Medical & Biological Engineering & Computing (2022) 60:1555–1568
https://doi.org/10.1007/s11517-022-02552-w

ORIGINAL ARTICLE

Extraction of low-dimensional features for single-channel common lung sound classification

M. Alptekin Engin[1] · Selim Aras[2] · Ali Gangal[3]

Received: 9 November 2021 / Accepted: 9 March 2022 / Published online: 4 April 2022

Abstract
In this study, feature extraction methods used in the classification of single-channel lung sounds obtained by automatic identification of respiratory cycles were examined in detail in order to extract distinctive features at the lowest size. In this way, it will be possible to design a system for the detection of lung diseases, completely autonomously. In the study, automatic separation and classification of 400 respiratory cycles were performed from the single-channel common lung sounds obtained from 94 people. Leave one out cross validation (LOOCV) was used for the calibration and validation of the classification model. The Mel frequency cepstrum coefficients (MFCC), time domain features, frequency domain features, and linear predictive coding (LPC) were used for classification. The performance of the features was tested using linear discriminant analysis (LDA), *k*-nearest neighbors (*k*-NN), support vector machines (SVM), and naive Bayes (NB) classification algorithms. The success of combinations of features was explored and enhanced using the sequential forward selection (SFS). As a result, the best accuracy (90.14% in the training set and 90.63% in the test set) was acquired using the *k*-NN for the triple combination, which included the standard deviation of LPC and the standard deviation and the mean of MFCC.

Keywords Lung sounds · Respiratory cycle · Automatic recognition · Feature extraction · Classification · Sequential forward selection

1 Introduction

The lung sounds that emerge due to the rapid change of air pressure and the vibrations of tissues are evident in the subsequent vibration of the chest wall. However, it is difficult to hear the lung sounds because they are weakened when they reach the chest wall. Long ago, doctors tried to overcome this hearing problem by putting their ears on the patient's chest. A French doctor, Laennec, recognized that putting his ear on the patient's chest did not work in a fat person, and he invented the stethoscope about 200 years ago [1]. The classic stethoscope has undergone very little change since it was invented. In order to diagnose lung diseases, a physician must have good hearing ability, good training, and sufficient experience. Physicians' interpretation while simultaneously listening to lung sounds is not up to a standard. However, with the development of electronic stethoscopes that can enable a standard decision mechanism, it is possible to reduce ambient noise and enhance the lung sounds. These important features make electronic stethoscopes superior to classic stethoscopes for hearing lung sounds. In addition, lung sounds can be recorded using an electronic stethoscope and can be used in the training of medical students. Such recordings also can be used in the computerized analysis of lung sounds. Auscultation is the process of listening to lung sounds with a stethoscope. Doctors who have recently graduated from medical school are known to have difficulty in recognizing common lung sounds due to their lack of experience [2]. However, electronic stethoscopes have the potential to automatically recognize lung sounds in a standard way with the aid of computer analysis.

Studies on the recognition of lung sounds generally are considered to be single-channel or multi-channel in terms of recording methods. However, multi-channel recording methods generally were used in these studies [3–9]. Although

✉ M. Alptekin Engin
maengin@bayburt.edu.tr

[1] Department of Electrical and Electronics Engineering, Bayburt University, 69000 Bayburt, Turkey

[2] Department of Electrical and Electronics Engineering, Ondokuz Mayis University, 55200 Samsun, Turkey

[3] Department of Electrical and Electronics Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

multichannel recording systems contain more information about lung sounds, single-channel recording systems are more suited to lung auscultation because of placement multi-microphones on the chest wall which is very troublesome. This is due to the fact that placing the microphones of the multichannel recording system on the chest is very difficult and can take a long time. In addition, when using the multi-channel recording method, it is difficult to place the microphones on fat or hairy people. Thus, single-channel recording methods are more suitable for developing a real-time application, so we used single-channel lung sounds in this study.

The process of recognizing lung sounds consists of three steps, i.e., determining the respiratory cycles, extracting the features, and classifying the features. Respiratory cycles are commonly determined manually for single-channel-recorded lung sounds. In fact, the first study to automatically determine the limits of breathing cycles for single-channel lung sounds was our previous work [2]. In multichannel recordings, non-auscultatory signals, such as trachea sounds [10], spirometer (airflow) signals [11], nasal breath sounds [12], and photoplethysmogram signals [13], are used to determine the breath cycles in the lung sounds simultaneously through another signal.

Various methods have been used to classify the features obtained from lung sound signals, e.g., *k*-nearest neighbors (*k*-NN), artificial neural networks (ANNs) [14], Gaussian mixture models (GMM) [15], and support vector machines (SVM) [16]. Yilmaz and Kahya used a dataset consisting of 27 healthy people and 21 pathological people to develop healthy and pathological classifications [3]. They divided the respiratory phase into inspiration and expiration. The maximum accuracies obtained in the expiration and inspiration phases of the respiratory cycle using the coefficients of the autoregressive (AR) model were 77.8% and 68.9%, respectively. Many studies on computerized analysis of lung sounds have been able to distinguish between healthy and pathological lung sounds. Classification of healthy and unhealthy individuals will not completely solve the problem, since lung sounds are of different types and each type addresses a different health problem. However, in order to develop a useful method, at least the classification of common lung sounds is needed. There have been relatively few studies related to the classification of multiple lung sounds. Rajkumar Palaniappan et al. [17] obtained records from the trachea, the left and right posterior points of the lung, using a digital stethoscope. They used a dataset with 120 records, i.e., 40 normal, 20 wheeze, 20 rhonchi, 20 fine crackle, and 20 coarse crackles. They used the attributes of Mel frequency cepstrum coefficients (MFCC); AR and the SVM classifier was used to classify the lung sounds. They achieved a classification accuracies of 88.72% for the AR coefficients and 89.68% for the MFCC coefficients. When past studies are examined, it is seen that MFCC is used as one of the most common methods to extract features from lung sounds. Commonly used methods other than MFCC are fast Fourier transform (FFT) [18], wavelet transform (WT) [19], and linear predictive coding (LPC) [20]. Daniel Chamberlain et al. [21] recorded wheeze and crackle sounds in 284 patients using an electronic stethoscope and a smartphone. Their algorithm achieved receiver operating characteristic (ROC) curves with area under the curves (AUCs) of 0.86 for wheeze and 0.74 for crackle using the semi-supervised, deep-learning method. Rizwana Zulfiqar et al. [22] applied Fourier analysis with artificial noise addition (ANA) to visually inspect abnormal breath sounds for deep convolutional neural networks (CNN) method. They achieved a classification accuracy of 100% through the AlexNet algorithm. In another study based on deep learning, Yoonjoo Kim et al. [23] detected abnormal respiratory sounds with 86.5% accuracy and area under the ROC curve (AUC) of 0.93. They measured overall respiratory sound classification accuracy as 60.3% for different groups, including medical students, interns, residents, and fellows. Nasreddine Belkacem et al. [24] proposed a theoretical end-to-end point-of-care system supported by artificial intelligence (AI) to classify and diagnose different respiratory diseases.

Among the numerous studies on lung sounds, the main purpose of this study is to show that it is possible to design a completely automatic device that can be used in the diagnosis of disease from single-channel lung sounds, which only needs the approval of doctors. Thus, it will be possible to shorten the diagnosis and treatment time of lung diseases. Therefore, we propose a fully automatic classification method for lung sounds recorded as single-channel without manual assistance for the first time in the literature. Initially, respiration cycles of all lung sounds in the database used were automatically detected. Then, 11 different feature extraction methods were used in the classification phase. Then, the most suitable combination of features was investigated to classify lung sounds at high performance using low-dimensional features via the sequential forward selection (SFS) method. In order to obtain the classification performance, we used the well-known and extensively used support vector machines (sVM), *k*-nearest neighbors (*k*-NN), linear discriminant analysis (LDA), and naive Bayes (NB) algorithms. As a result, the best accuracy value (90.14% in the training set and 90.63% in the test set) among the features used was acquired for the triple combination, which included the standard deviation of LPC and the standard deviation and the mean of MFCC.

2 Materials and methods

2.1 Data set description

The dataset is the most important factor because it can have a direct effect on the success of classification studies.

Gurung [25] criticized the datasets used in the studies involving the computer analysis of lung sounds and emphasized that the lack of standardization in the studies was the biggest obstacle in the development of a commercial product. In addition, databases that are widely used in previous studies were created from the lung sounds obtained by using the multi-channel recording method. However, microphones used in multi-channel recording systems are not practical due to the difficulty of placing them on the chest. In addition, since the databases used in many previous studies contain limited data, data augmentation has been applied to improve classification performance. Operations such as adding noise to the background of the sound in the data augmentation process cause mixing of normal lung sounds with other lung sounds. For these reasons, it is necessary to establish a suitable database for detailed feature research in the classification of the sounds.

In this study, the standards for recording were set in advance and a document of authorization for the study was obtained from the Ethics Committee of Karadeniz Technical University (KTU). Working with two physicians who were specialists in this field, the records were provided in accordance with the auscultation procedure. The records were labeled by physicians in the Department of Chest Diseases in the School of Medicine at KTU. We used a digital sound recorder and an electronic stethoscope (Thinklabs ds32a+) with a single-channel analog output to obtain the recordings. The electronic stethoscope was used in the diaphragm mode, which was suitable for the frequency range of the lung sounds (20–2000 Hz). To create a large dataset, we recorded lung sounds from 94 different people. The number of breath cycles we obtained in 94 people and the details of the data are shown in Table 1. Each record contains between three and seven respiratory cycles. Commonly heard lung sounds are defined as normal, rhonchi, fine crackles, coarse crackles, and wheeze [1]. However, the wheeze sound was ignored in the study because it can be heard in healthy individuals with forced expiration after a deep breath and maybe loud enough to be heard without a stethoscope [2].

Table 1 Recorded lung sounds and automatically determined breath cycles

Sound	Number of people	Number of cycles
Normal	30	100
Rhonchi	23	100
Fine crackle	20	100
Coarse crackle	21	100
Total database	94	400

2.2 Automatic recognition of respiratory cycles

The respiratory cycle is the sequence of events during which a human inhales and exhales a given volume of air through the respiratory system [1]. The main purpose at this stage is to determine the respiratory cycles with different characteristics for each individual, regardless of the duration. The automatic recognition of respiratory cycles begins with the spectrogram calculation. Figure 1 shows the flowchart of the process. In this method, the lung sounds are divided into frames of 10 ms that overlap each other. After applying the Hamming window to the frames, the N-point FFT is calculated. After the spectrogram is obtained, a bandpass filter is applied in the range of 80–1000 Hz to reduce the effect of noise in order to concentrate on the lung sounds.

The next step is to calculate the density of the spectral energy and apply a smooth filter to obtain smoother patterns. As a result of the initial processing, the breath cycles appear as repetitive, similar patterns, as shown in Fig. 2. The samples from these patterns were compared in the DTW algorithm and the boundary points were determined based on the similarity of the patterns. Then, these boundary points were applied to the original lung sound to determine respiratory cycles. Physicians have confirmed the suitability of all

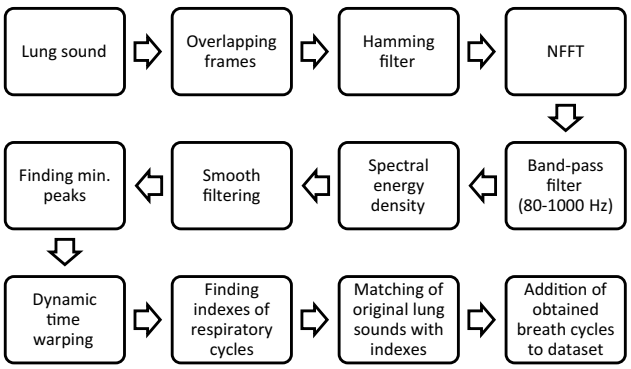

Fig. 1 Flowchart of automatic recognition of respiratory cycles from lung sounds

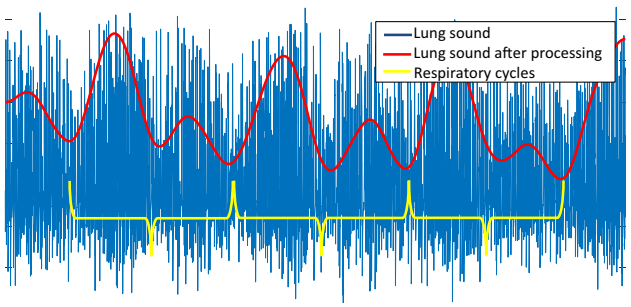

Fig. 2 Respiratory cycles and their patterns after processing

respiratory cycles that we have automatically identified and included in the dataset. Our previous work [2] provides more information about the process of automatically determining respiratory cycles.

2.3 Feature extraction

In this study, the most appropriate method for extracting the features of common, single-channel lung sounds was researched extensively using the Mel frequency cepstrum coefficients (MFCC), frequency domain features (FDF), time domain features (TDF), and linear predictive coding (LPC) features. The most suitable combination of features also was investigated using the sequential forward selection (SFS) method, which is described later in the classification section. Figure 3 shows an overview of how to obtain the features we used in this study. Before calculating all of the features, the signals from the respiration cycle were divided into overlapping frames and the features of each frame were calculated. In this study, we used the frames that overlapped each other as 16 ms As a result, a feature matrix with the size of *number of features* × *number of frames* were obtained for a respiratory cycle. However, the cycles in the dataset had different durations that depended on the person's rate of breathing and pathological findings. Hence, the number of frames obtained for each respiratory cycle changes, so the lengths of the feature matrix also change. However, to be able to classify lung sounds, the number of the features obtained from all of the respiratory cycles must be equal. To achieve this, we obtained equal-length feature vectors for all breath cycles by separately computing the various parameters for each column of the feature matrix, such as mean, standard deviation, Hjorth parameters (activity, mobility, complexity), kurtosis, and skewness. Thus, 11 features were derived from the calculated parameters.

2.3.1 Mel frequency cepstrum coefficients (MFCC)

In the Introduction, we discussed that auscultation is related to the physician's hearing ability. Since MFCC is a method developed based on the human auditory system, it was used as a feature extraction method in this study. Before beginning to calculate the features, such as MFCC, the respiratory cycle was divided into overlapping short frames, which were considered to have balanced acoustic characteristics of the signal. The power spectrum was obtained by fast Fourier transform (FFT) by applying Hamming window function to the frames to minimize discontinuities in the signals at the beginning and end of each frame. This power spectrum was passed through a Mel filter bank that consisted of 20 triangular filters. Thus, the spectrum was arranged according to the Mel scale, which imitates the hearing of the human ear. Then, the logarithm is calculated to reduce the sensitivity of the feature vectors. In the last stage, we applied the discrete cosine transform, and MFCC coefficients were calculated for each frame, resulting in a total of 12 coefficients. Thus, the MFCC coefficients obtained from a frame formed a row of the feature matrix. The number of columns is the same as the number of MFCC coefficients. The mean (*f1*), standard deviation (*f2*), Hjorth parameters (*f3*), kurtosis (*f4*), and skewness (*f5*) of the parameters of each column of the MFCC feature matrix were calculated, and five different feature vectors were obtained. The formulas for these five parameters and the other parameters we used in obtaining the next features are given in Table 2.

The arithmetic mean obtained by dividing the sum of the elements of a sequence by the number of elements is given by Eq. 1. Standard deviation (Eq. 2) describes how the data are spread according to the arithmetic mean. Hjorth parameters [26] were used predominantly in the electroencephalography (EEG) signals, which consist of mobility (Eq. 3), activity (Eq. 4), and complexity (Eq. 5). Kurtosis (Eq. 6) indicates the apex or flatness of the distribution, i.e., it indicates how far the height of the series is from the height of the normal series. Skewness (Eq. 7) is a parameter that shows how the samples in the series are scattered and how far away they are symmetrical.

2.3.2 Time domain features (TDF)

The most important advantage of TDF is that it can be calculated quickly without requiring any mathematical transformation. However, since they are calculated using the amplitude of the signal, they can be sensitive to noise and other disturbing effects. In this study, zero crossing rate (ZCR), energy, entropy of energy, activity, mobility, and complexity parameters were calculated in the time domain. ZCR is the rate at which the signal changes from positive to negative or negative to positive. The term $x_i(n)$ represents examples of the ith frame of the audio signal. ZCR is calculated as shown in Eq. 8, where N is the length of the frame, and *sgn* is the signum function. The short-time energy, which is one of the basic parameters in the processing of audio signals, is calculated as shown in Eq. 9 for the ith frame of the audio signal. To calculate the entropy of the energy that determines the irregularity of the time-energy content of a signal, the selected x_i frame was divided into M subframes of equal length. Then, the energy of each subframe is divided by the total energy of the frame, x_i, after it is calculated using Eq. 9. If the energy of the subframes $ej, j = 1, \ldots, M$, then the entropy of the energy (H_i) is calculated using Eq. 10. Hjorth parameters in the time domain were calculated using Eqs. 3, 4, and 5. Thus, six parameters in the time domain for each frame were calculated, and these six parameters were used

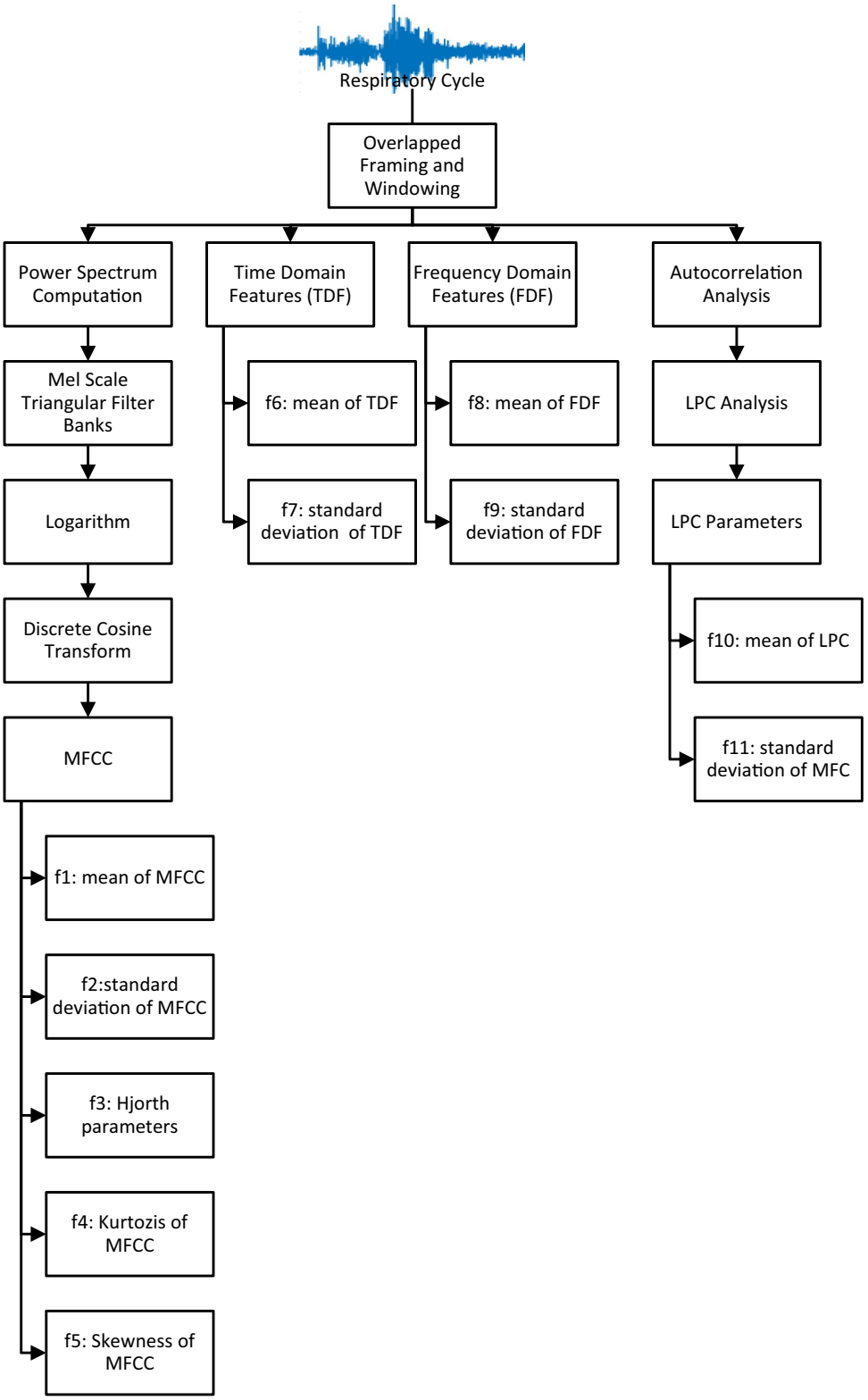

Fig. 3 Overview of how we obtained the features used in the study

Table 2 Parameters and their formulas we used to derive the features

Mean	(1)	Kurtosis	(6)	Spectral Centroid	(11)
$\overline{\mathrm{x}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{N} \mathrm{x_i}$		$K = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{X_i - \overline{X}}{\sigma}\right]^4$		$C_i = \frac{\sum_{k=1}^{N/2} kX_i(k)}{\sum_{k=1}^{N/2} X_i(k)}$	
Standard Deviation	(2)	Skewness	(7)	Spread of a Frame	(12)
$\sigma = \sqrt{\frac{1}{\mathrm{N}-1} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{x_i} - \overline{\mathrm{x}})^2}$		$S = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{X_i - \overline{X}}{\sigma}\right]^3$		$S_i = \sqrt{\frac{\sum_{k=1}^{N/2} (k - C_i)^2 X_i(k)}{\sum_{k=1}^{N/2} X_i(k)}}$	
Activity	(3)	Zero-Crossing Rate (ZCR)	(8)	Spectral Entropy	(13)
$\mathrm{Variance(X(n))} = \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\mathrm{X_n} - \overline{\mathrm{X}}\right)^2$		$\frac{1}{2N} \sum_{n=1}^{N} \left\lvert sgn\left[x_i(n)\right] - sgn\left[x_i(n-1)\right]\right\rvert$		$H = -\sum_{f=0}^{L-1} n_f . log_2(n_f)$	
Mobility	(4)	Energy	(9)	Spectral Flux	(14)
$Mobility = \sqrt{\frac{Variance\left(\frac{dX(n)}{dn}\right)}{Variance(X(n))}}$		$E(i) = \sum_{n=1}^{N} \left[x_i(n)\right]^2$		$SF_{(i,i-1)} = \sum_{k=1}^{\frac{N}{2}} \left(X_i(k) - X_{i-1}(k)\right)^2$	
Complexity	(5)	Entropy of Energy	(10)	Spectral Roll-off	(15)
$Complexity = \frac{mobility\left(\frac{dX(n)}{dn}\right)}{mobility(X(n))}$		$H_i = -\sum_{j=1}^{M} e_j . log_2(e_j)$		$\sum_{k=1}^{m} X_i(k) = 0.85 \sum_{k=1}^{N/2} X_i(k)$	

in each frame of the breath cycle to calculate the mean (*f6*) and standard deviation (*f7*) of the feature matrix.

2.3.3 Frequency domain features (FDF)

In the frequency domain, we calculated five feature matrices for each frame, i.e., spectral center, spectral spread, spectral entropy, spectral flux, and spectral roll-off. The mean (*f8*) and standard deviation (*f9*) of the feature matrices that were obtained were used as the features of the frequency domain. The spectral center was obtained by determining the center of gravity of the spectrum obtained after the Fourier transformation. For the calculation of the spectral properties, it is sufficient to work with the first half of the frequency spectrum and the spectral center (C_i) for the ith frame that is calculated using Eq. 11. The spectral spread gives information about how the spectrum is distributed around the center (Eq. 12). The filter for which spectral entropy is to be calculated is first divided into L subbands. Next, the normalizing energy (*nf*) of each fth subband (*f=0, …, L-1*) is found with the help of total spectral energy. Equation 13 was used to calculate the entropy of *nf*. Equation 14 is used to calculate the spectral flux, which is the spectral change between two consecutive frames. Here, $X_i(k)$ is expressed as the normalized coefficient of the discrete Fourier transform (DFT) in the ith frame. The frequency value that corresponds to a predetermined percentage value of the sum of the magnitudes in the frequency spectrum is called the spectral roll-off. The percentage value was chosen as 85% in this study. If we consider that the mth DFT coefficient corresponds to the spectral decay of the ith frame, the spectral roll-off can be calculated using Eq. 15.

2.3.4 Linear predictive coding (LPC)

LPC is one of the most popular methods used for sound recognition and encoding. The same procedure is used for framing and windowing in this method as well as for other methods of extracting features. The LPC coefficients were calculated by applying autocorrelation and LPC analysis. The LPC analysis is based on the idea that the instantaneous sound sample can be represented by a weighted linear combination of past *p* samples [27, 28]. The aim here was to calculate the predictive coefficient values, a_k, that minimize the square error function.

$$\mathrm{E_n} = \sum_{\mathrm{m}=-\infty}^{\infty} \left(\mathrm{s_n[m]} - \sum_{\mathrm{k}=1}^{\mathrm{p}} \mathrm{a_k s_n[m-k]}\right)^2 \tag{16}$$

The derivative of Eq. 16 is equated to zero ($\partial E / \partial a_k = 0$). After solving this equation, the generalized normal equations are obtained. Covariance analysis can be used instead of autocorrelation analysis to switch to LPC analysis, but we used an extensively used autocorrelation analysis instead. In the autocorrelation method, the signal is assumed to be zero except for the analyzed frame, and the infinite total expression is limited. In this case, the generalized normal equations can be written in matrix form (Eq. 17).

$$\begin{bmatrix} \mathrm{r(1)} \\ \mathrm{r(2)} \\ \cdots \\ \mathrm{r(p)} \end{bmatrix} = \begin{bmatrix} \mathrm{r(0)} & \mathrm{r(1)} & \cdots & \mathrm{r(p-1)} \\ \mathrm{r(1)} & \mathrm{r(0)} & \cdots & \mathrm{r(p-2)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{r(p-1)} & \mathrm{r(p-2)} & \cdots & \mathrm{r(0)} \end{bmatrix} \begin{bmatrix} \mathrm{a_1} \\ \mathrm{a_2} \\ \vdots \\ \mathrm{a_p} \end{bmatrix} \tag{17}$$

The predictive coefficients (a_k) are obtained from Eq. 17 using the recursive Levinson-Durbin algorithm. The mean (*f10*) and standard deviation (*f11*) of the feature matrix, which was obtained by calculating 12 LPC coefficients in each frame, were used as features.

2.4 Classification algorithms

Figure 4 shows the classification protocol used in this study. The input is a dataset that consisted of 400 labeled breath cycles. The output was the average of the performance of training and validation where this dataset is selected 100 times evenly and randomly. Repeating the dataset 100 times and taking the average of the performances allows the method to provide more reliable results. After the attributes were calculated in the training phase of classification, the leave one out cross validation (LOOCV) method was used to minimize the sample dependency. The total number of trials in the LOOCV technique was the number of samples in the training set. In each trial, one sample was left out, and the training phase was performed with the remaining samples.

In order to obtain the classification performance, we used the well-known and extensively used support vector machines (SVM), *k*-nearest neighbors (*k*-NN), linear discriminant analysis (LDA), and naive Bayes (NB) algorithms. To improve the classification performance, the success of feature combinations was investigated using the SFS method. Figure 5 shows the flow diagram of the SFS method.

Since the SVM, *k*-NN, LDA, and NB algorithms are well-known and have been used extensively, some features have been provided rather than detailed descriptions. LDA was first used in pattern recognition by Belhumeur [29]. The LDA algorithm constructs a probability density function model for each class feature. Then, the test data is classified by determining the maximum probability density function.

The *k*-NN method is a classical, simple, and supervised learning algorithm that does not require the probability distribution of classes. In this method, the test vector contains the *k*-neighbors closest to the current learning data. If the majority of these neighbors belong to a given class, that class is assigned as a result of recognition [30]. In our study, the Euclidean distance function was used to determine the best value of *k*. The *k*-value was searched in the range between 1 and the dimension of the feature vector using a step size of 1.

SVM, a method based on statistical learning theory, was developed by V. Vapnik et al. [31]. SVM performs the classification operation by determining the best separating plane for linearly decomposable data structures. The difference from other classifiers is that they try to find solutions that will reduce the probability of incorrect classification. In this study, the Gauss radial basic function was used as the seed function. Also, because the number of classes was more than two, the one-vs-one method was used.

In the naive Bayes classification method, all of the features that form the feature vector are considered to be statistically independent [32]. Briefly, this classifier computes the conditional probabilities of different classes given the values of the features, and, then, it selects the highest conditional probability

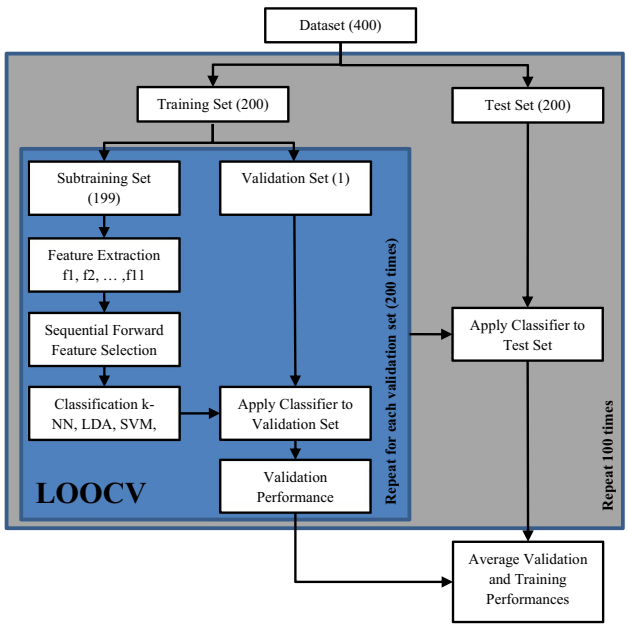

Fig. 4 Classification process

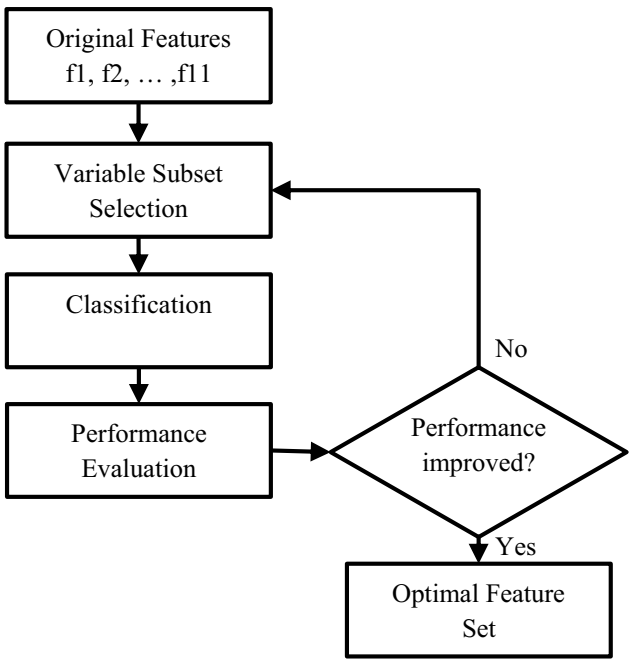

Fig. 5 Flowchart of the sequential forward selection (SFS) method

class. Despite the lean design and seemingly simplified assumptions, the naive Bayes classifier only requires a small amount of training data to estimate the classification parameters. So, in situations where the dataset is inadequate, it gives much better results than expected in real-world situations.

3 Results

In this study, we used common lung sounds recorded as a single channel from 94 people. Respiratory cycles in the lungs were determined automatically and a large dataset was created with 400 breath cycles. The two physicians with whom we worked verified that all of the respiratory cycles we determined were appropriate. In this study, we investigated 11 different feature extraction methods using MFCC, TDF, FDF, and LPC. We used LDA, *k*-NN, SVM, and NB algorithms for testing these features. In order to avoid the increase or loss of success caused by the random selection problem, the random selection of training and test clusters was repeated 100 times. We used the SFS method to investigate whether we could increase the success rate in the classification accuracy. In the SFS method, the successes of the attributes obtained from the training set are evaluated, and the success of the best combination of attributes is investigated as long as success is achieved. In this study, success increased up to triple combinations of features but since the performance fell in quadruple combinations of features, the calculations were terminated. Figure 6 shows the results we obtained.

When attributes were evaluated one at a time (Fig. 6a), the best performance in the training set was $88.02 \pm 2.16\%$ in the case of classifying the *f1* attribute with *k*-NN ($k = 1$). According to the SFS method, the success of the binary combination of the *f1* feature with the other features was investigated in the second step. Figure 6b shows that the best combination in binary form was $f1 + f2$, with $89.56 \pm 1.80\%$. This success was better than the result of $88.02 \pm 2.16\%$ we obtained earlier,

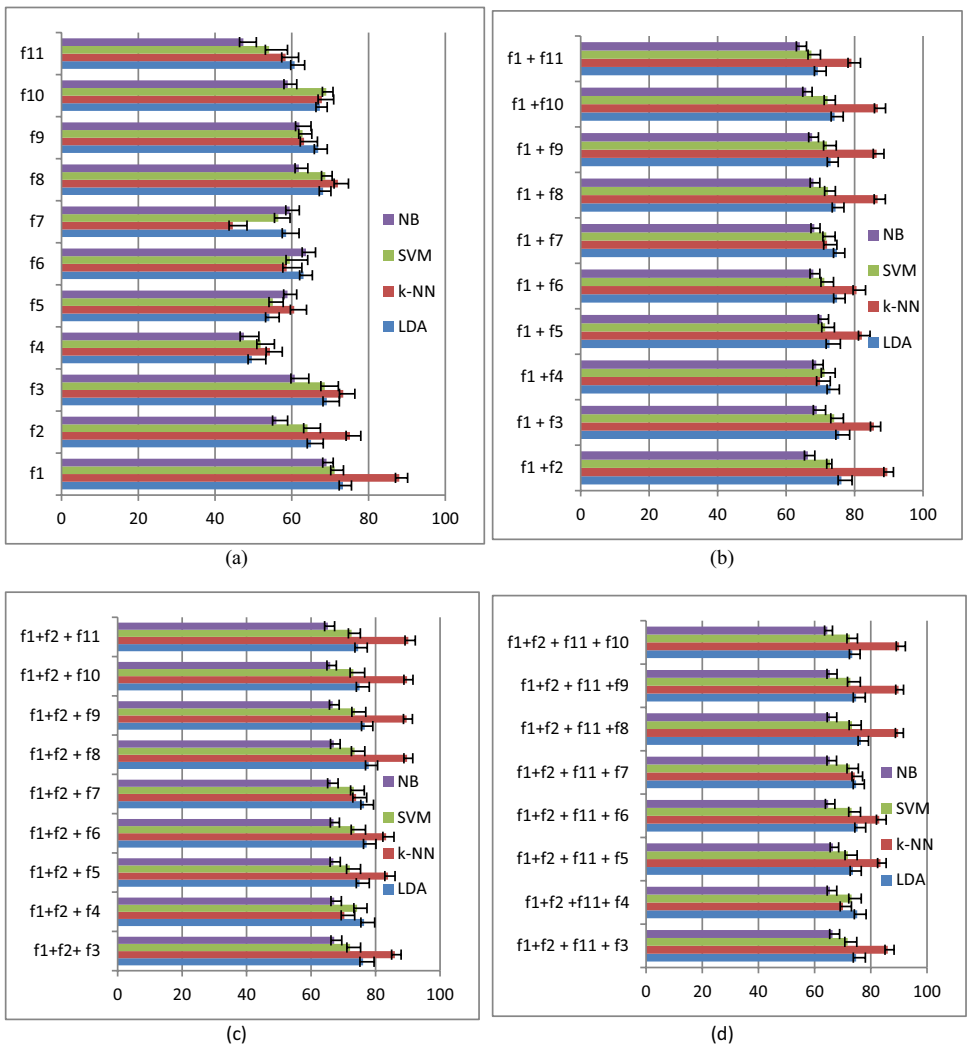

Fig. 6 Training accuracy results of the proposed method: **a** initial results for features; **b** binary combinations of the best feature and other features; **c** triple combinations of the best two features and other features; **d** quad combinations of the best three features and the other features

so we investigated the combination of triple features (Fig. 6c) where *f1* and *f2* were fixed. When investigating the success of triple-feature combinations, we obtained the best result achieved in this study. Accordingly, the best combination of features belonged to the *f1* + *f2* + *f11* community, which produced the result of 90.14 ± 2.18%. In the next stage (Fig. 6d), the highest performance was obtained with the combination of *f1* + *f2* + *f11* + *f10*, i.e., 90.06 ± 2.23%, but since this result was below the success of the previous combination, i.e., *f1* + *f2* + *f11*, feature selection according to the SFS method was completed. Thus, the SFS method allowed us to increase the success rate from 88.02 ± 2.16% to 90.14 ± 2.18%. This increase reflects the importance of the SFS method in feature selection. In addition to these results, the accuracy of the test results of the best feature combination (*f1* + *f2* + *f11*) is shown in Table 3.

When Table 3 is examined, it is seen that the *k*-NN classification algorithm is generally more successful than other classification algorithms. For the purpose of visually investigating the cause of this, a three-dimensional graph of the features of *f1*, *f2*, and *f11* is shown in Fig. 7. It is also visually presented how this combination of triple features separates the four classes. Figure 7 shows that the features obtained from the normal respiratory cycles were easily separated from the other features obtained from the pathological cycles. Essentially, this distribution shows that our proposed method also has the potential to determine the difference between pathological and healthy lung sounds even though the features of the pathologic classes of fine crackle, coarse crackle, and rhonchi were very close to each other. But, using the well-known *k*-NN classifier for classes with such a distribution, the highest performance was obtained for the four-class problem. Therefore, another good attribute of the proposed method was using *k*-NN, because it determined the nearest *k*-neighbors in the training set for the test data and determined the class with the highest rate as the test class.

Table 3 Accuracy of the test results of the proposed method

Features	LDA (%)	*k*-NN (%)	SVM (%)	NB (%)
f1 + *f2* + *f3*	75.59 ± 2.39	86.49 ± 2.43	72.90 ± 2.25	67.99 ± 2.43
f1 + *f2* + *f4*	76.98 ± 2.40	70.35 ± 2.62	74.40 ± 3.17	67.16 ± 2.33
f1 + *f2* + *f5*	75.26 ± 2.48	84.20 ± 2.46	71.80 ± 2.60	67.74 ± 2.49
f1 + *f2* + *f6*	76.42 ± 2.43	83.53 ± 2.65	73.72 ± 2.27	66.80 ± 2.47
f1 + *f2* + *f7*	76.58 ± 3.02	74.54 ± 2.38	73.95 ± 2.80	67.03 ± 2.33
f1 + *f2* + *f8*	77.30 ± 2.38	90.28 ± 2.12	73.69 ± 2.54	66.82 ± 2.35
f1 + *f2* + *f9*	76.06 ± 2.62	90.03 ± 2.30	73.27 ± 2.51	66.16 ± 2.48
f1 + *f2* + *f10*	75.62 ± 2.65	90.43 ± 2.10	73.26 ± 2.65	65.83 ± 2.25
f1 + *f2* + *f11*	75.05 ± 2.52	90.63 ± 2.23	73.00 ± 2.98	65.17 ± 2.90

The classification results of test data with features *f1* + *f2* + *f11* and *k*-NN classifier are provided as confusion matrix in Table 4. The confusion matrix gives detailed information about misclassifications. In a perfectly performing confusion matrix, all data are counted on the leading diagonal. When Table 4 is examined, normal sounds were classified perfectly. Rhonchi sounds had %3 misclassifications, fine crackle sounds had %15 misclassifications which were %10 classified as coarse crackle and coarse crackle had %19 misclassifications which were %11 classified as fine crackle.

The receiver operating characteristic (ROC) curves are shown in Fig. 8. The ROC curves show the true positive rate versus the false positive rate for our best feature combination case. A perfect result without misclassified points is a right angle to the upper left corner of the graph. The area under the curve (AUC) value is a measure of the overall quality of the classifier. Larger AUC values indicate better classification performance.

4 Conclusion and discussion

In this study, a first-time method was proposed for the fully automatic classification of single-channel recordings of common lung sounds. In the first part of the method, the lung sounds go through a process that we call automatic recognition of respiratory cycles. At the end of the initial process, respiratory cycles including the spectrogram are obtained as repetitive patterns. Using the dynamic time warping (DTW) algorithm, the boundaries of the recurrent respiratory cycles are determined and separated from the sound signal. Thus, a large dataset was created from 400 respiratory cycles containing 100 healthy, 100 fine rales, 100 coarse rales, and 100 rhonchi. To verify the automatic recognition of respiratory cycles, the lower and upper boundary points of the cycles indicated by the red dots in Fig. 9 were obtained by the automatic method. The boundaries

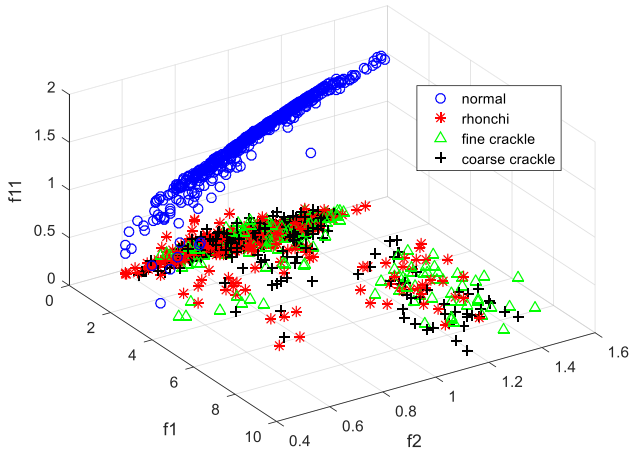

Fig. 7 3D graph of the combination of *f1* + *f2* + *f11*

Table 4 Confusion matrix

Confusion matrix		Predicted class				True positive rate	False negative rate
		Normal	Rhonchi	Fine crackle	Coarse crackle		
True class	Normal	100.0%	0.0%	0.0%	0.0%	100.0%	0.0%
	Rhonchi	0.0%	97.0%	2.0%	1.0%	97.0%	3.0%
	Fine crackle	1.0%	4.0%	85.0%	10.0%	85.0%	15.0%
	Coarse crackle	1.0%	7.0%	11.0%	81.0%	81.0%	19.0%

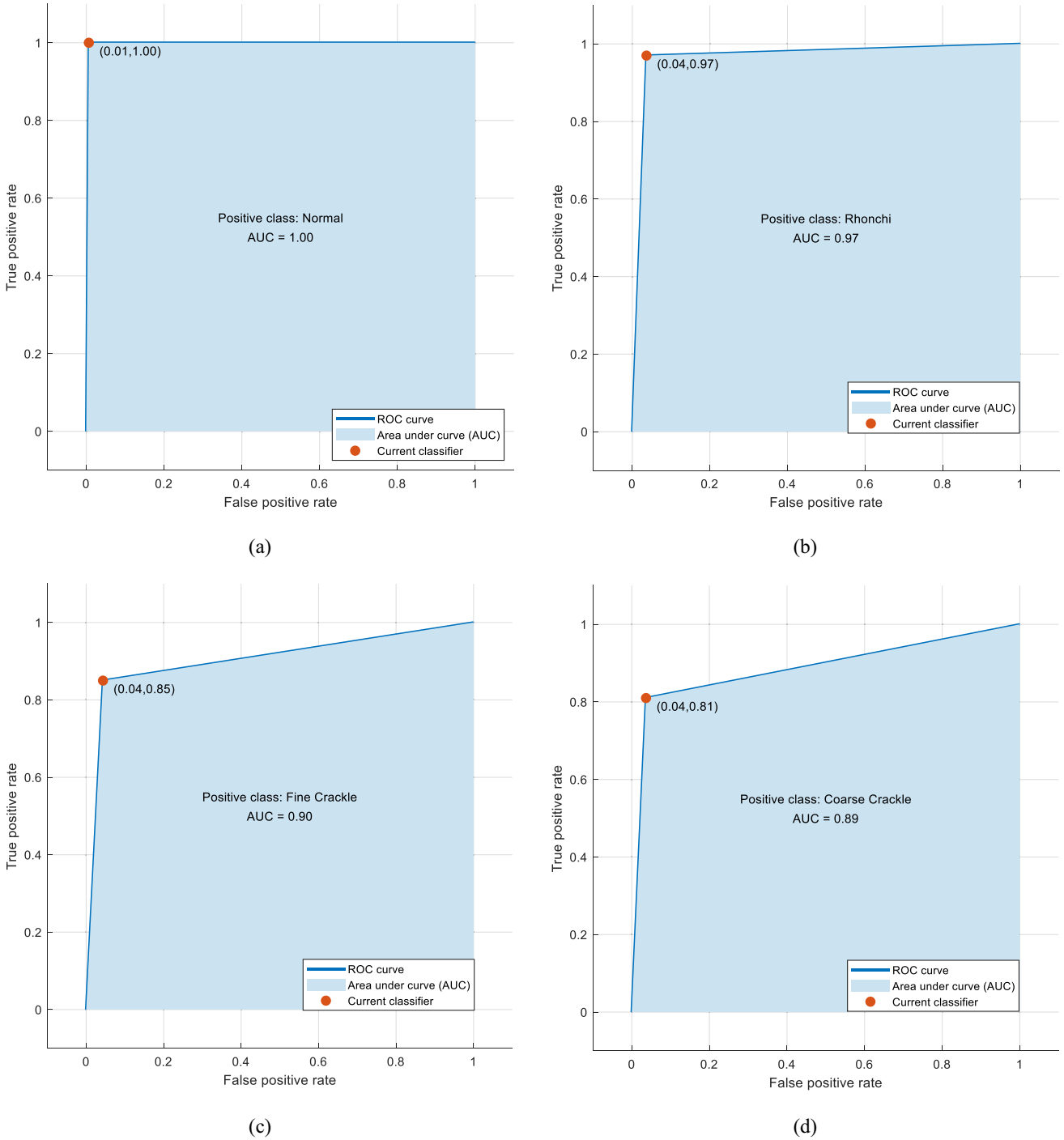

Fig. 8 ROC of the model for classifying lung sounds: **a** normal; **b** rhonchi; **c** fine crackle; **d** coarse Crackle, the mean AUC of 0.94

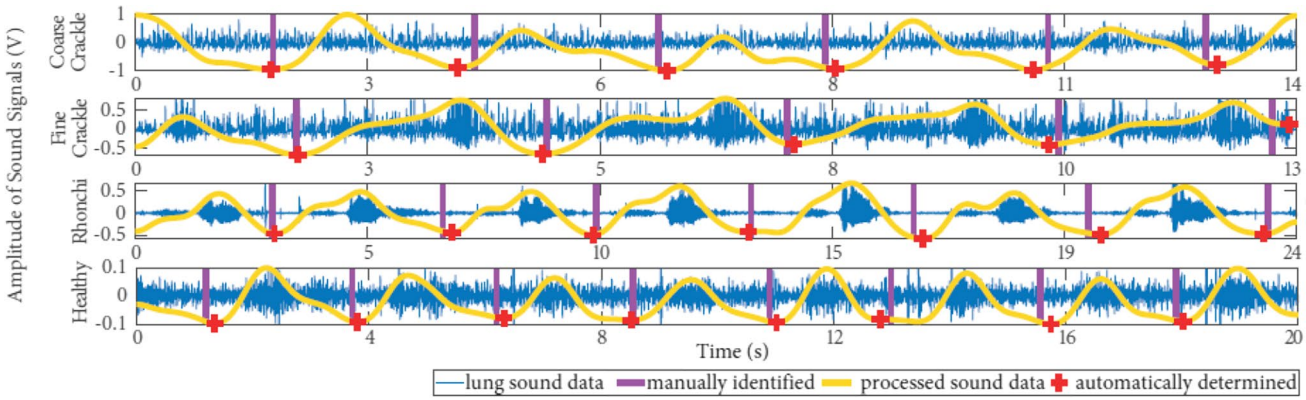

Fig. 9 Visual comparison of automatic and manual detection of the respiratory cycle

determined by the doctors are shown in magenta. When the results are examined, the automatic recognition method used and the determinations of the physicians are very close to each other. In addition, the boundaries determined by the physicians with the automatic determination method used were statistically compared in Table 5. Examining Table 5, the mean absolute error of the durations is 4.8% of the mean duration of all respiratory cycles. This value is quite low.

The standard deviation of absolute error rates can be considered high for mean values. However, since the results of the automated method are evaluated according to the boundaries set by the physicians, it is reasonable for the standard deviations to be higher than expected. Physicians' boundaries take into account that the time between respiratory cycles is short for some types of lung sounds and long for other types of sounds. Thus, small shifts in the boundary points determined by the physicians cause an increase in the standard deviation.

The problem of obtaining low-dimensional distinctive features, which is generally encountered in classification problems, is also important in the classification of common lung sounds. Obtaining these low-dimensional distinctive features makes it possible to perform classification with high performance and running fast. In the study, since lung sounds obtained with the single-channel recording method contain less information than multi-channel lung sounds, a detailed analysis was made on feature extraction. In the second part of the study, a comprehensive feature extraction consisting of Mel frequency cepstrum coefficients (MFCC), frequency domain features (FDF), time domain features (TDF), and linear predictive coding (LPC) was performed to obtain the appropriate features of the data set. In order to increase the reliability of this study, training and tests were repeated 100 times in classification. Also, leave one out cross-validation (LOOCV) method was used in classification. In addition, the sequential forward selection (SFS) method was used and the success rate was increased by investigating appropriate combinations of the features. Obtained features were tested using *k*-NN, SVM, LDA, and naive Bayes classification methods. Since this whole procedure was repeated 100 times for randomly determining training and test clusters, we presented the means of the results. The results indicated that the *k*-NN classifier was quite successful compared to other classification algorithms. However, the results for the combination of attributes were very close to each other. As a result, the best accuracy value (90.14% in the training set and 90.63% in the test set) among the features used was acquired for the triple combination, which included the standard deviation of LPC, the standard deviation, and the mean of MFCC.

Some of the disadvantages in comparing lung sound studies are that no standard method is followed for recording sounds and different preprocessings are applied to sounds to be classified. In Table 6, details about some studies in the literature and a comparison of the proposed study are given.

Table 5 Statistical comparison of automatic and manual detection of the respiratory cycle

Sound type	Mean of respiratory cycles (ms)	Standard deviation of respiratory cycles (ms)	Mean absolute errors (ms)	Standard deviation of absolute errors (ms)
Normal	2919.86	475.63	150.97	56.79
Rhonchi	2608.91	405.86	120.64	85.20
Fine crackle	2427.45	205.95	100.53	93.87
Coarse crackle	2258.19	380.49	115.95	83.58
Mean of results	2553.60	366.98	122.02	79.86

Table 6 Results of various lung sound classification studies reference

Reference	Data type	Dataset creation methods	Classification methods	Results
[3] Yılmaz and Kahya	27 healthy 21 pathological Multi-channel	Manuel	Autoregressive (AR) *k*-NN classifier	Accuracy; expiration 77.8% inspiration 68.9%
[33] Şen et al	20 healthy 20 pathological Multi-channel	Manuel	AR, Gaussian mixture model (GMM), SVM classifier	Accuracy; 90.0%
[17] Rajkumar Palaniappan et al	40 normal, 20 wheeze, 20 rhonchi, 20 fine crackles 20 coarse crackles Single-channel	Manuel	Mel frequency cepstrum coefficients (MFCC), AR and the SVM classifier	Accuracy; 88.72% AR 89.68% MFCC
[21] Chamberlain et al	171 normal 33 wheeze 19 crackle 4 wheeze and crackle Single-channel	Manuel	Semi-supervised deep learning	AUC; 86% Wheeze 74% Crackle
[23] Yoonjoo Kim et al	1222 normal 696 abnormal Single-channel	Manuel	Deep learning–based algorithm	Accuracy; 86.5% AUC; 93.0%
[23] Yoonjoo Kim et al	297 crackles 298 wheezes 101 rhonchi Single-channel	Manuel	Deep learning–based algorithm	Accuracy; 85.7% AUC; 92.0%
Proposed	100 normal 100 rhonchi 100 fine crackle 100 coarse crackle Single-channel	Automatic	LPC MFCC *k*-NN classifier	Accuracy; 90.63% AUC; 94.0%

When we compare our study with other studies, we can see that the proposed method has high performance and balance in terms of accuracy and AUC values as a result of the availability of sufficient training and test data in classification. The proposed study also indicated that recognition studies of single-channel lung sounds may be as successful as using multi-channel lung sounds. The method we proposed is the first step towards the development of a fully-automatic approach for recognizing single-channel lung sounds. We believe that the method we proposed will provide a useful basis for additional studies in this area. Another good attribute of the proposed method is that our study has the potential to transform a single-channel lung sound into a real-time application. The *k*-NN classification algorithm that we used in the proposed method has a lower computational requirement than other classification algorithms and this is one of the supporting factors for this approach. Also, finding and using appropriate features reduce feature vector size and computational complexity in real-time applications. In our future work, we will focus on the design of a device that implements fully automatic lung sound classification.

Acknowledgements The authors would like to thank TUBITAK (The Scientific and Technological Research Council of Turkey) for supporting this work with project number 116E003.

References

1. Lehrer S (2008) Understanding lung sounds with audio CD, 3rd ed. WB Saunders, London, England
2. Aras S, Öztürk M, Gangal A (2018) Automatic detection of the respiratory cycle from recorded, single-channel sounds from lungs. Turk J of Electr Eng Comput Sci 26:11–22. https://doi.org/10.3906/elk-1705-16
3. Yilmaz CA, Kahya YP (2006) Multi-channel classification of respiratory sounds. Conf Proc IEEE Eng Med Biol Soc 2006:2864–2867. https://doi.org/10.1109/IEMBS.2006.259385
4. Murphy R (2007) Computerized multichannel lung sound analysis. Development of acoustic instruments for diagnosis and management of medical conditions. IEEE Eng Med Biol Mag 26:16–19. https://doi.org/10.1109/memb.2007.289117

5. Sen I, Kahya YP (2005) A multi-channel device for respiratory sound data acquisition and transient detection. Conf Proc IEEE Eng Med Biol Soc 2005:6658–6661. https://doi.org/10.1109/IEMBS.2005.1616029
6. Islam MA, Bandyopadhyaya I, Bhattacharyya P, Saha G (2018) Multichannel lung sound analysis for asthma detection. Comput Methods Programs Biomed 159:111–123. https://doi.org/10.1016/j.cmpb.2018.03.002
7. Charleston-Villalobos S, Martinez-Hernandez G, Gonzalez-Camarena R et al (2011) Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. Comput Biol Med 41:473–482. https://doi.org/10.1016/j.compbiomed.2011.04.009
8. Messner E, Fediuk M, Swatek P et al (2020) Multi-channel lung sound classification with convolutional recurrent neural networks. Comput Biol Med 122:103831. https://doi.org/10.1016/j.compbiomed.2020.103831
9. Altan G, Kutlu Y, Gökçen A (2020) Chronic obstructive pulmonary disease severity analysis using deep learning on multichannel lung sounds. Turk J of Electr Eng Comput Sci 28:2979–2996. https://doi.org/10.3906/elk-2004-68
10. Huq S, Moussavi Z (2012) Acoustic breath-phase detection using tracheal breath sounds. Med Biol Eng Comput 50:297–308. https://doi.org/10.1007/s11517-012-0869-9
11. Tabata H, Hirayama M, Enseki M et al (2016) A novel method for detecting airway narrowing using breath sound spectrum analysis in children. Respir Investig 54:20–28. https://doi.org/10.1016/j.resinv.2015.07.002
12. Yahya O, Faezipour M (2014) Automatic detection and classification of acoustic breathing cycles. In: Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education. IEEE
13. Dabiri S, Masnadi Shirazi MA (2015) Estimation of respiratory rate from photoplethysmogram signal of sleep apnea patients: a comparative study of different methods. In: 2015 38th International Conference on Telecommunications and Signal Processing (TSP). IEEE
14. Waitman LR, Clarkson KP, Barwise JA, King PH (2000) Representation and classification of breath sounds recorded in an intensive care setting using neural networks. J Clin Monit Comput 16:95–105. https://doi.org/10.1023/a:1009934112185
15. Bahoura M (2009) Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Comput Biol Med 39:824–843. https://doi.org/10.1016/j.compbiomed.2009.06.011
16. Sen I, Saraclar M, Kahya YP (2015) A comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds. IEEE Trans Biomed Eng 62:1768–1776. https://doi.org/10.1109/TBME.2015.2403616
17. Palaniappan R, Sundaraj K, Lam CK (2016) Reliable system for respiratory pathology classification from breath sound signals. In: 2016 International Conference on System Reliability and Science (ICSRS). IEEE
18. Göğüş FZ, Karlık B, Harman G (2016) Identification of pulmonary disorders by using different spectral analysis methods. Int J Comput Intell Syst 9:595. https://doi.org/10.1080/18756891.2016.1204110
19. Koeipensri T, Boonchoo P, Sueaseenak D (2016) The development of biosignal processing system (BPS-SWU V1. 0) for learning and research in biomedical engineering. In: 9th Biomedical Engineering International Conference (BMEiCON). Laos, pp 1–4
20. Sankur B, Kahya YP, Çağatay Güler E, Engin T (1994) Comparison of AR-based algorithms for respiratory sounds classification. Comput Biol Med 24:67–76. https://doi.org/10.1016/0010-4825(94)90038-8
21. Chamberlain D, Kodgule R, Ganelin D et al (2016) Application of semi-supervised deep learning to lung sound analysis. Annu Int Conf IEEE Eng Med Biol Soc 2016:804–807. https://doi.org/10.1109/EMBC.2016.7590823
22. Zulfiqar R, Majeed F, Irfan R et al (2021) Abnormal respiratory sounds classification using deep CNN through artificial noise addition. Front Med (Lausanne) 8:714811. https://doi.org/10.3389/fmed.2021.714811
23. Kim Y, Hyon Y, Jung SS et al (2021) Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. Sci Rep 11:17186. https://doi.org/10.1038/s41598-021-96724-7
24. Belkacem AN, Ouhbi S, Lakas A et al (2021) End-to-end AI-based point-of-care diagnosis system for classifying respiratory illnesses and early detection of COVID-19: a theoretical framework. Front Med (Lausanne) 8:585578. https://doi.org/10.3389/fmed.2021.585578
25. Gurung A, Scrafford CG, Tielsch JM et al (2011) Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. Respir Med 105:1396–1403. https://doi.org/10.1016/j.rmed.2011.05.007
26. Hjorth B (1970) EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 29:306–310. https://doi.org/10.1016/0013-4694(70)90143-4
27. Rabiner LR, Juang B-H (1993) Fundamentals of speech recognition
28. Makhoul J (1975) Linear prediction: a tutorial review. Proc IEEE Inst Electr Electron Eng 63:561–580. https://doi.org/10.1109/proc.1975.9792
29. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19:711–720. https://doi.org/10.1109/34.598228
30. Yavuz E, Aydemir O (2016) Olfaction recognition by EEG analysis using wavelet transform features. In: 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA). IEEE
31. Sain SR, Vapnik VN (1996) The nature of statistical learning theory. Technometrics 38:409. https://doi.org/10.2307/1271324
32. Theodoridis S, Koutroumbas K (2014) Pattern recognition, 3rd edn. Academic Press
33. Şen I, Saraclar M, Kahya YP (2015) A comparison of DVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds. IEEE Trans Biomed Eng 62:1768–1776

M. Alptekin Engin received the B.S. degree in electrical and electronics engineering from Cumhuriyet University, Sivas, Turkey, in 2007. He received the M.S. and Ph. D. degrees from Ataturk University, Erzurum, Turkey, in 2010 and 2016 respectively, both in electrical and electronics engineering. His current research interests include the areas of image/video processing and multimedia compression.

Selim Aras received the B.S. degree in electrical and electronics engineering from Ondokuz Mayıs University, Samsun, Turkey, in 2002, the M.S. degree from the Ataturk University, Erzurum, Turkey, in 2009, and the Ph. D. degree from the Karadeniz Technical University, Trabzon, Turkey, in 2018, both in electrical and electronics engineering. His current research interests include the areas of image and biomedical signal processing.

Ali Gangal received the B.S. degree in electrical and electronics engineering from Karadeniz Technical University, Trabzon, Turkey, in 1986, He received the M.S. and Ph. D. degrees from Karadeniz Technical University, Trabzon, Turkey, in 1989 and 1995 respectively, both in electrical and electronics engineering. His current research interests include the areas of image/video and biomedical signal processing.