Intelligent Automation & Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.041392
Article

Recognition System for Diagnosing Pneumonia and Bronchitis Using Children's Breathing Sounds Based on Transfer Learning

Jianying Shi[1], **Shengchao Chen**[1], **Benguo Yu**[2], **Yi Ren**[3,*], **Guanjun Wang**[1,4,*] **and Chenyang Xue**[5]

[1]School of Information and Communication Engineering, Hainan University, Haikou, 570228, China
[2]School of Biomedical Information and Engineering, Hainan Medical College, Haikou, 571199, China
[3]Department of Pediatrics, Haikou Hospital of the Maternal and Child Health, Haikou, 570203, China
[4]Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, 430074, China
[5]School of Instrument and Electronics, North University of China, Taiyuan, 030051, China
*Corresponding Authors: Yi Ren. Email: renwing1981@163.com; Guanjun Wang. Email: wangguanjun@hainanu.edu.cn

Received: 20 April 2023; Accepted: 08 June 2023; Published: 11 September 2023

Abstract: Respiratory infections in children increase the risk of fatal lung disease, making effective identification and analysis of breath sounds essential. However, most studies have focused on adults ignoring pediatric patients whose lungs are more vulnerable due to an imperfect immune system, and the scarcity of medical data has limited the development of deep learning methods toward reliability and high classification accuracy. In this work, we collected three types of breath sounds from children with normal (120 recordings), bronchitis (120 recordings), and pneumonia (120 recordings) at the posterior chest position using an off-the-shelf 3M electronic stethoscope. Three features were extracted from the wavelet denoised signal: spectrogram, mel-frequency cepstral coefficients (MFCCs), and Delta MFCCs. The recognition model is based on transfer learning techniques and combines fine-tuned MobileNetV2 and modified ResNet50 to classify breath sounds, along with software for displaying analysis results. Extensive experiments on a real dataset demonstrate the effectiveness and superior performance of the proposed model, with average accuracy, precision, recall, specificity and F1 scores of 97.96%, 97.83%, 97.89%, 98.89% and 0.98, respectively, achieving superior performance with a small dataset. The proposed detection system, with a high-performance model and software, can help parents perform lung screening at home and also has the potential for a vast screening of children for lung disease.

Keywords: Deep learning; breath sounds; transfer learning; signal denoising

1 Introduction

Children's lungs are more vulnerable to disease because of their weaker immune systems than adults and lack of self-protection. They depend on their parents for help with prevention and treat-

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ment, and lung disease that is not detected and treated in time can have serious health consequences for children, especially during respiratory pandemics, which can have a sudden impact on life [1]. For this reason, the prevention and treatment of lung disease have received much attention. Early detection and treatment of lung disease can lead to more timely and effective care and reduce the likelihood of emergencies. The simple, non-invasive, low-cost stethoscope-based diagnostic technique provides valuable clinical information about the heart, lungs, and airways [2], however, the diagnosis may be disturbed by various factors such as environmental noise [3], besides, diagnosis is subjectivity due to the physician's expertise to recognize sounds [4]. Therefore, the development of computerized lung sound analysis systems has been extensively researched. The analysis system uses an electronic stethoscope to record lung sounds and applies a deep learning algorithm to classify the recorded lung sounds, which helps overcome the limitations of traditional auscultation and improves the quality of health monitoring [5]. However, large datasets of children's breath sounds are not available to satisfy the large amount of data required by deep learning models. To solve this tricky problem, transfer learning can be applied and it is possible to train with fewer datasets and require less computational cost, solving the problem of insufficient data and the problem of long training times.

Researchers have turned their attention to designing a cost-effective breath sound recognition model to not rely on real-time expert experience. Mining the time-frequency characteristics from the breath sound signal is usually used to explore the relationship between lung condition and breath sounds [6–8]. As an advanced machine learning technique, deep learning methods use deep neural networks to learn hierarchical features from low to high from raw input data [9], enabling the analysis of audio and images with superb predictive accuracy [10]. For the automatic processing of breath sounds, several algorithms have been proposed. Examples include the convolutional neural network (CNN) model to recognize types of breath sound data [11,12], and MFCC to analyze breath sounds [13].

Despite the progress made in previous studies, most studies of breath sounds have focused on adults and neglected children. Due to the significant difference between the breath sounds of children and adults, the results obtained from the adult dataset are likely to misdiagnose childhood diseases.

Based on the above problems, this paper aims to propose an intuitive recognition system for lung diseases by using transfer learning in children. The major contributions of this paper are as follows:

1. A database of children's breath sounds has been established. Reliable breath sounds of different types of children have been recorded in the hospital by specialist doctors.
2. A model with solid generalization ability for children's breath sound recognition within a small dataset is designed, and introduce a novel feature engineering strategy that extracts and fuses time and frequency information from the original breath sound signal at the same time to boost the recognition accuracy.
3. Software written in Python can intuitively present the recognition results of breath sounds recorded by the 3M™ Littmann Digital Stethoscope.

This paper presents a breath sound recognition model based on a transfer learning strategy for the early diagnosis and prevention of lung diseases in children. Experiments have shown that the proposed model can achieve excellent breath sound recognition performance through cross-domain knowledge transfer rather than training a complex weighted deep learning model from scratch. This proposed model could provide a novel and efficient diagnostic platform for computer-assisted breath sound clinics. The paper is organized as follows. Section 2 presents the literature review. Section 3 describes the dataset. Section 4 presents the proposed method. Section 5 is about the experiment. Section 6 puts forward the results obtained from this research. And the conclusions and suggestions for future research are finally discussed in Section 7.

2 Related Works

There have been many studies on breath sound classification systems. In [14], the disease-related relevant features of the lung sound signals are identified in terms of the statistical distribution parameters: the mean, the variance, the skewness, and the kurtosis. The feature set is fed to the classifier model to identify the corresponding classes. The significance of the developed features is validated by conducting several experiments using supervised and unsupervised classifiers. The experimental result is evaluated by statistical analysis. The developed method shows better results compared to the baseline methods and achieves a higher accuracy of 94.16%, sensitivity of 100% and specificity of 93.75% for an artificial neural network classifier. Islam et al. [15] proposed artificial neural network (ANN) and support vector machine (SVM) classifiers together for the classification of normal and asthmatic patients using multi-channel signals. The performance of the combined channels was found to be better than that of the individual channels. The best classification accuracy was 89.2% and 93.3% for the 2-channel and 3-channel combinations in the ANN and SVM classifiers respectively. The channel combination studies show the contribution of each lung sound acquisition region and their combination in asthma detection.

Acharya et al. [16] proposed a deep CNN-RNN model that classifies breath sounds based on mel spectrograms to identify breath sound anomalies for automated diagnosis of respiratory and pulmonary diseases. In addition, the local logarithmic quantization of the training weights is shown to significantly reduce memory requirements, and this type of patient-specific retraining strategy may be useful in the development of reliable long-term automated patient monitoring systems.

Shivapathy et al. [17] used Ensemble Experimental Mode Decomposition for noise reduction and the CNN-RNN model for classification. The classification models are used to classify anomalies in breath sounds such as wheezing and crackling. The data received by the acquisition is denoised using Ensemble Empirical Mode Decomposition. The features of the breathing sound are extracted and sent to the CNN-RNN model for training in order to classify them. The proposed classification model achieves an accuracy of 0.98, a sensitivity of 0.96 and a specificity of 1 for predicting the four classes. Gupta et al. [18] proposed a pre-processing technique that denoises respiratory sounds using the variational mode decomposition (VMD) technique. Different transfer learning models based on deep convolutional neural network architecture are used to classify. Since CNN model over-fit when the dataset size is small, transfer learning have been used for sound classification. The method can classify breath sounds into three classes with accuracy, precision, sensitivity and specificity of 98.8%, 97.7%, 100% and 97.6%, respectively. In [19], Stasiakiewicz et al. developed a classification system using wavelet packets, a genetic algorithm and a support vector machine (SVM) to identify healthy patients and patients with crackles. The system is designed and tested on a dataset consisting of healthy and sick patients with a sensitivity of about 95% and a specificity of 91%.

Haider et al. [20] presented a computerized method for the classification of asthma and chronic obstructive pulmonary disease (COPD) cases based on breath sounds. Empirical mode decomposition is used to denoise the breath sounds. To classify normal, COPD and asthma, different classifiers such as support vector machine (SVM), decision tree (DT), k-nearest neighbor (KNN) and discriminant analysis (DA) are used. In the study, DT classifier was used to discriminate between normal, asthma and COPD cases with a remarkable classification accuracy of 99.3%. In [21], the authors explores the use of deep learning techniques for the classification of lung sounds acquired from patients suffering from connective tissue diseases. A pre-processing pipeline for denoising and data augmentation is developed. Different convolutional neural networks achieved a high overall accuracy of 91% for the classification of pulmonary sounds. The algorithms can be easily supported by modern high-performance edge computing hardware. This study has significant implications for a large-scale screening campaign for interstitial lung disease in the elderly.

Table 1 briefly summarizes the sound types, methods and results of the authors' studies.

Table 1: Summary of reported work in the literature related to lung sound signal analysis

Mondal et al., 2018 [14] • Wheeze, crackle and normal • ANN • The method accuracy is 94.16%.
Islam et al., 2018 [15] • Normal and asthma • ANN, SVM • The best method accuracy is 89.2% and 93.3% for the 2-channel and 3-channel combinations in the ANN and SVM classifiers, respectively.
Acharya et al., 2020 [16] • Crackles, wheezes, a combination of them, and no adventitious respiratory sounds. • CNN-RNN • The proposed hybrid CNN-RNN model achieves a score of 66.31% for the four-class classification of breathing cycles for the ICBHI'17 scientific challenge respiratory sound database.
Shivapathy et al., 2021 [17] • Wheeze, crackle and normal • CNN, RNN • The model scores an accuracy of 0.98, sensitivity of 0.96 and specificity of 1.
Gupta et al., 2021 [18] • Wheeze, crackle and normal • CNN • The proposed method can classify sounds with accuracy, precision, sensitivity and specificity of 98.8%, 97.7%,100% and 97.6%.
Stasiakiewicz et al., 2021 [19] • Crackle and normal • SVM • The system has a sensitivity of approximately 95% and a specificity of 91%.
Haider et al., 2022 [20] • Asthma, COPD and normal • Decision tree • The method accuracy is 99.3%

(Continued)

Table 1 (continued)

Dianat et al., 2023 [21]

- Sounds acquired from patients affected by connective tissue diseases
- CNN
- Various convolutional neural networks have provided an overall accuracy as high as 91% in the classification of lung sounds

3 Methodology

Since there is no large database involving children's breath sounds in the public dataset, the research is carried out by using a 3M™ Littmann® 3200 electronic stethoscope, cooperating with Haikou Hospital of the Maternal and Child Health. The stethoscope is used in outpatient clinics to record a child's breathing sounds with a sampling time of 15 s, and only the signal is noted with its corresponding disease, without retaining any information about the patient, protecting the patient's privacy and ensuring the anonymity of the data. The data are collected by the doctors, which will ensure their reliability. And the pediatric expert advises that the sampling position should avoid the interference of heart sounds as much as possible. Besides, the acquisition environment should be quiet and the stethoscope should be placed in the corresponding stethoscope area.

The collected data can be divided into three classes: normal breath sound signals, pneumonia breath sound signals, and bronchitis signals.

The details of the proposed recognition model of children's breath sound signals are introduced. The signal processing steps are shown in Fig. 1, including preprocessing, and combining models through soft voting by using Fine-tuned MobileNetV2 and ResNet50 with random forest.

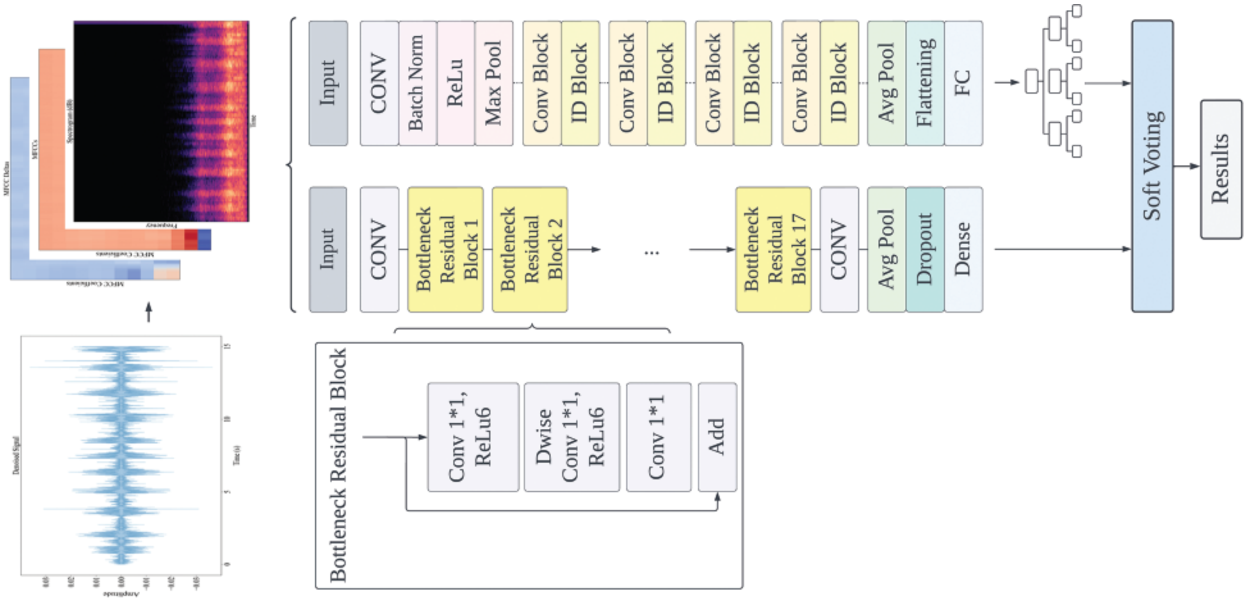


Figure 1: System diagram of the proposed method

3.1 Data Preprocessing

The experimental data are divided into three types of signals: normal breath sound signal, pneumonia breath sound signal, and bronchitis signal. In order to balance the experimental data, the number of randomly selected data of each type is consistent. The training set has 84 signals in each category, 252 signals in total; The test set has 36 signals of each type, a total of 108 signals.

The waveforms of normal breath sound, pneumonia, and bronchitis are shown in Fig. 2. It can be seen from the waveform that the signals are periodic, and the period of three kinds of heart sound signals is obvious. The breath sounds contain a large amount of physiological and pathological information in the lungs, and each pathological abnormal breath sound corresponds to a patient who may have some lung disease.

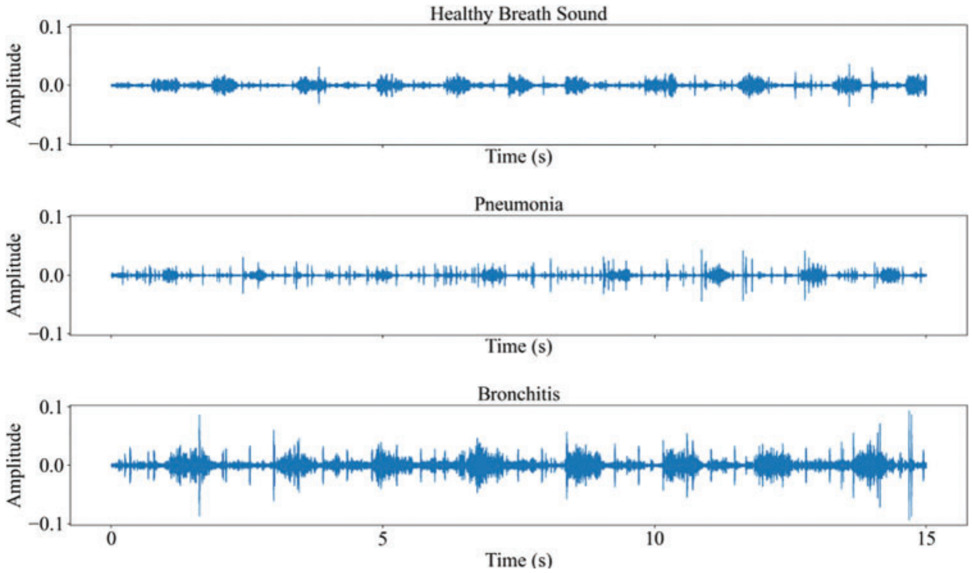


Figure 2: The waveforms of breath sound signals

In the process of breath sound signal acquisition, in addition to heart sound, equipment noise, and hospital environment noise will also be mixed into the recording. Without a high-quality signal, the subsequent signal analysis will be affected, so it is necessary to preprocess the collected breath sound signal. The frequency of breath sound is within the range of [50–2500 Hz] [22]. Considering the breath sound rate, the fourth-order Butterworth bandpass filter [23] with frequencies of 50 and 2500 Hz is used to filter the breath sound. This interval ensures that the main components of these sounds are retained, while high-frequency noise and low-frequency baseline fluctuations are reduced. The waveform of the original breath sound signal is shown in Fig. 3, and the filtering effect is shown in Fig. 4.

The spectrum of the raw breath sound signal is shown in Fig. 5, and the filtering signal is shown in Fig. 6.

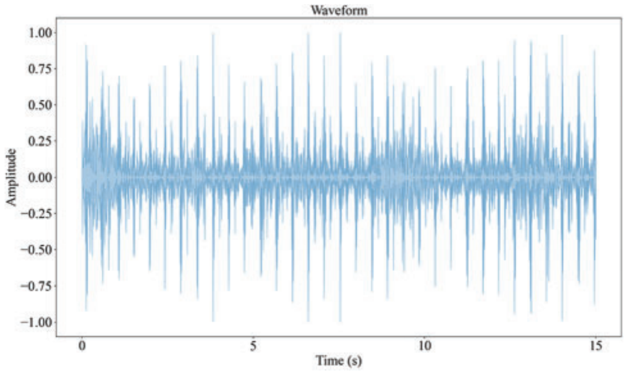


Figure 3: Raw signal

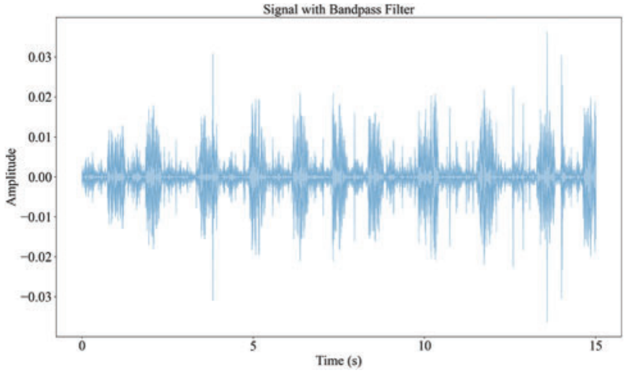


Figure 4: Signal with bandpass filter

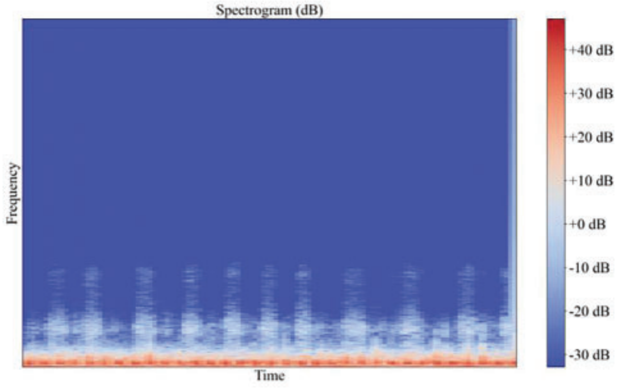


Figure 5: The spectrum of the raw breath sound signal

From the above two pictures, it can be seen that after high-pass filtering, the noise signal has significantly weakened in the low-frequency region. The change before and after filtering can be clearly seen in the waveform and spectrum. If there is no filtering processing, the noise with a high energy value will greatly affect the recognition effect of the relatively weak lung sound signal.

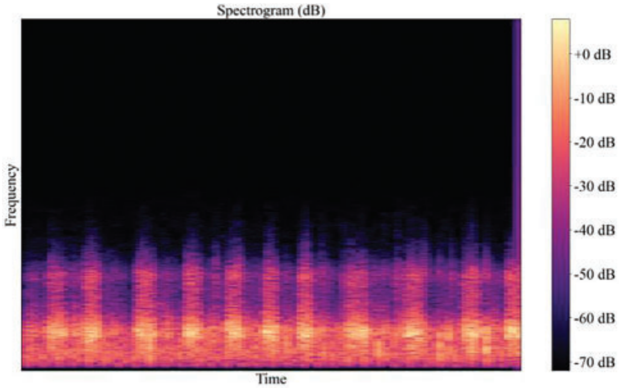


Figure 6: The spectrum of the filtered breath sound signal

3.2 Denoise

The breath sound signal contains important information about time and frequency. If only a bandpass filter is used and the frequency part of the signal is analyzed, the information about the time of each frequency occurrence will be ignored. This problem can be solved by wavelet transform. Wavelet transform can adapt to the requirements of time-frequency signal analysis more automatically and solve the problems that Fourier transform cannot when processing non-stationary signals [24].

Discrete wavelet transform (DWT) is a linear transformation method, which operates on coefficient vectors of an integer power of 2 in length and converts them into vectors with different values of the same length [25].

Assume that $s(n)$ is the original signal and the frequency range is from 0 to π rad/s. DWT of time domain signal $s(n)$ is defined as:

$$W_x(a,b) = \sum\nolimits_n \frac{1}{\sqrt{a}} s(n)\, \psi^*\left(\frac{n-b}{a}\right) \tag{1}$$

The most appropriate way to view the noise effect added to the signal is to add white noise. After denoising with different mother wavelet and decomposition levels, the performance can be measured by comparing the denoised signal with the raw signal.

3.3 Feature Extraction

In order to adapt to the format of the input into the model, the features obtained from the signal are processed. Because the input signal of the image classification model is in RGB format, a three-dimensional array (width, height, channels), which means there are 3 channels, representing the three RGB color channels. Therefore, the features are combined by using three diagrams, as shown in Fig. 7. The feature calculation method will be described in detail below.

3.3.1 Spectrogram

Spectrum is a visual representation of the spectrum of the signal changing with time. The spectrum diagram can be generated by optical spectrometer, a group of band-pass filters and a transform [26].

Short-Time Fourier Transform (STFT) of the signal can be obtained by adding windows to the signal and performing discrete Fourier transform (DFT) on each window.

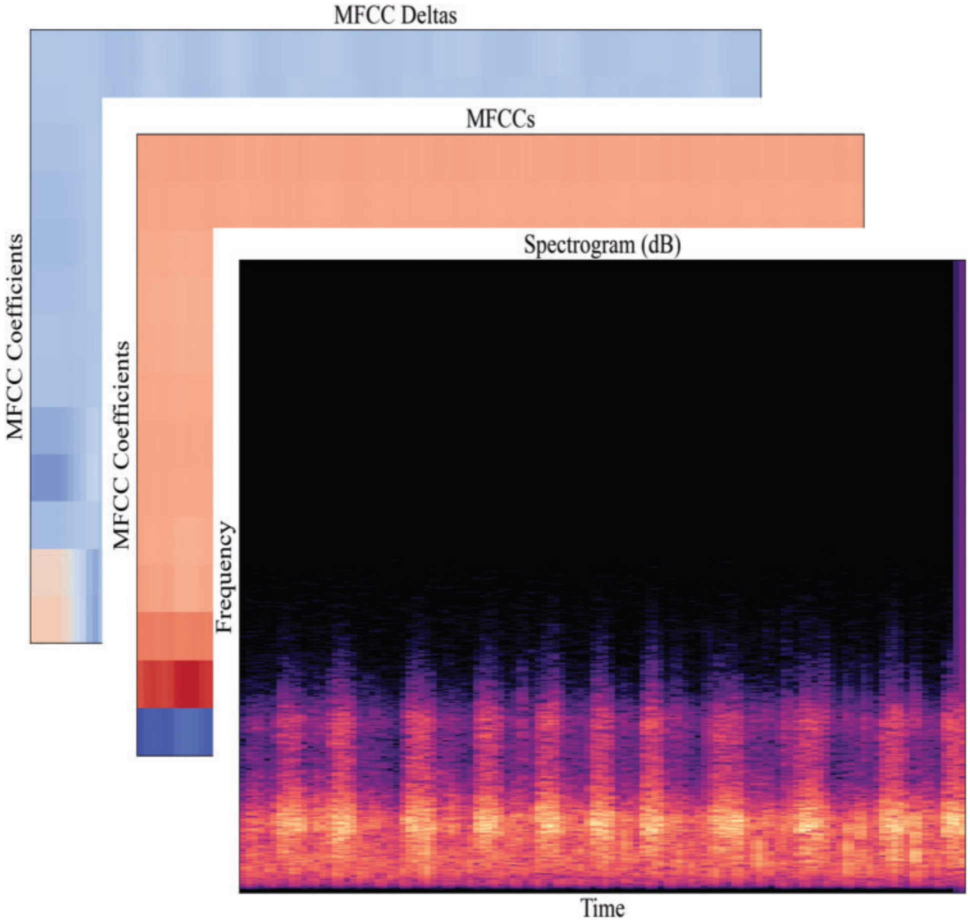


Figure 7: Features to input

The short-time Fourier transform formula is:

$$STFT\{x_n\}(h,k) = X(h,k) = \sum_{n=0}^{N-1} x_{n+h}\omega_n e^{-i2\pi\frac{kn}{N}} \tag{2}$$

where x_n is the input signal, ω_n is the window. STFT describes the evolution of frequency components with time.

3.3.2 MFCCs

Mel-Frequency Cepstrum is a linear transformation of the logarithmic energy spectrum based on the nonlinear mel scale of sound frequency. MFCCs are the coefficients that make up the Mel-frequency cepstral. It is derived from the cepstrum of an audio segment. The conversion formula from Hertz to Mel scale is as follows [27]:

$$m = 2595log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

where f is the original frequency in Hz.

3.3.3 Delta MFCCs

The MFCCs feature vector only describes the power spectrum envelope of a single frame, but the signal has dynamic information. Delta coefficients are used to identify the dynamics of the signal power spectrum. The delta coefficients are computed using the following formula.

$$d_t = \frac{\sum_{n=1}^{v_1} n(c_{t+n} - c_{t-n})}{2\Sigma_{n=1}^{N} n^2} \tag{4}$$

where d_t is a delta coefficient from frame t computed in terms of the static coefficients c_{t-n} to c_{t+n}. n is usually taken to be 2.

3.4 Transfer Learning Models for Classification

3.4.1 Fine-Tuned MobileNetV2

MobileNetV2, a convolutional neural network, builds upon the ideas from MobileNetV1 [28], using depthwise separable convolution as efficient building blocks. The MobileNetV2 models are much faster in comparison to MobileNetV1, it is a very effective target detection and segmentation feature extractor.

Therefore, MobileNetV2 is used as the basic model and the final output layer is changed to the one that meets the classification needs. A Dropout layer is added before the classification layer for regularization, and the input data is set to adjust the image size to 120 × 120. Freezing all layers in the network except the classification layer prevents the weights in those layers from being re-initialized. The next step is to add a new trainable layer to transform the old features into the prediction of the new data set. The structure of the improved MobileNetV2 is shown in Fig. 8.

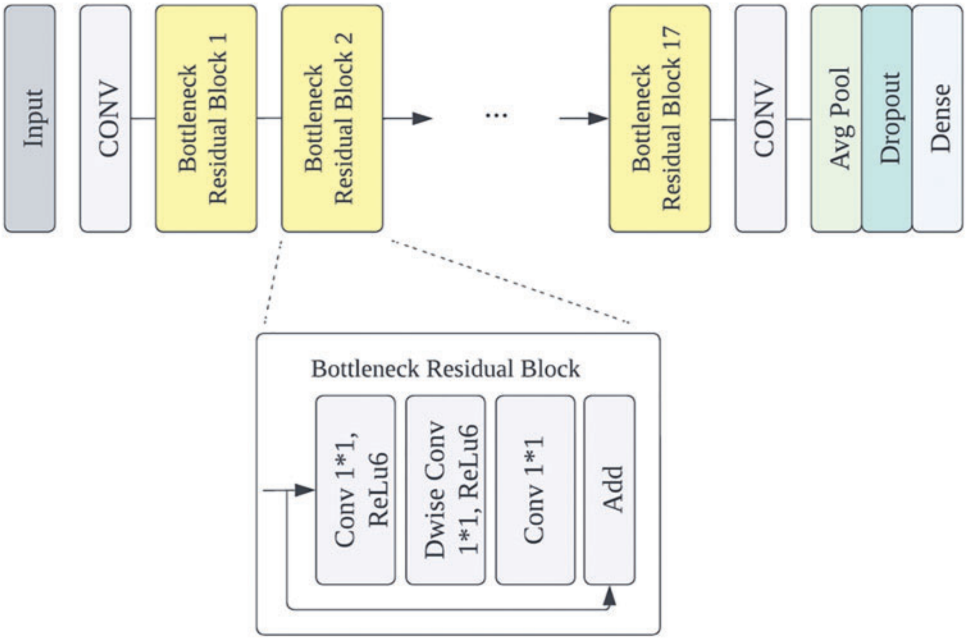

Input
CONV
Bottleneck Residual Block 1
Bottleneck Residual Block 2
...
Bottleneck Residual Block 17
CONV
Avg Pool
Dropout
Dense
Bottleneck Residual Block
Conv 1*1, ReLu6
Dwise Conv 1*1, ReLu6
Conv 1*1
Add

Figure 8: Model structure diagram

After training, fine-tuning, which is used to freeze a few network layers for feature extraction and jointly train the unfrozen layer of the pre-training model and the newly added classifier layer, is used to improve the performance of the model. Unfreezing the base model and training the entire model end-to-end with a low learning rate is the last step. Fine-tuning allows the past knowledge in the target field and re-learn new knowledge to be applied to the model. A low learning rate will improve the performance of the model on the new data set while preventing over-fitting.

ImageNet is a large visual database designed for visual object recognition software research. The pre-trained ImageNet weights are used for the MobileNetV2.

3.4.2 ResNet50 with Random Forest

ResNet50 deep learning model is regarded as the pre-trained model for feature extraction in transfer learning, and then the random forest is used to classify. Resnet50 is characterized by deep structure and residual connectivity, which makes it one of the best feature extractor options [29].

In this paper, the ResNet50 model is employed to preprocess the image and obtain the basic features, and then these features are transferred to the random forest, that is, combined with the traditional classifier to classify the data without retraining the basic model. Random forest is a supervised learning algorithm that combines the output of multiple decision trees to achieve a single result, which can be used for classification and regression. Pre-trained convolutional neural networks have important characteristics for signal classification. The model structure is shown in Fig. 9.

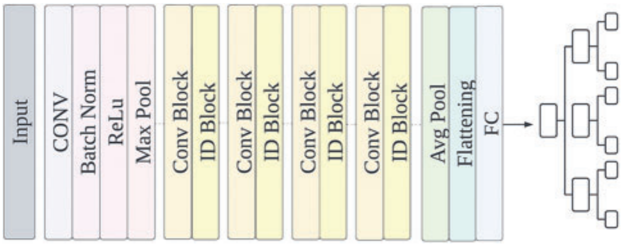


Figure 9: Modified ResNet50 model structure

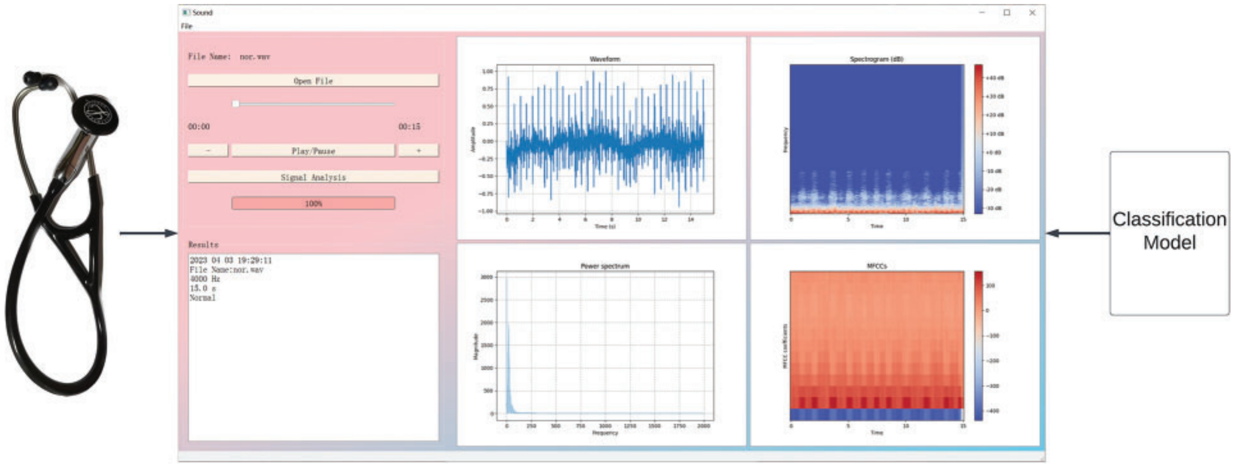
3.4.3 Soft Voting

A voting classifier is an ensemble classifier whose input is two or more estimators that combine models from different classification algorithms with individual weights to obtain confidence and classify data based on majority votes. Voting classifiers, which are built by combining different classification models, perform better and can compensate for the weaknesses of individual classifiers on a given dataset.

Soft voting was chosen as it gives more weight to those models with high probabilities and performs better than hard voting.

3.5 Breath Sounds Disease Recognition System

With an off-the-shelf 3M electronic stethoscope, software was designed to help doctors and patients analyze the breath sounds recorded by the stethoscope. The software, with the proposed classification model, can display information about the recorded breath sounds and the recognition results. The overall structure is shown in Fig. 10.



File Name: nor.wav

```
2023 04 03 19:29:11
File Name:nor.wav
4000 Hz
15.0 s
Normal
```

Figure 10: Breath sounds disease recognition system structure

A sample of a normal breath sound is shown in Fig. 10. Once the audio file of the breath sound signal is opened in the system and the "Signal Analysis" button is clicked, the program will process the input audio signal and convert it into spectrograms, which is the same as the training set, and the extracted features will be classified by the proposed model to get the classification results. The relevant waveform plot of the signal will be displayed on the right side of the application, the analysis time, file name, signal frequency, signal duration, and the diagnosis result appear below. The software also includes functions to play sounds and control sound volume.

In conclusion, normal, pneumonia and bronchitis breath sounds were recorded using a 3M electronic stethoscope. For the recognition and prevention of lung diseases in children, effective processing and analysis of the breath sounds collected by the electronic stethoscope is crucial. In addition, a data set of children's breath sounds could also contribute to the development of related research. Moreover, the combination of algorithms and software makes the system reliable, safe, and intuitive to use. Supported by advanced artificial intelligence technology, the system can provide efficient and comprehensive disease monitoring for children, and doctors and parents can easily access the highest quality diagnostic capabilities.

4 Experiment

4.1 Performance Evaluation Criteria

Signal-to-noise ratio (SNR), Fit, and Correlation coefficient criteria were used to evaluate the performance of the denoising method. The SNR formula is as follows:

$$SNR\left(x_d, y\right) = 10\log_{10}\frac{\sum_i^N x_d\left(i\right)^2}{\Sigma_i^N\left(y\left(i\right) - x_d\left(i\right)\right)^2} \tag{5}$$

where N is the number of signal samples, x_d is the desired signal, and y is the denoised signal.

Fit can obtain basic information between the desired signal and the denoised signal and ensures that important information about the signal is not lost during the denoising process. The Fit formula is as follows [30]:

$$Fit = 100 \times \left(1 - \frac{\Sigma_i^N \left(y(i) - x_d\left(i\right) \right)^2}{\Sigma_i^N \left(x_d\left(i\right) - \frac{1}{N} \Sigma_i^N x_d\left(i\right) \right)^2} \right) \tag{6}$$

The correlation coefficient between two variables means predicting the value of one in relation to the other. The correlation coefficient formula can be seen below:

$$corr\left(x_d, y\right) = \left(\frac{cov\left(x_d, y\right)}{\sigma_{x_d}\sigma_y} \right) = \left(\frac{E\left[\left(x_d - \mu_{x_d}\right)\left(y - \mu_y\right) \right]}{\sigma_{x_d}\sigma_y} \right) \tag{7}$$

The performance evaluation metrics included accuracy, precision, recall, specificity, and F1 score. The confusion matrix-based definitions for each of these metrics are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Specificity = \frac{TN}{FP + TN} \tag{11}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

where the true positives (TP) and the true negatives (TN) represent the amount of the correctly classified audio signals, while the false positives (FP) and false negatives (FN) represent the number of the incorrectly classified signals.

4.2 Setup

After the breath sound signal passes through the band-pass filter, the white noise of different decibels is added to determine the wavelet function with the best filtering effect. The breath sound signal after denoising obtains the spectrogram, MFCCs feature map, and first-order difference MFCCs feature map, and the image is entered into the MobileNetV2 and ResNet50 networks through three channels with the size of [120, 120]. The weight parameters are fine-tuned in the MobileNetV2 network; the weight parameters of ResNet50 remain unchanged for feature extraction and are combined with random forest. Finally, the class probability vectors of the two classifiers are fused by the soft voting algorithm to get the classification result.

5 Results and Discussions

Mother wavelet, similar to breath sound signals for detection [31–33], such as Daubechies (Db) wavelet family, Symlets (Sym) wavelet family and Coiflets (Coif) wavelet family is selected. The tested signal is polluted by white noise with an SNR of 5 dB as an initial value to test the performance of the proposed denoising technique. For each layer decomposition, hard-thresholding and soft-thresholding are used to analyze the denoising performance of the resulting breath sound signal. Table 2 shows the SNR results of wavelet decomposition layers ranging from 2 to 9, using hard threshold and soft threshold.

Table 2: Performance parameters

No. levels	Coif2						Coif5					
	Soft			Hard			Soft			Hard		
	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)
2	11.02	92.03	96.26	10.97	91.96	96.18	11.05	92.11	96.28	11.02	92.04	96.25
3	14.04	96.03	98.08	13.82	95.82	97.96	14.06	96.05	98.09	14.00	95.99	98.06
4	16.57	97.79	98.91	16.03	97.48	98.75	16.62	97.81	98.91	16.62	97.81	98.92
5	17.49	98.20	99.11	16.10	97.52	98.77	17.80	98.33	98.92	16.69	97.84	98.93
6	**17.57**	**98.23**	**99.12**	16.10	97.52	98.77	**17.88**	**98.36**	**99.17**	16.70	97.85	98.94
7	17.57	98.23	99.12	16.10	97.52	98.77	17.88	98.36	99.18	16.70	97.85	98.94
8	17.57	98.23	99.12	16.10	97.52	98.77	17.89	98.36	99.18	16.70	97.85	98.94
9	17.57	98.23	99.12	16.10	97.52	98.77	17.89	98.36	99.18	16.70	97.85	98.94
No. levels	Db11						Db14					
	Soft			Hard			Soft			Hard		
	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)
2	11.03	92.05	96.24	11.02	92.04	96.23	10.96	91.93	96.21	10.99	91.99	96.24
3	14.00	95.99	98.04	14.02	96.01	98.05	13.90	95.91	98.02	13.96	95.95	98.04
4	16.61	97.80	98.91	16.65	97.82	98.92	16.62	97.81	98.92	16.47	97.73	98.88
5	17.87	98.35	99.17	16.73	97.86	98.94	17.77	98.32	99.16	16.59	97.79	98.91
6	17.89	98.36	99.18	16.73	97.86	98.94	17.80	98.33	99.17	16.58	97.79	98.91
7	**17.90**	**98.37**	**99.18**	16.73	97.86	98.94	**17.81**	**98.33**	**99.17**	16.58	97.79	98.91
8	17.90	98.37	99.18	16.73	97.86	98.94	17.81	98.33	99.17	16.58	97.79	98.91
9	17.90	98.37	99.18	16.73	97.86	98.94	17.81	98.33	98.16	16.59	97.79	98.91
No. levels	Db4						Sym4					
	Soft			Hard			Soft			Hard		
	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)
2	11.06	92.11	96.26	10.98	91.97	96.20	11.04	92.07	96.26	10.97	91.95	96.21
3	14.06	96.04	98.08	13.95	95.94	98.03	13.98	95.97	98.04	13.97	95.96	98.04
4	16.62	97.81	98.91	16.43	97.71	98.86	16.58	97.78	98.91	16.40	97.69	98.86
5	17.56	98.23	99.12	16.48	97.73	98.88	17.52	98.22	99.11	16.44	97.71	98.87
6	**17.61**	**98.25**	**99.13**	16.48	97.74	98.88	**17.57**	**98.24**	**99.12**	16.44	97.71	98.87
7	17.61	98.25	99.13	16.48	97.74	98.88	17.57	98.24	99.12	16.44	97.71	98.87
8	17.61	98.25	99.13	16.48	97.74	98.88	17.57	98.24	99.12	16.44	97.71	98.87
9	17.61	98.25	99.13	16.48	97.74	98.88	17.57	98.24	99.12	16.44	97.71	98.87

According to Table 2, the degree of decomposition and the type of threshold are important parameters that influence the SNR when a wavelet family is selected. Fig. 11 shows a histogram of the SNR values obtained when comparing different wavelet families with soft thresholding. It can be seen from the figure that the Coif5, Db11 and Db14 perform better.

The 6th layer decomposition of Coif5 wavelet adopts soft threshold and has the highest SNR of 17.88 dB, Fit of 98.36%, Corr of 99.17%. The 7th layer decomposition of Db11 wavelet uses soft threshold and has the highest SNR, 17.90 dB with Fit being 98.37% and Corr being 99.18%. The 7th layer decomposition of Db14 wavelet uses soft threshold and has the highest SNR, which is 17.81 dB. Fit is 98.33%. Corr is 99.17%.

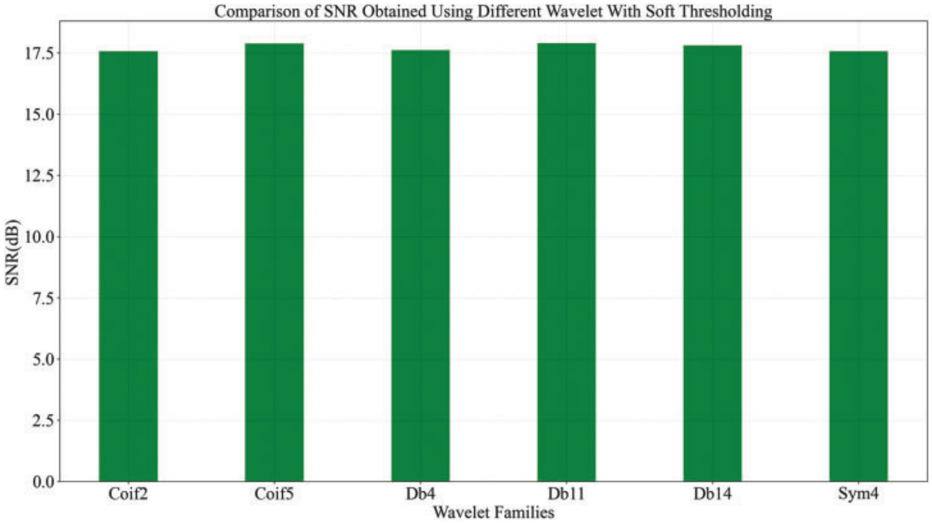


Figure 11: Comparison of SNR obtained using different wavelet with soft thresholding

As shown in Fig. 12, the first row is the raw signal diagram, the signal diagram after adding noise, and the diagram of the 6-layer denoising signal decomposed by Coif5 wavelet; the second row is the raw signal diagram, the signal diagram after adding noise, and the 7-layer denoising signal decomposed by Db11 wavelet; the third row is the original signal diagram, the noise-added signal diagram, and the Db14 wavelet decomposition 7-layer denoising signal diagram.

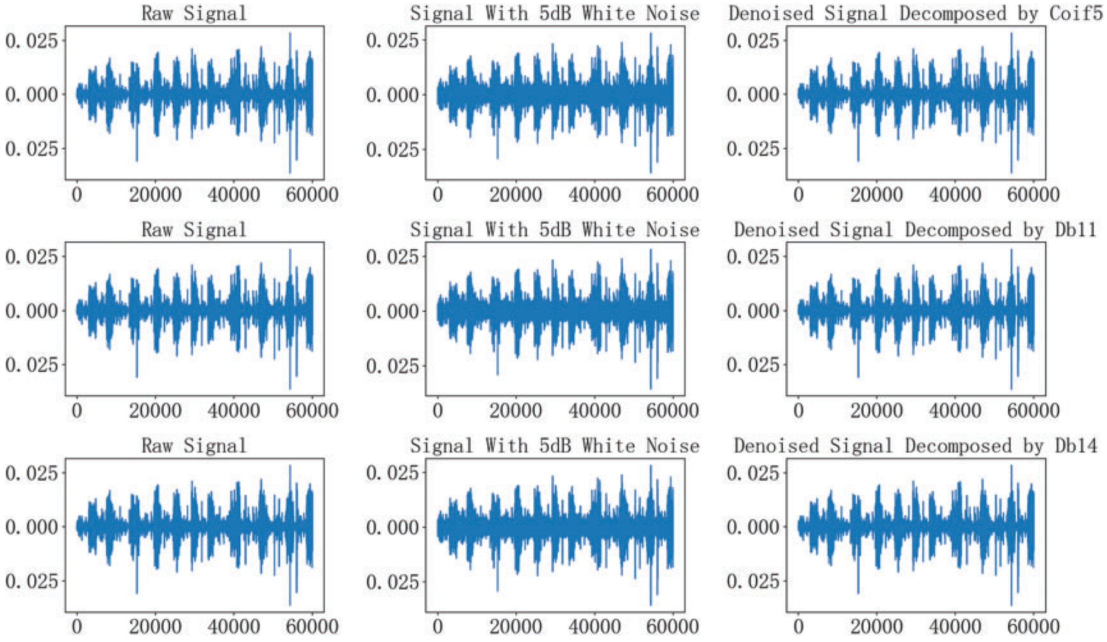


Figure 12: The raw signal, signal with noise, denoised signal by Coif5, Db11 and Db14

The results in Table 3 show that Db14 outperforms other mother wavelets when adding white noise pollution with SNR of 10, 15 and 20 dB.

Table 3: Signal denoising by different mother wavelets

Mother wavelets	SNR = 10			SNR = 15			SNR = 20		
	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)	SNR (dB)	Fit (%)	Corr (%)
Coif5	21.51	99.28	99.64	25.27	99.70	99.85	28.92	99.87	99.94
Db11	21.41	99.27	99.64	25.19	99.69	99.84	28.91	99.87	99.93
Db14	**21.53**	**99.29**	**99.65**	**25.29**	**99.71**	**99.85**	**28.93**	**99.87**	**99.94**

From the above experimental results, it can be seen that Db11 wavelet and Db14 wavelet have the best performance in breath sound denoising. The difference in noise removal performance between Db11 and Db14 can be ignored.

Based on the collected data, we used EfficientNet [34], VGG16 [35], RNN [36], MobileNetV2 [37] and ResNet50 [38] models to classify the breath signals. Table 4 summarizes the performance of the different classifiers. The experimental results show that MobileNetV2 and ResNet50 have better performance.

Table 4: Performance evaluation metrics of different classifiers

Classifier	Accuracy	Precision	Recall	Specificity	F1 score
EfficientNet	85.81%	91.01%	65.03%	91.01%	0.78
VGG16	89.81%	92.02%	68.01%	96.04%	0.82
RNN	88.60%	90.01%	57.11%	97.02%	0.77
MobileNetV2	89.82%	89.89%	89.82%	94.91%	0.90
ResNet50	91.05%	90.09%	89.82%	94.91%	0.90

MobileNetV2 is a lightweight model that has a significantly smaller number of parameters in a deep neural network. ResNet50 is a deep architecture, which makes it more suitable for image recognition. Although ResNet is much deeper than VGG16, the model size is actually much smaller due to the use of global average pooling rather than fully connected layers. Both models are chosen for the subsequent experiments because of their excellent performance, and the experiments aim to modify the structure of these two models.

In order to modify the network structure of MobileNetV2, two sets of comparison experiments were conducted with the same dataset. One experiment is to modify the output layer without modifying the weight parameters, and the other experiment is to fine-tune the MobileNetV2 network. These samples take 120 epochs for the dataset to pass through the model. The experimental parameters use the AdamOptimizer optimization function, the learning rate is set to 0.0001, and the multi-classification cross-entropy loss function is employed.

When only the output layer is modified without modifying the weight parameters, the accuracy rate of the training set is 92.19%, and the accuracy rate of the test set is 89.82%. To illustrate the performance of the model, both accuracy and loss graphs are shown in Fig. 13. Fig. 14 shows the confusion matrix of the proposed model that describes how many signals are correctly classified among the test signals.

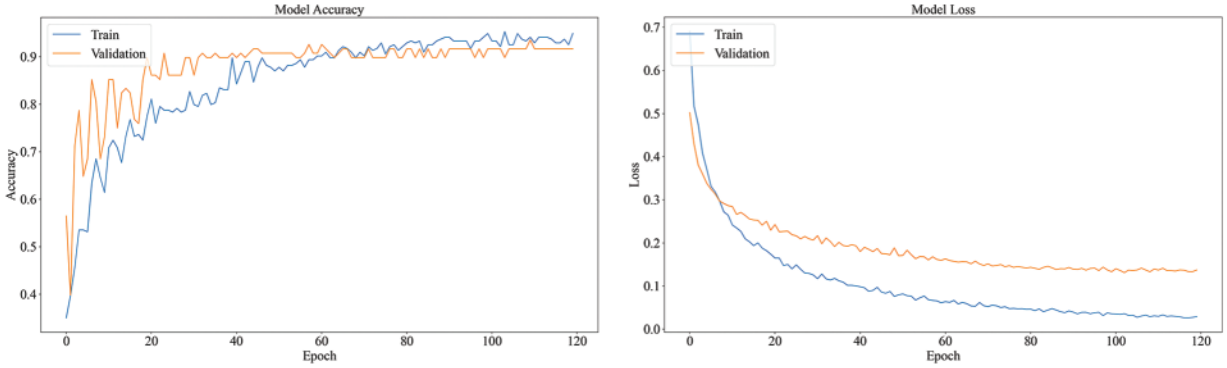


Figure 13: Accuracy and loss graph of the MobileNetV2

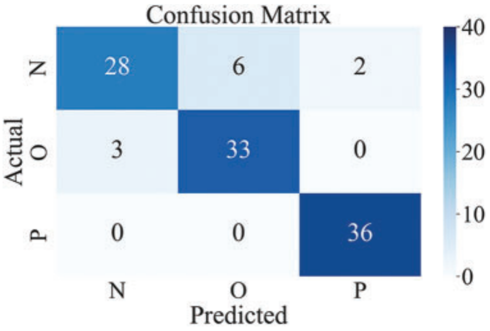


Figure 14: Confusion matrix of the MobileNetV2

The classification performance of the model for each category is shown in Table 5.

Table 5: The precision, recall, specificity and F1 score for the MobileNetV2

	Precision	Recall	Specificity	F1 score
N	90.32%	77.78%	95.83%	0.84
O	84.62%	91.67%	91.67%	0.88
P	94.74%	100%	97.22%	0.97

The model performs well for classifying pneumonia, and recall reaches 100%; However, for normal breath sound recognition, the recall is poor. The specificity of normal breath sounds is lower than the specificity of pneumonia, and the F1 score of normal breath sounds is also the lowest of the three categories. After fine-tuning the MobileNetV2 model, the accuracy rate is better in the training, and the accuracy of the test is 96.01%. Accuracy and loss graphs are shown in Fig. 15. Fig. 16 shows the confusion matrix of the fine-tuning MobileNetV2.

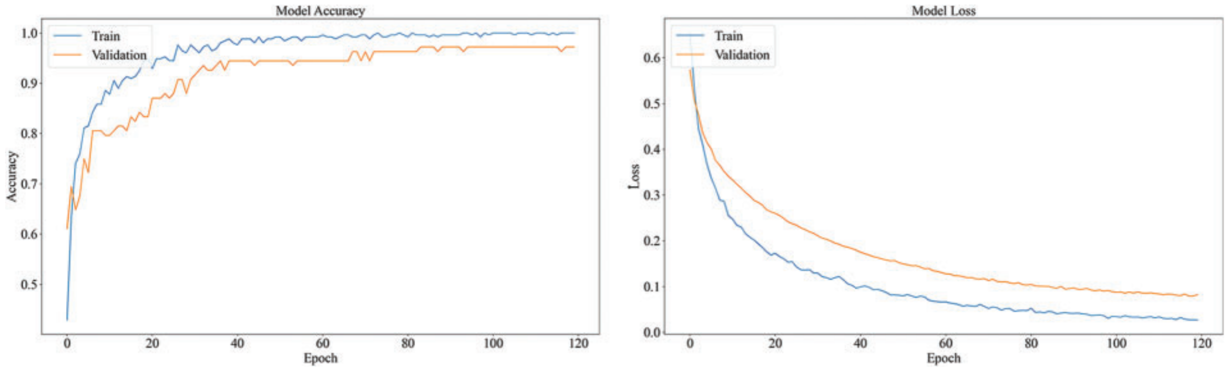


Figure 15: Accuracy and loss graph of the fine-tuning MobileNetV2

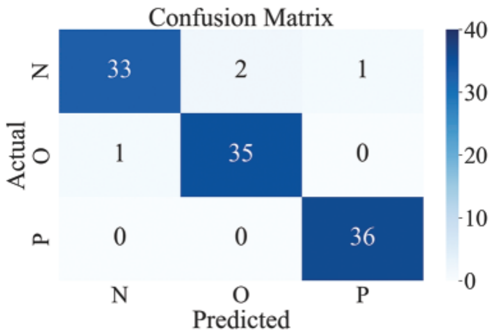


Figure 16: Confusion matrix of the fine-tuned MobileNetV2

The classification performance of the model for each category is shown in Table 6.

Table 6: The precision, recall, specificity and F1 score for fine-tuning MobileNetV2

	Precision	Recall	Specificity	F1 score
N	97.06%	91.67%	98.61%	0.94
O	94.60%	97.22%	97.22%	0.96
P	97.30%	100%	98.61%	0.98

The fine-tuned MobileNetV2 model has improved in all aspects. In particular, it greatly improves the recognition effect of normal breathing sounds. Precision has increased by 6.74%, and Recall by 13.89%. The fine-tuned MobileNetV2 model performs better in classifying breath sounds. There is a gap between the loss function curves of the training set and the test set of the model without transfer learning, and the gap between the two loss function curves of the model after transfer learning is reduced, which shows that transfer learning alleviated the overfitting of the model.

For ResNet50, the Softmax function is used in the top layer as the activation function, and the experiment takes 120 epochs for the dataset during training. The experimental parameters use the AdamOptimizer optimization function, the learning rate is set to 0.0001, and the multi-classification cross-entropy loss function is adopted.

The accuracy of ResNet50 in classifying data is 91.05%. Accuracy and loss graphs are shown in Fig. 17. Fig. 18 shows the confusion matrix of the ResNet50.

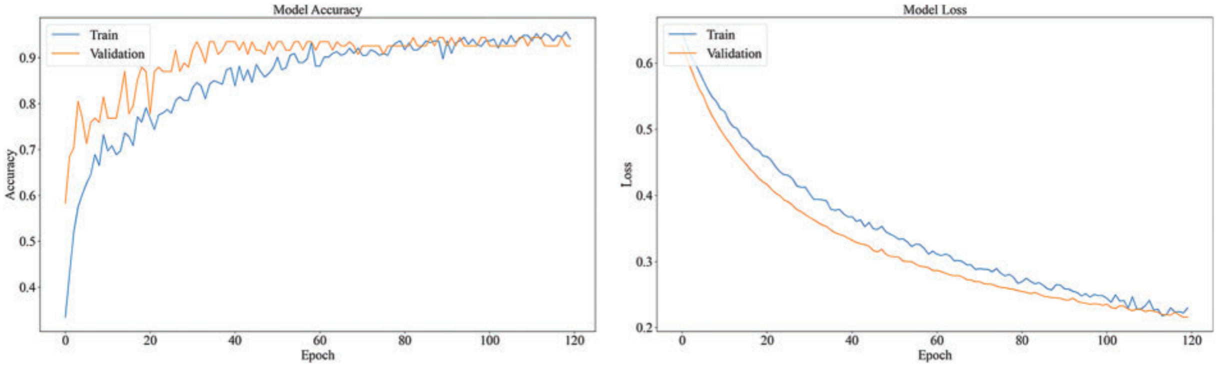


Figure 17: Accuracy and loss graph of the ResNet50

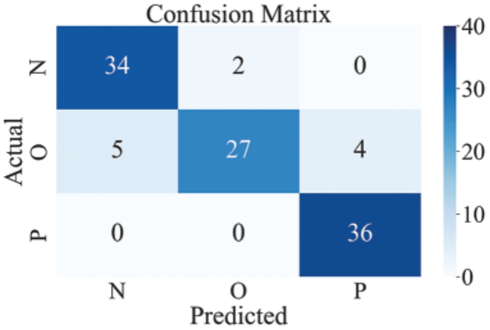


Figure 18: Confusion matrix of the ResNet50

The classification performance of the model for each category is shown in Table 7. The model has a better performance for pneumonia, which is better than the performance of the other two categories.

Table 7: The precision, recall, specificity and F1 score for ResNet50

	Precision	Recall	Specificity	F1 score
N	87.18%	94.44%	93.06%	0.91
O	93.10%	75.00%	97.22%	0.83
P	90.00%	100%	94.44%	0.95

The proposed method uses the feature maps of ResNet50 and passes them to a random forest (ResNetRF) to classify the data, the accuracy is 94.45%. The confusion matrix is shown in Fig. 19. More details about the classification performance of ResNetRF are shown in Table 8.

The ResNet50 is used as a feature extractor to extract features, then we use random forest for classification, as a result, the classification performance has been greatly improved compared with the ResNet50. ResNetRF is better at identifying Pneumonia, precision and specificity have reached 100%, and Recall is 97.22%. The proposed model combines the two models through soft voting.

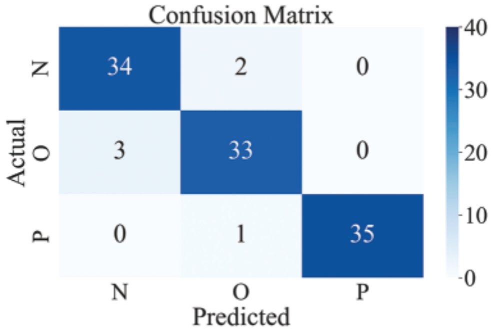


Figure 19: Confusion matrix of the ResNetRF

Table 8: The precision, recall, specificity and F1 score for ResNetRF

	Precision	Recall	Specificity	F1 score
N	91.89%	94.44%	95.83%	0.93
O	91.67%	91.67%	95.83%	0.91
P	100.00%	97%	100.00%	0.98

The classification performance comparison of models is shown in Table 9.

Table 9: The classification performance comparison of models

Classifier	Accuracy	Precision	Recall	Specificity	F1 score
MobileNetV2	89.82%	89.89%	89.82%	94.91%	0.90
ResNet50	91.05%	90.09%	89.82%	94.91%	0.90
Fine-tuning MobileNetV2	96.01%	96.32%	96.30%	98.15%	0.96
ResNetRF	94.45%	94.52%	94.44%	97.22%	0.94
Proposed network	**97.96%**	**97.83%**	**97.89%**	**98.89%**	**0.98**

The proposed model has the best performance, with an accuracy of 97.96%, which is 9.84% higher than MobileNetV2, 2.95% higher than fine-tuning MobileNetV2, 4.82% higher than ResNet50, and 3.51% higher than ResNetRF. Compared with other models, the precision, recall, specificity, and F1 score of the model are also the best. The proposed model is significantly better than single models. Fine-tuning MobileNetV2 is also better than MobileNetV2, which shows that the transfer learning of model parameters can effectively improve the accuracy of breath sound classification. The combination of ResNet50 and Random Forest works better than the ResNet50 model.

In order to verify the influence of noise components on the classification results of breath sounds, the noise components in breath sounds were not removed in the signal preprocessing, and the above experiment was repeated by using the proposed model. The results are shown in Table 10.

Table 10: The classification performance

	Accuracy	Precision	Recall	Specificity	F1 score
With noise	80.90%	66.00%	60.00%	86.00%	0.63
Without noise	97.96%	97.83%	97.89%	98.89%	0.98

From the experimental results, it can be seen that noise has a great influence on the recognition accuracy of the breath sound classification. When the experiment is performed without removing the noise, the classification performance of the breath sound data decreases. Experiments show that noise is an important interference factor in breath sound classification, and using wavelet transformation to remove noise greatly improves classification accuracy.

With the continuous development of artificial intelligence technology, intelligent diagnosis is widely used for providing objective and accurate results. However, most studies have focused on physiological data from adults, while children with weak immune systems have been neglected. And the lack of public datasets on children's breath sounds has limited the development of deep learning studies on children's breath sounds. To solve this problem, we collaborated with a hospital where normal, pneumonia and bronchitis breath sounds of children were collected by doctors. The proposed system with an off-the-shelf stethoscope incorporates a transfer learning-based model which can achieve superior performance with a small dataset, along with software for displaying analysis results. In addition, the system will meet the need for accurate recognition and analysis of children's breath sounds for early diagnosis of lung disease.

Through extensive experimental comparisons of commonly used wavelets, the wavelet suitable for this dataset was selected for denoising. The comparison of the classification results with and without noise demonstrates that noise can reduce the performance of the classification model, therefore denoising is an essential step in this recognition system. The transfer learning technique is suitable for this study since it can train the model with a small amount of data and still achieve a high level of performance, overcoming the limitations imposed by data scarcity. The MobileNetV2 network structure and the ResNet50 network structure have better classification performance for children's breath sounds compared to other network structures. The performance of the model combined with fine-tuned MobileNetV2 and modified random forest improves further for breath sounds recognition after the soft voting method.

The establishment of the dataset will facilitate further research on children's breath sounds. Furthermore, the performance metrics of the transfer learning-based children's breath sound recognition model proved its reliability even with a small dataset, and the software can display the results of breath sound recognition. The electronic stethoscope used in the system is available to the public instead of being specially designed, therefore the cost is reduced as it is spread over many users. Not to mention that the stethoscope has a wide audience to prove its effectiveness. Prevention and early intervention in pediatrics have long been far-reaching goals for health planners and academics. The proposed system for children is designed to detect emerging problems and risk factors and offer treatment early in life. Early detection and treatment can lead to better treatment outcomes, due to the fact that the disease may be in its early stages and be more responsive to treatment. Early detection and treatment can also prevent the development of diseases and reduce the risk of complications.

This research, however, is subject to several limitations. In this study, only three types of respiratory sounds have been studied, nevertheless, in reality there are other types of lung disease in children, so increasing the dataset is always necessary. In future work, more patient breath sound data will be obtained, large databases will be established and the data will be kept up to date, which will improve the generalization ability of the classification model. In addition, more functions of the software can be developed to make the system more convenient to use.

6 Conclusions

Different types of respiratory sound signals generated by the human respiratory system correspond to different respiratory conditions. In recent years, the development of artificial intelligence recognition signal technology has made the demand for respiratory sound signal monitoring increasingly strong. This paper takes the breath sound signal collected in the hospital as the research object, uses the band-pass filtering and wavelet denoising methods, studies the application of wavelet technology in the breath sound signal denoising, and adopts the deep learning and migration learning recognition model to design the recognition module. The spectrum and MFCCs features are selected for the breath sound signals of clinical medicine, and the dynamic MFCCs first-order differential parameters are innovatively added. The extracted three feature images are used to construct feature vectors as input samples for training, and the final research purpose is to accurately identify the breath sound samples to be tested and obtain the desired recognition and classification results. Furthermore, the software can display relevant information about the breath signal, making the results more intuitive. The system helps parents with lung screening and is a vital tool for diagnosing and preventing breath disease in children.

Acknowledgement: The authors are grateful to G. W. and Y. R. for their constructive suggestions and comments on the manuscript.

Funding Statement: This research was funded by the Scientific Research Starting Foundation of Hainan University (KYQD1882) and the Flexible Introduction Scientific Research Starting Foundation of Hainan University (2020.11-2025.10).

Author Contributions: Conceptualization, J.S. and G.W.; methodology, J.S.; resources, Y.R.; writing—original draft preparation, J.S.; writing—review and editing, S.C. and B.Y.; supervision, C.X., G.W. and Y.R.; funding acquisition, G.W. and Y.R. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Ethics Approval: The procedures followed in this study strictly comply with the ethical standards formulated by the Ethics Committee of the Haikou Hospital of the Maternal and Child Health, Haikou, Hainan, China. This study was approved by the Ethics Committee of the Haikou Hospital of the Maternal and Child Health (approval number [2022] 03005). Informed consent was obtained from all participants before study enrolment.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] U. A. Bhatti, Z. Zeeshan, M. M. Nizamani, S. Bazai, Z. Yu *et al.,* "Assessing the change of ambient air quality patterns in Jiangsu Province of China pre-to post-COVID-19," *Chemosphere*, vol. 288, no. 1, pp. 132569, 2022.

[2] A. K. Abbas and R. Bassam, "Phonocardiography signal processing," *Synthesis Lectures on Biomedical Engineering*, vol. 4, no. 1, pp. 1–194, 2009.

[3] D. Kumar, P. Carvalho, M. Antunes and J. Henriques, "Noise detection during heart sound recording," in *2009 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Minnesota, MN, USA, pp. 3119–3123, 2009.

[4] S. A. Taplidou and L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," *Computers in Biology and Medicine*, vol. 37, no. 8, pp. 1073–1083, 2007.

[5] U. A. Bhatti, M. Huang, D. Wu, Y. Zhang, A. Mehmood *et al.,* "Recommendation system using feature extraction and pattern recognition in clinical care systems," *Enterprise Information Systems*, vol. 3, no. 1, pp. 329–351, 2019.

[6] M. M. Azmy, "Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine," in *2015 IEEE Jordan Conf. on Applied Electrical Engineering and Computing Technologies (AEECT)*, Mövenpick Resort, Jordan, pp. 1–5, 2015.

[7] R. Palaniappan, K. Sundaraj and S. Sundaraj, "A comparative study of the SVM and K-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals," *BMC Bioinformatics*, vol. 15, pp. 1–8, 2014.

[8] S. Z. H. Naqvi, M. Arooj, S. Aziz, M. U. Khan, M. A. Choudhary *et al.,* "Spectral analysis of lungs sounds for classification of asthma and pneumonia wheezing," in *2020 Int. Conf. on Electrical, Communication, and Computer Engineering (ICECCE)*, New York, NY, USA, pp. 1–6, 2020.

[9] U. A. Bhatti, Z. Yu, J. Chanussot, Z. Zeeshan, L. Yuan *et al.,* "Local similarity-based spatial-spectral fusion hyperspectral image classification with deep CNN and gabor filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[10] U. A. Bhatti, H. Tang, G. Wu, S. Marjan and A. Hussain, "Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence," *International Journal of Intelligent Systems*, vol. 2023, no. 1, pp. 1–28, 2023.

[11] M. Aykanat, Ö. Kılıç, B. Kurt and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, pp. 1–9, 2017.

[12] Q. Chen, W. Zhang, T. Xiang, X. Zhang, S. Chen *et al.,* "Automatic heart and lung sounds classification using convolutional neural networks," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA)*, Jeju, Korea, pp. 1–4, 2016.

[13] B. Dalal, K. Zhang and A. S. Mohammad, "Lung sounds classification using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 58–69, 2018.

[14] A. Mondal, P. Banerjee and H. Tang, "A novel feature extraction technique for pulmonary sound analysis based on EMD," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 199–209, 2018.

[15] M. A. Islam, I. Bandyopadhyaya, P. Bhattacharyya and G. Saha, "Multichannel lung sound analysis for asthma detection," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 111–123, 2018.

[16] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 3, pp. 535–544, 2020.

[17] R. Shivapathy, S. Saji and N. S. Haider, "Wearables for respiratory sound classification," *Journal of Physics: Conference Series*, vol. 1937, no. 1, pp. 012055, 2021.

[18] S. Gupta, M. Agrawal and D. Deepak, "Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning," *Biomedical Signal Processing and Control*, vol. 70, pp. 102–947, 2021.

[19] P. Stasiakiewicz, A. P. Dobrowolski, T. Targowski, N. Gałązka-Świderek, T. Sadura-Sieklucka *et al.,* "Automatic classification of normal and sick patients with crackles using wavelet packet decomposition and support vector machine," *Biomedical Signal Processing and Control*, vol. 67, no. 3, pp. 102521, 2021.

[20] N. S. Haider and A. K. Behera, "Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD)," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 42–59, 2022.

[21] B. Dianat, P. L. Torraca, A. Manfredi, G. Cassone, C. Vacchi *et al.,* "Classification of pulmonary sounds through deep learning for the diagnosis of interstitial lung diseases secondary to connective tissue diseases," *Computers in Biology and Medicine*, vol. 160, pp. 106928, 2023.

[22] S. Reichert, R. Gass, C. Brandt and E. Andrès, "Analysis of respiratory sounds: State of the art," *Clinical Medicine. Circulatory, Respiratory and Pulmonary Medicine*, vol. 2, pp. CCRPM–S530, 2008.

[23] F. Hassan and A. Javed, "Voice spoofing countermeasure for synthetic speech detection," in *2021 Int. Conf. on Artificial Intelligence (ICAI)*, Rio de Janeiro, Brazil, pp. 209–212, 2021.

[24] M. F. Syahputra, S. I. G. Situmeang, R. F. Rahmat and R. Budiarto, "Noise reduction in breath sound files using wavelet transform based filter," in *IOP Conf. Series: Materials Science and Engineering*, Tianjin, China, pp. 012040, 2017.

[25] S. R. Messer, J. Agzarian and D. Abbott, "Optimal wavelet denoising for phonocardiogram," *Microelectronics Journal*, vol. 32, no. 12, pp. 931–941, 2001.

[26] S. Ervin, D. Igor and S. LJubisa, "Quantitative performance analysis of scalogram as instantaneous frequency estimator," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3837–3845, 2008.

[27] D. O'Shaughnessy, "Speech communication: Human and machine," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3837–3845, 1987.

[28] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang *et al.,* "Mobilenets: Efficient convolutional neural networks for mobile vision applications," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, pp. 4510–4520, 2017.

[29] J. Sharma, O. C. Granmo and M. Goodwin, "Deep CNN-ELM hybrid models for fire detection in images," in *27th Int. Conf. on Artificial Neural Networks*, Rhodes, Greece, pp. 245–259, 2018.

[30] D. Gradolewski, G. Magenes, S. Johansson and W. J. Kulesza, "A wavelet transform-based neural network denoising algorithm for mobile phonocardiography," *Sensors*, vol. 19, no. 4, pp. 957, 2019.

[31] M. F. Syahputra, S. I. G. Situmeang, R. F. Rahmat and R. Budiarto, "Noise reduction in breath sound files using wavelet transform based filter," in *IOP Conf. Series: Materials Science and Engineering*, vol. 190, no. 1, pp. 012040, 2017.

[32] Y. Shi, Y. Li, M. Cai and X. D. Zhang, "A lung sound category recognition method based on wavelet decomposition and BP neural network," *International Journal of Biological Sciences*, vol. 15, no. 1, pp. 195, 2019.

[33] Y. Xu, C. Zhang, Z. Xu, J. Zhou, K. Wang *et al.,* "A generic parallel computational framework of lifting wavelet transform for online engineering surface filtration," *Signal Processing*, vol. 165, no. 1, pp. 37–56, 2019.

[34] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Int. Conf. on Machine Learning*, Long Beach, CA, USA, pp. 6105–6114, 2019.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd Int. Conf. on Learning Represe (ICLR 2015)*, San Diego, CA, USA, pp. 1–14, 2015.

[36] K. Cho, B. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares *et al.,* "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1724–1734, 2014.

[37] M. Sandler, A. Howaed, M. Zhu, A. Zhmoginov and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp. 4510–4520, 2018.

[38] K. He, X. Zhang, S. Ren and J. Sun, "MobileNetV2: Inverted residuals and linear bottlenecks," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778, 2016.