classification of respiratory signals using cepstral features. *International Journal of Applied Engineering Research*; 10 (28): 22121-22125.

[13]. Haider NS, Singh BK, Periyasamy R, Behera AK. 2019. Respiratory sound based classification of chronic obstructive pulmonary disease: a risk stratification approach in machine learning paradigm. *Journal of Medical Systems*; 43: 255. doi: 10.1007/s10916-019-1388-0.

[14]. Pramono RXA, Bowyer S, Rodriguez-Villegas E. 2017. Automatic adventitious respiratory sound analysis: A systematic review. *Plos One*; 12(5): 1-43. doi: 10.1371/journal.pone.0177926

[15]. Bokov P, Mahut B, Flaud P, Delclaux C. 2016. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Computers in Biology and Medicine*; 70 (2016): 40-50. doi:19 10.1016/j.compbiomed.2016.01.002.

[16]. Bahoura M. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. 2009. *Computers in Biology and Medicine*; 39 (2009): 824-843. doi: 10.1016/j.compbiomed.2009.06.011

[17]. Göğüş FZ, Karlık B, Harman G. 2016. Identification of pulmonary disorders by using different spectral analysis methods. *International Journal of Computational Intelligence Systems*; 9 (4): 595-611. doi:10.1080/18756891.2016.1204110

[18]. Kandaswamy A, Kumar CS, Ramanathan RP, Jayaraman S, Malmurugan N. 2004. Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*; 34 (2004): 523-537. doi: 10.1016/S0010-27 4825(03)00092-1

[19]. Oweis RJ, Abdulhay EW, Khayal A, Awad A. 2015. An alternative respiratory sounds classification system utilizing artificial neural networks. *Biomedical Journal*; 38 (2): 153-161. doi: 10.4103/2319-4170.137773.

[20]. Lozano M, Fiz JA, Jane R. 2016. Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency. *IEEE Journal Of Biomedical And Health Informatics*; 20 (2): 486-497. doi: 10.1109/JBHI.2015.2396636

[21]. Pasterkamp H, Kraman SS, Wodicka GR. 1997. State of the Art Respiratory Sounds Advances Beyond the Stethoscope. *American Journal of Respiratory and Critical Care Medicine*; 156(3): 974-987. doi: 10.1164/ajr-ccm.156.3.9701115.

[22]. Palaniappan R, Sundaraj K, Ahamed N, Arjunan A, Sundaraj S. 2013. Computer-based respiratory sound analysis: a systematic review. *IETE Technical Review*; 30 (3): 248-256. doi: 10.4103/0256-4602.113524

[23]. Dokur Z. 2009. Respiratory sound classification by using an incremental supervised neural network. *Pattern Analysis and Applications*; 12 (4): 309-319. doi: 10.1007/s10044-008-0125-y

[24]. Sovijarvi ARA, Vanderschoot J, Earis JE. 2000. Standardization of computerized respiratory sound analysis. *European Respiratory Review*; 10 (77): 585.

[25]. Ben Nouma B, Mitiche A, Ouakrim Y, Mezghani N. 2019. Pattern Classification by the Hotelling Statistic and Application to Knee Osteoarthritis Kinematic Signals. *Machine Learning and Knowledge Extraction*; 1 (3): 768-784. doi:10.3390/make1030045

[26]. Vannuccini L, Earis JE, Helistö P, Cheetham BMG, Rossi M et al. 2000. Capturing and preprocessing of respiratory sounds. *European Respiratory Review*; 10 (77): 616-620.

[27]. Huang NE, Shen Z, Long SR,Wu MC, Shih HH et al. 1998. The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. *Royal Society of London Proceedings Series A*; 454: 903-995. doi: 10.1098/rspa.1998.0193

[28]. İçer S, Gengeç Ş. 2014. Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digital Signal Processing*; 28 (2014): 18-27. doi: 10.1016/j.dsp.2014.02.001

[29]. Hu X, Peng S, Hwang WL. 2012. EMD revisited: A new understanding of the envelope and resolving the model mixing problem in AM-FM signals. *IEEE Transactions on Signal Processing*; 60(3): 1075-1086. doi:13 10.1109/tsp.2011.2179650

[30]. Deering R, Kaiser JF. The use of a masking signal to improve Empirical Mode Decomposition. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, USA, 2005. doi:10.1109/icassp.2005.1416051

[31]. Palaniappan R, Sundaraj K. Respiratory sound classification using cepstral features and support vector machine. 2013 IEEE Recent Advances in Intelligent Computational Systems, Trivandrum, India, 2013. doi:10.1109/raics.2013.6745460

[32]. Aykanat M, Kılıç Ö, Kurt B, Saryal S. 2017. Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*; 2017(1). doi: 10.1186/s13640-017-0213-2

[33]. Rioul O, Vetterli M. 1991. Wavelets and Signal Processing. *IEEE Signal Processing Magazine*; 8 (4): 14-38. doi: 10.1109/79.91217

[34]. Gengeç Ş. Akciğer Seslerinden İşaret İşleme Teknikleri Kullanılarak Özellik Çıkarma ve Sınıflandırma. Erciyes Üniversitesi, 2012. (in Turkish)

[35]. Subasi A. 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*; 32 (4): 1084-1093. doi:10.1016/j.eswa.2006.02.005

[36]. Reichert S, Gass R, Brandt C, Andres E. 2008. Analysis of Respiratory Sounds: State of the Art. *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine*; 2008 (2): 45-58. doi: 10.4137/ccrpm.s530

[37]. Vapnik VN. Statistical Learning Theory. New York: Wiley; 1998.

[38]. Başer F, Apaydın A. 2015. Comparison Between Logistic Regression and Support Vector Machines for Classification Purposes. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi B- Teorik Bilimler*; 3 (2): 53-65 (in Turkish with an abstract in English) doi: 10.20290/btdb.67263.

[39]. Naves R, Barbosa BHG, Ferreira DD. 2016. Classification of lung sounds using higher-order statistics: A divide-and-conquer approach. *Computer Methods and Programs in Biomedicine*; 129 (2016): 12-20. doi:36 10.1016/j.cmpb.2016.02.013

[40]. Bhatia N, Vandana. 2010. Survey of Nearest Neighbor Condensing Techniques. *International Journal of Computer Science and Information Security*; 8 (2): 302-305.

[41]. Tu J V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*; 49 (11):1225-1231. doi: 10.1016/s0895-4356(96)00002-9

[42]. Özmen Ö, Khdr A, Avcı E. 2018. Sınıflandırıcıların kalp hastalığı verileri üzerine performans karşılaştırması. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*; 30(3): 153–159. (in Turkish with an abstract in English) doi:10.13140/rg.2.2.24732.95365.