Detection of Respiratory Sounds Based on Wavelet Coefficients and Machine Learning

FEI MENG[1], YAN SHI[1,2], NA WANG[1], MAOLIN CAI[1], AND ZUJING LUO[3]

[1]School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
[2]State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310058, China
[3]Beijing Engineering Research Center of Respiratory and Critical Care Medicine, Department of Respiratory and Critical Care Medicine, Beijing Institute of Respiratory Medicine, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China

Corresponding authors: Yan Shi (shiyan@buaa.edu.cn) and Na Wang (lion_na@buaa.edu.cn)

ABSTRACT Respiratory sounds reveal important information of the lungs of patients. However, the analysis of lung sounds depends significantly on the medical skills and diagnostic experience of the physicians and is a time-consuming process. The development of an automatic respiratory sound classification system based on machine learning would, therefore, be beneficial. In this study, 705 respiratory sound signals (240 crackles, 260 rhonchi, and 205 normal respiratory sounds) were acquired from 130 patients. We found that similarities between the original and wavelet decomposed signals reflected the frequency of the signals. The Gaussian kernel function was used to evaluate the wavelet signal similarity. We combined the wavelet signal similarity with the relative wavelet energy and wavelet entropy as the feature vector. A 5-fold cross-validation was applied to assess the performance of the system. The artificial neural network model, which was applied, achieved the classification accuracy and classified the respiratory sound signals with an accuracy of 85.43%.

INDEX TERMS Respiratory sound, relative wavelet energy, wavelet entropy, wavelet similarity, cross validation, artificial neural network.

I. INTRODUCTION

Because respiratory sounds convey important lung information of patients, the auscultation of lung sounds is a fundamental component of a pediatric lung disease diagnosis, similar to the diagnosis of pneumonia, bronchitis, and sleep apnea [1]–[3]. Crackles and rhonchi are the most common adventitious lung sounds. Crackles are explosive sounds caused by fluid bubbles in the tracheal or bronchial tubes, and rhonchi are caused by the obstructed pulmonary airways when air flows through these tubes. Estimation of crackles and rhonchi is vital in lung diagnosis. However, the auscultation depends greatly on the medical skills and diagnostic experience of the physician, which are difficult to acquire. With the development of computer-based respiratory sounds, automatic lung sound recognition based on machine learning has an important clinical significance for the diagnosis of lung abnormalities [4].

There are three main methods used in the feature extraction of respiratory sounds, i.e., statistics in the time-frequency domain, wavelet coefficients, and cepstrum coefficients.

The associate editor coordinating the review of this manuscript and approving it for publication was Shahzad Mumtaz.

Statistics in the time-frequency domain are intuitive features of lung sounds. Naves *et al.* use higher order statistics to extract features. Naves *et al.* [5] employ temporal–spectral dominance-based features. Auto-regressive (AR) models have also been widely used in the classification of lung sounds [6], [7]. However, certain individual statistics in the time and frequency domains cannot reveal the time-frequency properties of such sounds. Statistics in the wavelet domain have also been widely used in respiratory sound classification. Chang and Lai [8] chose the mean, average energy, and standard deviation of the wavelet coefficients in every wavelet layer and the ratio of the absolute mean values of adjacent sub-bands as the feature vectors. A new type of feature extraction method based on the fast wavelet transform is presented in [9]. The frequency distribution of lung sounds cannot be characterized by the simple statistics of the wavelet coefficients. Cepstrum coefficients, particularly Mel frequency cepstrum coefficients (MFCCs), which are used to evaluate the formants in the spectrum, have been widely applied to research on speech recognition. Numerous researchers have applied MFCC to lung sound recognition in recent years [10]–[13]. However, the formants of respiratory sounds are not obvious, and the vector dimensions of MFCCs

are extremely high, which means a large dataset is needed during the training process. In conclusion, a low-dimensional feature vector that evaluates the frequency distribution of the respiratory sounds is needed. Therefore, the feature vector for respiratory sound classification must be studied.

Clinical respiratory sounds are difficult to acquire in practice, and the sample set of lung sounds is often small. Therefore, researchers have generally chosen traditional machine learning methods, such as an artificial neural network (ANN) [9], [14], hidden Markov model (HMM) [15], [16], support vector machine (SVM) [17], [18], or k-Nearest Neighbor [4], instead of a deep learning method for the classification of lung sounds. Chamberlain *et al.* [19] attempted to recognize wheezes and crackles through deep learning conducted on 11,627 sounds recorded from 11 different auscultation locations on 284 patients. However, the signal samples have strong correlations, and the classification model is not generalized. Because of the small sample dataset, the feature vectors play a more important role than the selection of a classification model for lung sound classification. Renard *et al.* [20] evaluated the discriminatory ability of different types of features, including WT and MFCC used in former studies, based on the evaluation index MCC, ROC, AUC, and F1 score. They found that certain individual features show good results in wheeze recognition, whereas combinations of features increase the accuracy. Haider *et al.* [21] investigated the chronic obstructive pulmonary disease (COPD) using lung sounds and obtained excellent results. They improved the classification results to 100% by combining lung sound parameters with spirometry parameters. Ashok *et al.* [22] proposed a new method for classifying normal and abnormal lung sounds using an ELM network. The proposed method achieves a classification accuracy of 92.86%. Jaber *et al.* [23] proposed a telemedicine framework for lung sound based on the telemedicine framework. Messner *et al.* [24] proposed a multi-channel lung sound classification method. They selected the convolutional recurrent neural network and obtained the F1 score approximates to 92%. Rizal *et al.* [25] use multi-scale Hjorth descriptors for lung signal classification and achieves a high accuracy. The 2017 Int. Conf. on Biomedical Health Informatics (ICBHI) is an important the open-source lung sound dataset. Numerous researches are based on the ICBHI dataset [26]–[29]. Perna and Tagarelli [26] investigated the open-source lung sound dataset of the Int. Conf. on Biomedical Health Informatics (ICBHI), by combining MFCC feature extraction methods with an RNN, the results of which outperformed other competing ICBHI methods. Acharya and Basu [27] propose a deep CNN-RNN model for lung sound classification based on MFCC.

In this study, it was found that if the frequency spectrum of the original signal concentrates on the same frequency range of a certain wavelet sub-signal, the origin signal is similar to the wavelet sub-signal. Therefore, in this research, the frequency properties of the signals can be characterized based on the signal similarity between the wavelet sub-signals and the original signals. The Gaussian kernel function is selected to evaluate the signal similarity as a part of the feature vector. The Gaussian kernel function measures the signal similarity and scales the signal similarity into a range of zero to one.

In addition, the relative wavelet energy (RWE) and wavelet entropy (WE) are used. RWE and WE were first proposed by Rosso *et al.* [30] in research into brain electrical signal processing. RWE and WE are widely used in EEG signal processing [31], [32] and have been introduced into ECG signal processing [33]. However, a few researchers have used RWE and WE in audio signal processing. In this research, RWE is used to measure the energy distribution in different wavelet bands, and WE is applied to evaluate the RWE distribution.

In this paper, 705 lung sound signals with 240 crackle signals, 260 rhonchus signals, and 205 normal respiratory sound signals are acquired from 130 patients. All signals are divided into 5 groups, with 48 crackle signals, 52 rhonchus signals, and 41 normal respiratory sound signals in each group. A 5-fold cross-validation is applied to assess the performance of the system. In every step of the training process, four groups were chosen as the training dataset, and the remaining group is chosen as the test set. A 15-D feature vector is obtained with seven dimensions of the relative wavelet energy, one dimension of the wavelet entropy, and seven dimensions of the Gaussian kernel functions. Three classification methods, i.e., an SVM, a KNN, and an ANN, were tested. The results show that an ANN has the highest classification accuracy of 85.43%.

The rest of this paper is organized as follows: Section 2 briefly introduces the lung sound signal acquisition scheme. In Section 3, a wavelet transform and multi-resolution wavelet decomposition are introduced. In Section 4, the feature extraction method is described. The feature vector comprises the wavelet energy, wavelet entropy, and Gaussian kernel function. In Section 5, an SVM, an ANN, and a KNN are used to design the classifier. The results are presented in Section 6. Finally, Section 7 summarizes the procedure followed by the algorithm and proves the validity of the characteristic vector choice.

II. LUNG SOUND SIGNAL ACQUISITION

All breath signals were acquired from the pediatric department in the China-Japan Friendship Hospital in China. This research methodology was approved by the Institutional Ethical Committee of the China-Japan Friendship Hospital and informed consent was obtained from the participants. For the participants who are minors, the research method was approved by their parents. All abnormal lung sound signals were collected from patients having pneumonia or bronchitis. All normal signals were acquired from patients with other diseases, such as heart or stomach diseases. The 705 signals were collected using a 3M Littmann electronic stethoscope on 44 patients having lung crackles, 50 patients

with rhonchi, and 36 patients having normal respiratory sounds. Five or six respiratory cycles are selected from every patient. The respiratory cycles from one patient were collected on different days. The sampling frequency of the signals was 4 000 Hz. The signal collection equipment is shown in Figure 1.

FIGURE 1. Respiratory sound recorded equipment.

The signals were pre-processed to reduce environmental noises. All signals were filtered using the algorithm mentioned in [34]. The algorithm is an integrated serial filter consisting of a Chebyshev band-pass filter, a wavelet de-noising filter, and an adaptive filter. The signal-to-noise ratio (SNR) of the filtered signals was greater than five.

The duration of the signals ranged from 1 to 3 s. The signals were segmented according to the respiratory cycle. All signals have periodic inspiratory and expiratory phases. The signals are segmented based on their amplitude. First, a signal amplitude threshold was set. Signals higher than the threshold were respiratory sounds, and signals lower than the threshold were noises. Second, the signals were segmented using an alternation of noises and respiratory sound components.
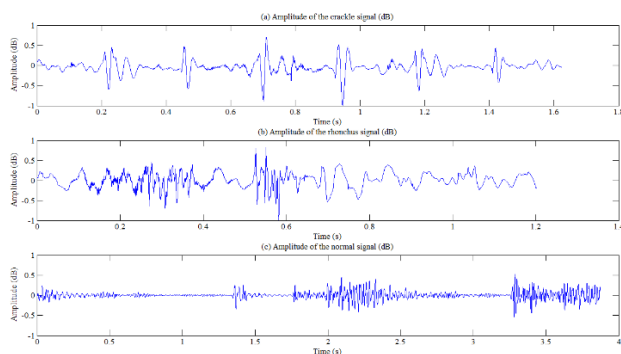

FIGURE 2. Typical respiratory sound signals.

III. WAVELET ANALYSIS
A. WAVELET TRANSFORM
A Fourier transform (FT) is a traditional method used to study the frequency of a signal. However, FT provides the
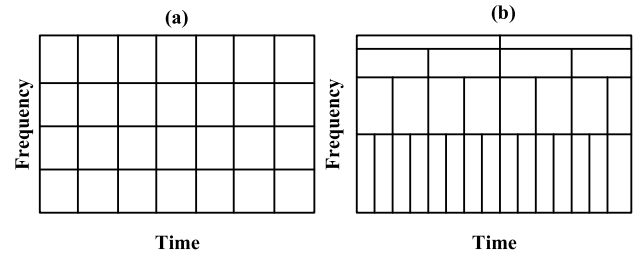

FIGURE 3. (a) Time-Frequency window of STFT, (b) Time-Frequency window of WT.

frequency information of the signal, not the frequency information within the time location. Therefore, the FT does not provide sufficient frequency information of non-stationary signals. A short time Fourier transform (STFT) utilizes window functions to segment the non-stationary signals into short-term sub-signals. The short-time sub-signals are considered stationary signals. The FT of every short-time sub-signal determines the frequency components of the sub-signal time location. In addition, the signals are mapped into two dimensions, i.e., the frequency and time domain. However, the sinusoidal frequency of an STFT has a constant time-frequency window, as shown in Figure 3 (a), limiting the promotion of STFT in the time-frequency analysis. To solve this problem, a wavelet transform (WT) is proposed. The WT determines the frequency components within the wavelet domain. The time-frequency windows of the wavelet domain are variable, as shown in Figure 3 (b). This guarantees the time domain resolution at low-frequency scales and the frequency domain resolution at high-frequency scales. The WT is as follows

$$CWT_f(a, b) = \int_{-\infty}^{+\infty} f(x)\overline{\varphi_{a,b}(x)}dx \qquad (1)$$

where a is the dyadic dilation, b is the dyadic position, and the wavelet mother function φ(x) is defined as:

$$\varphi_{a,b}(x) = |a|^{-\frac{1}{2}} \varphi(\frac{x - b}{2})a, b \in R \qquad (2)$$

The discrete wavelet transform (DWT) is as follows:

$$DWT_f(m, k) = \int_{-\infty}^{+\infty} f(x) \cdot \varphi_{m,k}(x)dx$$
$$= a^{-\frac{m}{2}} \int_{-\infty}^{+\infty} f(x) \cdot \varphi(a^{-m}t - kb)dx \qquad (3)$$

B. MULTI-RESOLUTION WAVELET DECOMPOSITION
The MALLAT algorithm is a simple algorithm of WT. Wavelet coefficients are obtained through the multi-resolution decomposition of the signals in the MALLAT algorithm [35], [36]. The process used by the algorithm is shown in Figure 4. The signal is decomposed into a high-frequency part by a high-pass filter h[n] and a low-frequency part by a low-pass filter g[n]. The relationship between the two filters is as follows:

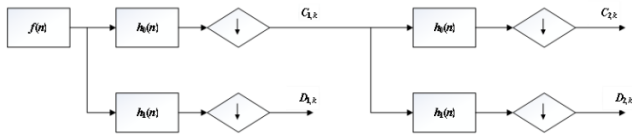$$g[k] = (-1)^k h[-k + 1] \qquad (4)$$

FIGURE 4. Structure of MALLAT algorithm.

After the down-sampling process, the wavelet coefficients in layer j are obtained. The high-frequency part of the signal is transformed into the detail coefficients Dj,k, and the low-frequency part is transformed into the approximation coefficients Aj,k, where j and k are the dyadic dilation and dyadic position, respectively.

The same decomposition process is repeated on the approximation coefficients Aj,k to obtain the detailed coefficients Dj+1,k and the approximation coefficients Aj+1,k at a higher resolution. The relationship between the wavelet coefficients at different resolutions is

$$A_{j+1,k} = \sum_n g[n]A_{j,n+2k} \tag{5}$$

$$D_{j+1,k} = \sum_n h[n]D_{j,n+2k} \tag{6}$$

After n iterations, n groups of detail coefficients Dj,k, j = 1, 2,...,n and one group of approximation coefficients An,k are obtained. The frequency range of the wavelet coefficients in layer j is as follows:

$$\frac{1}{2^{j+1}} \cdot f_s \leq f_D^j \leq \frac{1}{2^j} \cdot f_s \tag{7}$$

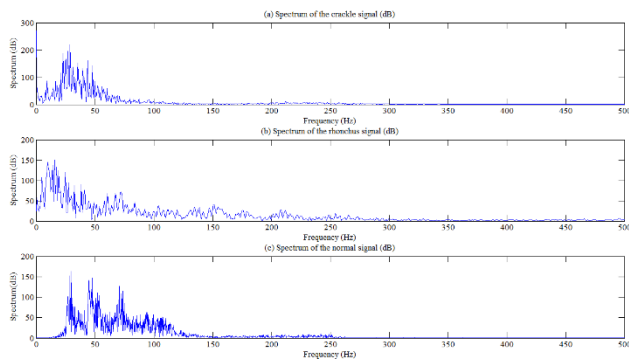where f_s is the sampling rate of the original signal.

FIGURE 5. Frequency spectrum of the crackle signal, the rhonchi signal and the normal signal.

IV. FEATURE EXTRACTION

Rhonchi, crackles, and normal respiratory sounds have different frequency distributions, as shown in Figure 5. Therefore, the statistics of the wavelet coefficients are selected to evaluate the frequency components in the wavelet domain. Authors use permutation entropy as the wavelet selection criteria like [37]. We compare the Daubechies (db) series (db1 ∼ db7), coif series (coif1 ∼ coif2) and sym

series(sym2 ∼ sym7). The coif2 has the least permutation entropy. The signal is decomposed into six wavelet layers using the coif2 wavelet base. The sampling frequency of the signals collected is 4000 Hz. The frequency ranges of the wavelet coefficients in different layers are presented in Table 1. In this research, the relative wavelet energy, wavelet entropy, and similarity between the original and sub-signals are chosen as the elements of the feature vector.

TABLE 1. Frequency ranges of wavelet coefficients in different wavelet layers.

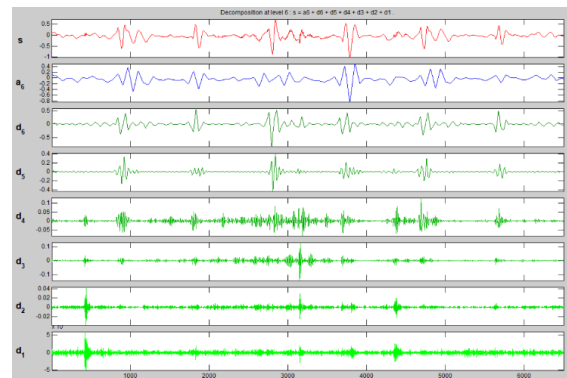| Coefficient layer	Frequency range(Hz)
CD1	1000-2000
CD2	500-1000
CD3	250-500
CD4	125-250
CD5	62.5-125
CD6	31.25-62.5
CD7	15.625-31.25
CA7	15.625-31.25

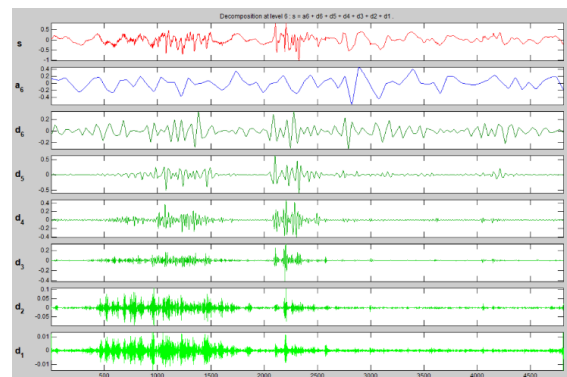**FIGURE 6.** Wavelet sub-signals and the original signal of a crackle signal.

FIGURE 7. Wavelet sub-signals and the original signal of a rhonchus signal.

A. WAVELET SIMILARITY

Figures 6–8 show the wavelet sub-signals and original signals of the crackles, rhonchi, and normal respiratory sounds, respectively. It was found that if the wavelet sub-signal had
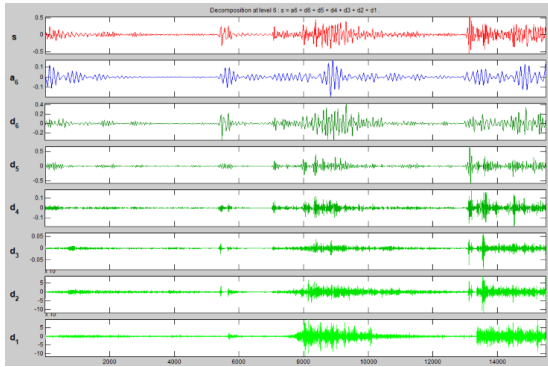
FIGURE 8. Wavelet sub-signals and the original signal of a normal signal.

the same frequency component of the original signal, the sub-signal of that wavelet layer would be similar to the original signal. As shown in Figures 5 and 6, the frequency of the crackle signal concentrates on the lower frequency, and the original crackle signal is similar to the wavelet sub-signals a6 and d6. As shown in Figures 5 and 8, the normal lung signal frequency concentrates within the higher frequency, and the original signal is similar to d4, d5, and d6. As shown in Figure 5, it was found that the Rhonchus signal had a wide frequency distribution. In addition, as shown in Figure 7, the wavelet sub-signals d3, d4, and d5 are similar to the original signal in the inspiratory of the original signal, whereas the sub-signal d6 is similar to the expiratory of the original signal. Because the frequency distributions of the crackles, rhonchi, and normal lung sounds are different, the similarities between the sub-signals and the original signals were chosen as the elements of the feature vector.

In this study, the kernel function was used to measure the similarity of the original and sub-signals. A kernel function is a symmetric function that measures the correlation of different signals, and is defined as follows:

$$k(x, x') = \varphi(x)^T \varphi(x') \qquad (8)$$

where x and x' are different signals, and $\varphi(x)$ is the function of signal x. Kernel functions are often used with the SVM method, which are proportional to the signal similarity. In this, the Gaussian kernel function, a type of homogeneous kernel, is used through the following:

$$k(x, x') = \exp(-\frac{\|x - x'\|^2}{\sigma^2}) \qquad (9)$$

which is proportional to the Euclidean distance of the two signals. In addition, the standard deviation σ is chosen as one. The wavelet correlation coefficients are not used because the Gaussian kernel function scales the similarity values into the range of zero to one. The normalized values are helpful for the training step.

Because the kernel function is influenced by the amplitude of the signals, the sub-signals need to first be normalized. The normalization method is based on the power of the signals.

The power of a discrete signal is given by the following:

$$P = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2, \qquad (10)$$

where $x[n]$ is the sampling point of the signal, and N is the length of the signal acquired. The original signal is normalized by the following:

$$s_2 = \sqrt{\frac{P_1}{P_2}} \cdot s \qquad (11)$$

where P_1 is the power of a standard signal, and s and P_2 are the signal and power of the signal to be processed, respectively. In this way, the amplitudes of the signals are normalized to a similar scale.

The similarity between the original signal and the sub-signal in the wavelet layer i is calculated as follows:

$$SIM_i = \exp(-\|s - s_i\|^2) \qquad (12)$$

where s_i is the sub-signal of wavelet layer i, and s is the original signal. In addition, SIM_is are chosen as the first seven elements of the feature vector.

B. RELATIVE WAVELET ENERGY
The energy of the wavelet coefficients is an intuitive statistic describing the frequency distribution of the wavelet coefficients in different wavelet layers.

$$WEN_i = \sum_k W_k^{i2} \qquad (13)$$

where W_{ki} is the kth wavelet coefficient in layer i. The energy is defined as the wavelet energy. However, the wavelet energy is proportional to the energy of the original signals. Therefore, the relative wavelet energy (*RWE*) is introduced as follows:

$$RWE_i = \frac{WEN_i}{WEN_{total}} \qquad (14)$$

where WEN_i is the wavelet energy of layer i, and WEN_{total} is the total energy calculated as

$$RWE_{total} = \sum WEN \qquad (15)$$

where *RWE*s are chosen as the next seven elements of the feature vector.

C. WAVELET ENTROPY
The Shannon entropy [38] measures the disorder of the probability distribution of a random process. The definition of entropy is described as follows:

$$S_{WT} = \sum_i p_i \ln(\frac{1}{p_i}) \qquad (16)$$

where p_i is the probability of a random process. If the probability distribution is more concentrated, the Shannon entropy is lower. By contrast, if the probability distribution is more dispersed, the Shannon entropy is higher.
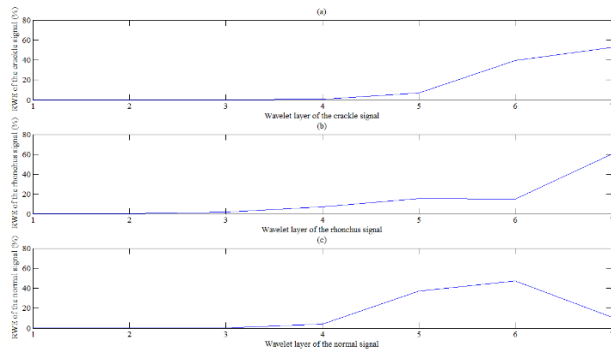
FIGURE 9. RWE of crackle, rhonchus and normal signals.

The idea of entropy is introduced to measure the disorder of the *RWE* distribution [20]. If the probability p_i is replaced by RWE_i, the Shannon entropy is defined as the wavelet entropy (*WE*). As shown in Figure 9, the *RWE* of the rhonchus concentrates on the sixth approximation coefficients, and the *RWE* of the sixth approximation coefficients is higher than 60%. The *RWE* of the crackle concentrates on the sixth approximation coefficients and the sixth detail coefficients. However, the *RWE* distribution of the normal lung sound signal is more disperse and therefore the *WE* of the crackle, rhonchus, and normal respiratory sound signals are different. In addition, the *WE* is chosen as the last element of the feature vector.

D. STATISTICAL SIGNIFICANCE ANALYSIS

E. Not all the features extracted in this research are equally important for the lung sound recognition. If a feature has similar statistical distributions among the crackles, rhonchus and normal lung sounds, the feature should not be included in the feature vector. Therefore, statistical significance analysis should be conducted to reduce the computation time and increase the classification accuracy. Authors firstly conduct the normality test by the Kolmogorov-Smirnov test. The result is shown in Table 2. All the features do not follow the normal distribution. Therefore, the Non-parametric test should be used for checking the significance level of the features. Authors use the Mann-Whitney U test at 95% confidence interval to check the statistical significances of the extracted features. Because there are three classes, the Non-parametric tests are carried by pairs. The results are shown in the following table. RWE in detail layer 6 and SIMR in detail layer 2,3,4,5 are found not statistically significant between crackles and wheezes with p>0.05. But the features are statistically significant between abnormal signals and normal signals. Therefore, the features are reserved for recognizing the normal lung sounds. RWE in detail layer 4 is reserved for the same reason. Therefore, all the features are reserved for the classification.

V. CLASSIFICATION DESIGN

A. SUPPORT VECTOR MACHINE

SVM is a non-linear classifier that maximizes the margin of the samples in the hyper plane. The margin is the distance

TABLE 2. Statistical significance analysis of features.

| Features	Normality test (p value)	Mann-Whitney U test between wheezes and crackles (p value)		
		Crackles/ Wheezes	Crackles/ Normal	Wheezes/ Normal
RWE in detail layer 1	0.000	0.000	0.000	0.003
RWE in detail layer 2	0.000	0.000	0.000	0.000
RWE in detail layer 3	0.000	0.000	0.000	0.000
RWE in detail layer 4	0.000	0.044	0.324	0.003
RWE in detail layer 5	0.000	0.048	0.000	0.000
RWE in detail layer 6	0.000	0.343	0.000	0.000
RWE in approximation layer 6	0.000	0.004	0.000	0.000
WE	0.000	0.000	0.000	0.000
SIMR in detail layer 1	0.000	0.057	0.000	0.000
SIMR in detail layer 2	0.000	0.537	0.000	0.000
SIMR in detail layer 3	0.000	0.850	0.000	0.000
SIMR in detail layer 4	0.000	0.268	0.000	0.000
SIMR in detail layer 5	0.000	0.086	0.000	0.000
SIMR in detail layer 6	0.000	0.015	0.000	0.000
SIMR in approximation layer 6	0.000	0.000	0.000	0.000
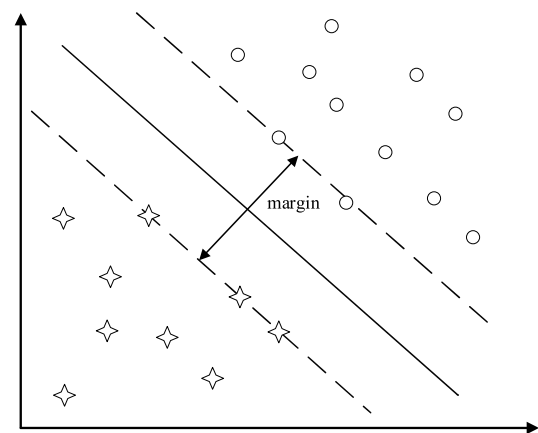
FIGURE 10. Image of the margin.

between a straight line and the two points closest to the line, as shown in Figure 10. The maximal margin in the hyper plane is called the max margin hyperplane (MMH). The sample points lying on the MMH are called support vectors. A classifier based on support vectors is called an SVM.

The hyper plane is defined as:

$$y = \omega x + b \qquad (17)$$

where x is the feature vector, ω is the weight vector, b is the bias, and y is the output. The hyper plane classifying the samples applies the following formula:

$$y\,(\omega x + b) \geq 1 \qquad (18)$$

The outputs of the SVM are $+1$ and -1. The distance from a sample point to the hyper plane is

$$\gamma = \left(\frac{\omega x + b}{\|\omega\|}\right) y \qquad (19)$$

To maximize the distance, $\|\omega\|$ is minimized. The training target of SVM is shown as

$$\min_{\omega, b} \frac{1}{2}\,\|\omega\|^2$$
$$s.t.\ y_i\,(\omega x_i + b) \geq 1 \quad i = 1, 2, \ldots, n \qquad (20)$$

A Lagrangian is selected for optimization:

$$L\,(b, \alpha, \omega) = \frac{1}{2}\,\|\omega\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i\left(\omega^T x_i + b\right) - 1 \right] \qquad (21)$$

By setting the derivatives of L to zero with respect to ω and b, ω is obtained as follows:

$$\omega = \sum_{i=1}^{n} \alpha_i y_i x_i \qquad (22)$$

The training target is reformulated as follows:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y^{(i)} y^{(i)} \alpha_i \alpha_j \left\langle x_i, x_j \right\rangle$$
$$s.t.\ \alpha_i \geq 0 \quad i = 1, 2, \ldots, n$$
$$\sum_{i=1}^{n} \alpha_i y^{(i)} = 0 \qquad (23)$$

To solve the non-separable case, the regularization factors C are introduced and reformulated Eq. (23):

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y^{(i)} y^{(i)} \alpha_i \alpha_j \left\langle x_i, x_j \right\rangle$$
$$s.t.\ C \geq \alpha_i \geq 0 \quad i = 1, 2, \ldots, n$$
$$\sum_{i=1}^{n} \alpha_i y^{(i)} = 0 \qquad (24)$$

To reduce the operational complexity of the inner products, the kernel functions are used to replace the inner product:

$$k\,(x_i, x_j) = \left\langle x_i, x_j \right\rangle \qquad (25)$$

The regular method used to obtain the coefficients is the sequential minimal optimization (SMO) algorithm [39].

For the SVM parameters, a context-aware support vector machine (C-SVM) is selected as the SVM type and a radial basis function (RBF) as the kernel function. The coefficient γ in the RBF function is 0.6667. The cost C is set to five and the minimum step size is 0.0001. The mean support vector is 334.8.

B. BP NEURAL NETWORK

The structure of a linear classifier can be defined as follows:

$$y\,(x, \omega) = f\left(\sum_{i=1}^{M} \omega_i \phi_i\,(x)\right) \qquad (26)$$

where ω_i indicates the coefficients of the linear classifier, $\phi_i(x)$ is the nonlinear function of input data x, and function $f\,(\cdot)$ is the nonlinear activation function.

In the BP neural network, the nonlinear function $\phi_i(x)$ can be regarded as the same model of (22). The model of the BP neural network is defined as follows:

$$y\,(x, \omega) = f\left(\sum_{j=1}^{M} \omega_{kj} h\left(\sum_{i=1}^{D} \omega_{ji} \eta\,(x)\right)\right) \qquad (27)$$

if the nonlinear function $\eta(\cdot)$ is defined as

$$\eta(x_i) = x_i \qquad (28)$$

the network is regarded as a two-layer-network. If the number of layers increases, the function $\eta(\cdot)$ has the same form as (23). The structure of a two-layer BP neural network is shown in Figure 11, which is divided into an input layer, a middle layer, and an output layer. The input layer has 15 nodes for the 15-dimensional feature vector, the middle layer has 500 nodes, and the output layer has 3 nodes. The outputs of the layer are the probabilities of occurrence of every type of sound. A rectified linear unit (ReLU) is selected as the activation function. The ReLU is defined as follows:

$$f\,(x) = \max\,(x, 0) \qquad (29)$$

The Softmax function is selected as the activation function of the output layer. The Softmax function is defined as

$$s\,(i) = \frac{e^i}{\sum_j e^j} \qquad (30)$$

The target of the training process is to minimize the loss function. Cross entropy is chosen as the loss function.

$$L = -\sum_{i=1}^{N} \left(y^{(i)} \log\left(\hat{y}^i\right) + \left(1 - y^i\right) \log\left(1 - \hat{y}^i\right)\right) \qquad (31)$$

where y^i is the real label of the data, and $\hat{y}^i$ is the predicted label of the data. The coefficients are updated by the back propagation. For the optimizer, the root mean square prop (RMSProp) is chosen.

C. KTH NEAREST NEIGHBORS

The principle of the KNN depends on the Euclidean distance of the different points in an N-dimensional hyperplane, which is defined as follows:

$$y = \sqrt{\sum_{i=1}^{N} \left(x_{1,i} - x_{2,i}\right)^2} \qquad (32)$$

where $x_{1,i}$ and $x_{2,i}$ are the coordinates of two points. To classify a new sample, the K-nearest points between the

Input layer   Middle layer   Output layer

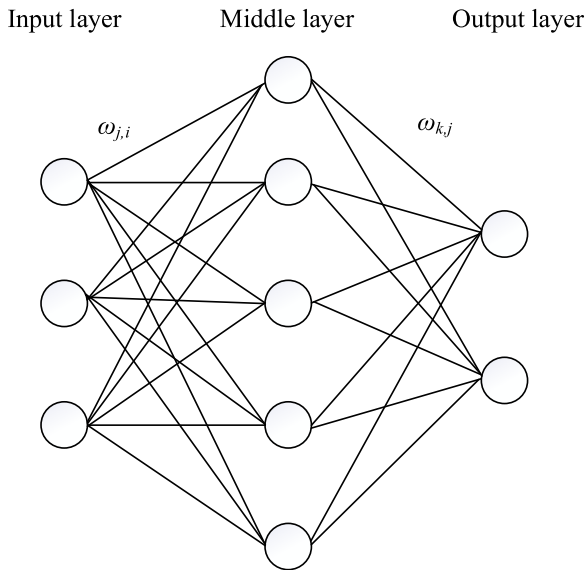$\omega_{j,i}$ $\omega_{k,j}$

FIGURE 11. Structure of BP neutral network.

sample points are chosen. The sample can be trusted to belong to the classification, which repeats the most in K points. In this research, the number of neighbors K is five.

VI. RESULTS AND DISCUSSION

A. ACCURACY OF MODELS

To avoid an over-fitting, a k-fold cross-validation process is designed for proving the correction of the scheme. The sampling set is divided into five groups, with 58 rhonchi samples, 42 crackle samples, and 41 normal lung sound samples. During every step of training, four groups are chosen as the training group, and the remaining group is chosen as the test group. The classification accuracy of the model is shown in Figure 12. The average classification accuracies of the SVM, ANN, and KNN are 69.50, 85.43, and 68.51%, respectively. The ANN model is chosen as the classifier of this research.

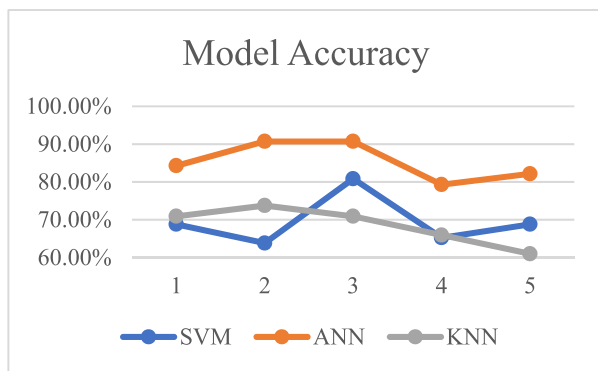**FIGURE 12.** Classification accuracies of three classifiers.

B. FURTHER ESTIMATION OF THE ANN MODEL

Accuracy, sensitivity, specificity, AUC (micro), AUC (macro), and F1 score are selected as the performance measures. The 5-fold cross-validated models are evaluated
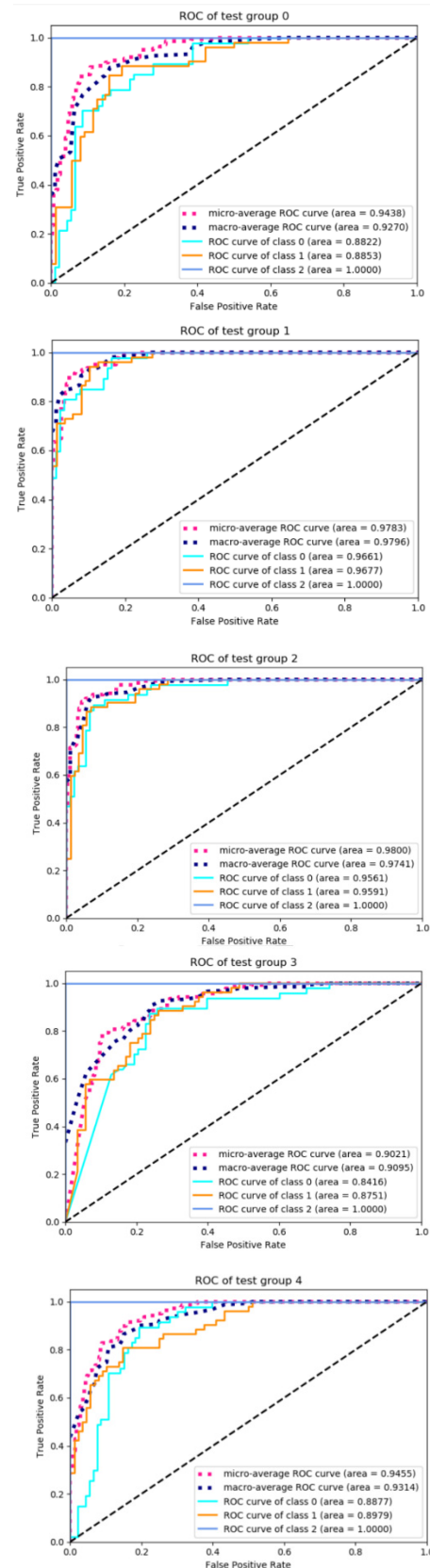
FIGURE 13. ROC curves of models.

and the results are presented in Table 3. The system has multiple classifications. Therefore, the sensitivity and

TABLE 3. Performance measures of models.

model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (micro)	AUC (macro)	F1 Score
model0	84.28	84.94	91.83	0.9438	0.9270	0.8500
model1	90.71	90.92	95.12	0.9783	0.9796	0.9107
model2	90.71	91.19	95.20	0.9800	0.9741	0.9121
model3	79.29	80.73	89.50	0.9021	0.9095	0.8027
model4	82.14	83.02	90.80	0.9455	0.9314	0.8289
average	85.42	86.16	92.49	0.9499	0.9443	0.8608

specificity of every category are calculated and the sensitivity and specificity are obtained based on the mean values. The average sensitivity of all models is 86.16%, and the average specificity of all models is 90.49%. The receiver operating characteristic (ROC) curves are shown for all models in Figure 13. We demonstrate the ROC curves of all classes, the macro-average ROC curve, and the micro-average ROC curve together. The area under the curve (AUC) are obtained for the macro- and micro-averages. All AUCs are higher than 90%. The F1 score was also used to evaluate the models. The average F1 score was 0.8608. The models showed good performances for the three categories applied to the test dataset. The models do not overfit the dataset. The true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are calculated by the mean value of every category. For example, the crackles, the patients with crackles were considered the positive group were evaluated, whereas the patients with rhonchi or normal lung sounds were considered as the negative group. In addition, three groups of TP, FP, TN, and FN were obtained. The final TP, FP, TN, and FN values were calculated using the mean value of the three groups.

C. DEEP LEARNING RESULT

Deep learning is widely used in sound classification. In this research, the long short-term memory (LSTM) deep learning model, which is widely used in sound recognition, is implemented for lung sound classification. The data are divided into five groups similar to machine learning methods. The network has 50 LSTM cells and a dense layer with the activation function of the soft-max function. The loss function is the categorical cross-entropy. The sensitivity, specificity, and accuracy of the training and test sets are presented in Table 3. The sensitivity, specificity, and accuracy are 100% for the training data. However, they are much lower on the test set, particularly sensitivity. All sensitivities are lower than 70% on the test set. Therefore, the deep learning model overfits when a small sample dataset is used.

TABLE 4. Performance measures of deep learning models.

Group	Train data			Test data		
	Acc (%)	Sensitivity (%)	Specificity (%)	Acc (%)	Sensitivity (%)	Specificity (%)
1	100.0	100.0	100.0	75.88	64.58	81.48
2	100.0	100.0	100.0	72.10	59.07	78.71
3	100.0	100.0	100.0	74.00	61.97	80.10
4	100.0	100.0	100.0	76.83	66.73	82.22
5	100.0	100.0	100.0	74.47	62.60	80.39

D. COMPARISON WITH SIMILAR APPROACHES

Most studies classifying respiratory sounds have only considered two types of respiratory sounds. Xaviero *et al.* [20] studied normal lung sounds and wheezes. Ashok *et al.* [22] studied respiratory sounds to distinguish between normal and abnormal subjects. However, crackles and wheezes are the most common abnormal lung sounds. Crackles and wheezes are symptoms of different respiratory disorders. Therefore, it is necessary to distinguish lung sounds from among crackles, wheezes, and normal lung sounds. Sengupta *et al.* [14] classified respiratory sounds into crackles, wheezes, and normal lung sounds. However, they used only 72 cycles as the training data. A multi-classification requires more training samples than the binary classification because the models overfit with small training sets.

In addition, the feature vector used in this study is low dimensional, and the structure of the neural network is simple. There are only 9503 parameters in the model, which is a much smaller number than in deep learning models, such as those developed by Perna and Tagarelli [26] and Chamberlain *et al.* [19]. The model is only 93 kilobytes (kb) in size and has a low calculation complexity. Therefore, the model is easily transplanted into a small auscultation device based on the micro-programmed control unit.

Haider *et al.* [21] use the median frequency and linear predictive coefficients combined with the spirometry parameters to classify the normal patients and COPD. The classification accuracy reaches 100%. The results prove that the respiratory sound classification may improve when combined with other medical indexes. Therefore, we will perform pulmonary disease studies in the future by combining respiratory sounds with other medical indexes.

E. RESUTL OF FEATURE VECTORS WITHOUT SMIs

The novelty of this research is the determination of the similarity between the sub-signals in different wavelet layers, and the original signals reflecting the frequency distribution of the signals. Therefore, the Gaussian kernel function is used to evaluate the signal similarities. The classification results are compared with a feature vector with and without wavelet sub-signal similarities. The results are presented in Table 5. From the results, the proposed wavelet sub-signal similarities increase the classification accuracy.

TABLE 5. Performance of the feature vector without SMIs.

Group	Accuracy	Sensitivity	Specificity
1	77.14%	68.75%	82.70%
2	78.09%	69.49%	83.27%
3	78.57%	68.04%	83.70%
4	79.04%	69.81%	84.07%
5	76.19%	65.28%	81.62%
average	77.81%	68.27%	83.07%

F. TEST RESULT WITH OPEN SOURCE RESPIRATORY SOUND DATABASES

The algorithm was tested using the 2017 ICBHI dataset. The signals are divided into two categories, normal and abnormal

sounds. There are signals in the training group as well as in the valid and test groups. The classification results are shown in Table 6. The model was proved to be effective on the ICBHI database.

TABLE 6. Test result of ICBHI data set.

Group	Accuracy	Sensitivity	Specificity
1	100.00%	100.00%	100.00%
2	100.00%	100.00%	100.00%
3	99.45%	100.00%	99.11%
4	100.00%	100.00%	100.00%
5	100.00%	100.00%	100.00%
average	99.89%	100.00%	99.82%

VII. CONCLUSION

In this paper, a classification model was proposed to classify crackles, rhonchi, and normal lung sounds of patients. The sample contained 705 signals acquired from 130 patients from a AAA hospital in China. The feature vector comprised the relative wavelet energy in seven wavelet layers, the wavelet entropy, and the Gaussian kernel functions to measure the wavelet similarity between the wavelet sub-signals and the original signal in seven layers. When compared, the artificial neural network showed the highest classification accuracy of 85.43% among the methods using SVM, AAN, and KNN. It was found that the similarity between the wavelet deposition sub-signals and original signal reflects the time-frequency characteristics of the signals. The statistics were chosen as the elements of the feature vector classifying the normal and abnormal lung sounds. However, some limitations to the methods exist that need to be improved. Because the Gaussian kernel function used in this study is related to the amplitude of the sub-signals, a normalization step is conducted before obtaining the signal similarities. This step may lead to an accumulation of errors. Therefore, a new statistic which measures the signal similarity should be proposed. In addition, the respiratory sounds can be combined with other medical parameters, such as spirometry parameters in intelligent disease recognition. Furthermore, multi-signal integration studies should be conducted in medical sound signal research. Although there are several wavelet families including Daubechies, in this study, a coif2 wavelet function was chosen because it achieves good performance under practical situations. Different wavelet functions and their classification accuracies will be considered in future studies.

REFERENCES

[1] Z. Moussavi, "Fundamentals of respiratory sounds and analysis," *Synth. Lectures Biomed. Eng.*, vol. 1, no. 1, pp. 1–68, Jan. 2006.

[2] R. G. Loudon and R. L. H. Murphy, "Lung sounds," in *Encyclopedia of Medical Devices and Instrumentation*. Hoboken, NJ, USA: Wiley, 2006.

[3] S. Reichert, R. Gass, A. Hajjam, C. Brandt, E. Nguyen, K. Baldassari, and E. Andrès, "The ASAP project: A first step to an auscultation's school creation," *Respiratory Med. CME*, vol. 2, no. 1, pp. 7–14, 2009.

[4] T. E. A. Khan and P. Vijayakumar, "Separating heart sound from lung sound using LabVIEW," *Int. J. Comput. Electr. Eng.*, vol. 2, no. 3, pp. 524–533, 2010.

[5] R. Naves, B. H. G. Barbosa, and D. D. Ferreira, "Classification of lung sounds using higher-order statistics: A divide-and-conquer approach," *Comput. Methods Programs Biomed.*, vol. 129, pp. 12–20, Jun. 2016.

[6] F. Jin, S. Krishnan, and F. Sattar, "Adventitious sounds identification and extraction using temporal-spectral dominance-based features," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 11, pp. 3078–3087, Nov. 2011.

[7] B. Sankur, Y. P. Kahya, E. Ç. Güler, and T. Engin, "Comparison of AR-based algorithms for respiratory sounds classification," *Comput. Biol. Med.*, vol. 24, no. 1, pp. 67–76, Jan. 1994.

[8] G.-C. Chang and Y.-F. Lai, "Performance evaluation and enhancement of lung sound recognition system in two real noisy environments," *Comput. Methods Programs Biomed.*, vol. 97, no. 2, pp. 141–150, Feb. 2010.

[9] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," *Comput. Biol. Med.*, vol. 34, no. 6, pp. 523–537, Sep. 2004.

[10] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 1, pp. 83–88, Jan. 1999.

[11] P. Mayorga, C. Druzgalski, R. L. Morelos, O. H. Gonzalez, and J. Vidales, "Acoustics based assessment of respiratory diseases using GMM classification," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 6312–6316.

[12] M. Bahoura and C. Pelletier, "Respiratory sounds classification using cepstral analysis and Gaussian mixture models," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2004, pp. 9–12.

[13] M. Bahoura and H. Ezzaidi, "Hardware implementation of MFCC feature extraction for respiratory sounds analysis," in *Proc. 8th Int. Workshop Syst., Signal Process. Appl. (WoSSPA)*, May 2013, pp. 226–229.

[14] N. Sengupta, M. Sahidullah, and G. Saha, "Lung sound classification using cepstral-based statistical features," *Comput. Biol. Med.*, vol. 75, pp. 118–129, Aug. 2016.

[15] M. Yamashita, S. Matsunaga, and S. Miyahara, "Discrimination between healthy subjects and patients with pulmonary emphysema by detection of abnormal respiration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 693–696.

[16] S. Matsunaga, K. Yamauchi, M. Yamashita, and S. Miyahara, "Classification between normal and abnormal respiratory sounds based on maximum likelihood approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 517–520.

[17] F. Jin, F. Sattar, and D. Y. T. Goh, "New approaches for spectro-temporal feature extraction with applications to respiratory sound classification," *Neurocomputing*, vol. 123, pp. 362–371, Jan. 2014.

[18] G. Serbes, C. O. Sakar, Y. P. Kahya, and N. Aydin, "Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 3314–3317.

[19] D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, and R. R. Fletcher, "Application of semi-supervised deep learning to lung sound analysis," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 804–807.

[20] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "Evaluation of features for classification of wheezes and normal respiratory sounds," *PLoS ONE*, vol. 14, no. 3, Mar. 2019, Art. no. e0213659.

[21] N. S. Haider, B. K. Singh, R. Periyasamy, and A. K. Behera, "Respiratory sound based classification of chronic obstructive pulmonary disease: A risk stratification approach in machine learning paradigm," *J. Med. Syst.*, vol. 43, no. 8, 2019, Art. no. 255.

[22] A. Mondal, P. Bhattacharya, and G. Saha, "Detection of lungs status using morphological complexities of respiratory sounds," *Sci. World J.*, vol. 2014, pp. 1–9, Jan. 2014.

[23] M. M. Jaber, S. K. Abd, P. M. Shakeel, M. A. Burhanuddin, M. A. Mohammed, and S. Yussof, "A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms," *Measurement*, vol. 162, Oct. 2020, Art. no. 107883.

[24] E. Messner, M. Fediuk, P. Swatek, S. Scheidl, F.-M. Smolle-Jüttner, H. Olschewski, and F. Pernkopf, "Multi-channel lung sound classification with convolutional recurrent neural networks," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103831.

[25] A. Rizal, R. Hidayat, and H. A. Nugroho, "Lung sound classification ussing Hjorth descriptor measurement on wavelet sub-bands," *J. Inf. Process. Syst.*, vol. 15, no. 5, 2019.

[26] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks," in *Proc. IEEE 32nd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2019, pp. 50–55.

[27] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 535–544, Jun. 2020.

[28] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Inf. Sci. Syst.*, vol. 8, no. 1, pp. 1–8, Dec. 2020.

[29] Y. Ma, X. Xu, Q. Yu, Y. Zhang, and G. Wang, "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-ResNet deep learning algorithm," in *Proc. BioCAS*, 2019, pp. 1–4.

[30] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Başar, "Wavelet entropy: A new tool for analysis of short duration brain electrical signals," *J. Neurosci. Methods*, vol. 105, no. 1, pp. 65–75, Jan. 2001.

[31] R. Q. Quiroga, O. A. Rosso, E. Başar, and M. Schürmann, "Wavelet entropy in event-related potentials: A new method shows ordering of EEG oscillations," *Biol. Cybern.*, vol. 84, no. 4, pp. 291–299, Mar. 2001.

[32] J. Yordanova, V. Kolev, O. A. Rosso, M. Schürmann, O. W. Sakowitz, M. Özgören, and E. Basar, "Wavelet entropy analysis of event-related potentials indicates modality-independent theta dominance," *J. Neurosci. Methods*, vol. 117, no. 1, pp. 99–109, May 2002.

[33] B. Natwong, P. Sooraksa, C. Pintavirooj, S. Bunluechokchai, and W. Ussawawongaraya, "Wavelet entropy analysis of the high resolution ECG," in *Proc. 1st IEEE Conf. Ind. Electron. Appl.*, May 2006, pp. 1–4.

[34] F. Meng, Y. Wang, Y. Shi, and H. Zhao, "A kind of integrated serial algorithms for noise reduction and characteristics expanding in respiratory sound," *Int. J. Biol. Sci.*, vol. 15, no. 9, pp. 1921–1932, 2019.

[35] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Academic, 1999.

[36] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[37] V. Vakharia, M. B. Kiran, N. J. Dave, and U. Kagathara, "Feature extraction and classification of machined component texture images using wavelet and artificial intelligence techniques," in *Proc. 8th Int. Conf. Mech. Aerosp. Eng. (ICMAE)*, Jul. 2017, pp. 140–144.

[38] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[39] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

YAN SHI received the Ph.D. degree in mechanical engineering from Beihang University, Beijing, China. He is currently a Professor with the School of Automation Science and Electrical Engineering, Beihang University. His research interests include intelligent medical devices and energy-saving technologies of pneumatic systems.
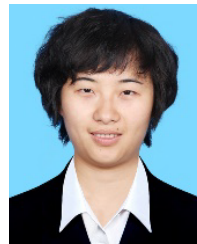
NA WANG received the Ph.D. degree in electric motor and electric apparatus from Beihang University, Beijing, China, in 2014. From 2015 to 2017, she was a Postdoctoral Researcher in mechanical and electronic engineering and a Lecturer with the Engineering Training Center, Beihang University. Her research interests include servo system control and artificial intelligence.

MAOLIN CAI received the Ph.D. degree from the Tokyo Institute of Technology. He is currently a Professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. His research interests include intelligent medical devices, research on the technology of high-efficiency, large-scale compressed air energy storage, and so on.

FEI MENG received the bachelor's degree from the Harbin Institute of Technology, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. His research interests include respiratory sound signal processing and sputum condition recognition.
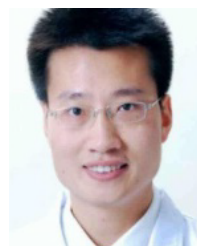
ZUJING LUO received the Ph.D. degree in respiratory medicine. He is currently an Attending Physician with the Department of Respiratory and Critical Care Medicine, Beijing Chaoyang Hospital, Beijing, China. His research interests include mechanical ventilation and respiratory failure.

• • •