IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 May 2024, accepted 20 May 2024, date of publication 23 May 2024, date of current version 31 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404657

RESEARCH ARTICLE

Classification and Recognition of Lung Sounds Based on Improved Bi-ResNet Model

CHENWEN WU, NA YE, AND JIALIN JIANG
College of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China
Corresponding author: Na Ye (731443570@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 72061022, and in part by the Natural Science Foundation of Gansu Province under Grant 21JR7RA293.

ABSTRACT Lung sound classification is an important diagnostic task in the medical field. By analyzing respiratory sounds, doctors can help diagnose various respiratory system diseases. Chronic respiratory diseases worldwide are usually associated with abnormal lung sounds, which are clinically related to conditions such as bronchitis or chronic obstructive pulmonary disease. In recent years, the outbreak of COVID-19 has once again sparked research into lung sound classification. However, due to the environmental noise and heart sounds mixed in abnormal lung sounds, further improvements are still needed for accurate classification. In this paper, an improved Bi-ResNet network structure model is proposed to enhance the accuracy of lung sound classification and fully utilize feature extraction information. The model still processes the extracted lung sound features in parallel, but by introducing skip connections and increasing the use of direct connections, it allows information to be directly transmitted and fully integrates original and processed features within the network. This improved structure enables the model to learn features from the data at a deeper level, enhancing the expressiveness of the features. Additionally, the improved Bi-ResNet model combines convolutional neural networks (CNN) and residual networks (ResNet), and uses two types of features, the lung sound short-time Fourier transform (STFT) and wavelet transform (Wavelet), for model training and analysis. This comprehensive approach captures lung sound data information more comprehensively, differentiating between different types of lung sounds and providing better diagnostic assistance to doctors, thereby promoting early diagnosis and treatment of respiratory system diseases. Through experiments, the proposed model achieved a classification accuracy of 77.81% on the Int. Conf. on Biomedical Health Informatics (ICBHI) 2017 dataset, representing a 25.02% improvement over the Bi-ResNet model, with an F1 score of 71.05%.

INDEX TERMS Lung sound classification, Bi-ResNet model, deep learning, Fourier transform, wavelet transform.

I. INTRODUCTION

Chronic Respiratory Disease (CRD) is one of the four major chronic diseases worldwide, characterized by reduced respiratory function caused by various chronic non-communicable diseases. According to statistics from the World Health Organization, approximately 400,000 people worldwide die from chronic respiratory diseases each year, with the majority of deaths attributed to Chronic Obstructive Pulmonary Disease (COPD). COPD is a chronic progressive respiratory system disease that is mainly characterized by airflow limitation, difficulty in breathing, coughing, and increased sputum production [1], [2]. Severe COPD can lead to a decrease in quality of life, reduced ability to work, and even disability. Currently, there are over 300 million COPD patients worldwide, and this number is expected to continue to increase by 2030.

Despite the significant burden that Chronic Obstructive Pulmonary Disease (COPD) poses on individual and societal health, the diagnosis and treatment of COPD still face numerous challenges in many countries, including China. Epidemiological survey data shows that there are

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

TABLE 1. The three low problems of chronic obstructive lung disease.

The three lows of chronic obstructive pulmonary disease	Diagnosis rate	Outpatient cure rate	Incidence of acute exacerbations within 1 year
Results	23.61%-30%	50%	65%

approximately 100 million COPD patients in China, with a prevalence rate of 13.7% among individuals aged 40 and above. COPD ranks third in terms of mortality and disease burden, but it has not received enough attention [3]. The three major bottlenecks currently impeding COPD management in China are insufficient diagnosis, non-standardized treatment, and poor control levels, as indicated by the three lows (as shown in Table 1). Firstly, early symptoms of COPD are similar to other respiratory system diseases, making them easily overlooked or misdiagnosed. Secondly, many patients still lack effective treatment after diagnosis, leading to disease progression and serious consequences. Additionally, due to the uneven distribution of medical resources, some regions have insufficient access to timely diagnosis and treatment for COPD patients [4].

In recent years, the application of machine learning and deep learning technologies in the field of medicine has made significant progress, providing new opportunities for early diagnosis and classification of Chronic Obstructive Pulmonary Disease (COPD). Machine learning is a method that can automatically discover patterns and make predictions by analyzing and learning from large amounts of data. It can be applied to the processing and analysis of respiratory sound data, assisting doctors in accurately diagnosing COPD and classifying different types of patients. By training and testing on a large amount of respiratory sound data, machine learning can automatically classify and identify respiratory sounds, thereby eliminating the subjectivity in doctors' judgments [5], [6]. Common machine learning algorithms include Support Vector Machine (SVM), Neural Network (NN), and Random Forest (RF). However, due to the limited number of COPD samples and the issue of imbalanced classification, traditional machine learning algorithms may lead to misjudgment and overfitting when processing COPD data.

Compared to traditional machine learning algorithms, deep learning algorithms can better handle complex medical data and improve the accuracy of diagnosis and classification of Chronic Obstructive Pulmonary Disease (COPD), assisting doctors in predicting pulmonary diseases as an auxiliary [7], [8]. Deep learning is a machine learning method based on artificial neural networks, which can automatically extract features and learn patterns through multi-layered neural network models. Deep learning has achieved significant results in medical image processing, disease prediction, and drug development, and is gradually being applied to the classification of lung sounds. Deep learning algorithms can automatically perform feature extraction and training in lung sound classification, reducing the workload of human involvement and improving the accuracy and speed of classification. However, there are still some challenges, such as designing suitable deep learning models, optimizing model performance, and fully utilizing the extracted features.

In general, machine learning and deep learning technologies provide new methods and tools for the early diagnosis and classification of Chronic Obstructive Pulmonary Disease (COPD). Through further research and development, these technologies are expected to become important adjuncts in the management of COPD, improving the accuracy of diagnosis and the effectiveness of treatment, and reducing the burden of COPD on individuals and society's health. However, it is important to note that machine learning and deep learning technologies still need to be validated and improved in clinical practice to ensure their safety and efficacy.

To address the above problems, the Convolutional Neural Networks (CNN) architecture is combined with the ability to overcome the diversity of the speech signal itself and the Residual Network (ResNet) module as a better classifier than the CNN for mutual fusion and adaptation in speech recognition. **Therefore, it is essential to assist healthcare professionals in using deep learning to identify abnormal lung sounds to predict and diagnose lung diseases accurately in the future.** This study proposes an improved Bi-ResNet model with a data augmentation approach to improve the recognition accuracy of abnormal lung sounds. **The model effectively solves the problem of using feature information to extract richer features, improving the accuracy of lung sound classification and the model's precision.**

The primary contribution of this study is as follows:

(1) Feature extraction: the information extraction from lung sound signals using short-time Fourier transform and wavelet transform.

(2) Data augmentation: a non-linear mixed data augmentation method was used to increase the amount of data on abnormal lung sounds to help model training.

(3) Feature fusion: after parallel processing of the extracted features, the ResNet module correctly identifies abnormal lung sounds.

(4) Improved Bi-ResNet model: the model improves the structure of the ResNet model by introducing shortcut connections and residual networks, and then uses a bilinear CNN model combined with residual blocks for feature processing and fuses multidimensional features to make full use of the information of the original extracted features. Finally, the fused features are classified using the residual block which improves the classification performance

and also allows the network to be trained at a deeper level.

The rest of the study is structured as follows: Section II describes related research work, Section III elaborates on data preprocessing techniques and feature extraction, and Section IV elucidates the improved Bi-ResNet model and the optimizer and loss function. Section V arranges the validation of the model's effectiveness. Finally, Section VI summarizes the study.

II. RELATED WORK

This study primarily concerns traditional and deep learning-based lung sound classification methods.

A. MACHINE LEARNING-BASED CLASSIFICATION METHODS

The earliest classification methods applied to lung sound classification were based on traditional machine learning. In the early days, they primarily focused on feature extraction and simple classification of lung sounds. Machine learning algorithms can automate the identification of abnormal breath sounds to assist doctors in accurate diagnosis and treatment, enabling early screening and prevention. However, machine learning algorithms can help doctors analyze breath sound data to improve the effectiveness of respiratory system treatment. For example, machine learning algorithms can automatically classify and analyze breath sound data to study lung disease pathogenesis and treatment options. Earlier approaches using machine learning to classify lung sounds relied primarily on Vector Quantization (VQ) techniques with the k-nearest neighbor (KNN) method, producing less accurate classification results. For example, Bahoura and Pelletier [9] used the VQ machine classification method with Mel-scale Frequency Cepstral Coefficients (MFCC) feature extraction to achieve a roaring tone accuracy of 77.5%, higher than other feature extraction methods. Jindal et al. [10] used the KNN machine classification method to add the burst tone parameters to the vector space, improving detection accuracy. Previous lung sound classification studies include machine learning methods, probabilistic statistical classification methods, and support vector machine (SVM)-based lung sound classification. Bahoura and Pelletier [11] discovered that the combination of MFCC+GMM methods had the highest sensitivity and accuracy using the Gaussian Mixture Model (GMM) to classify the lung sounds into two main categories: normal and wheeze. Moreover, the Hidden Markov Model (HMM) has been used to differentiate normal from abnormal lung sounds [12] or normal from emphysematous sounds [13], both in the context of maximum likelihood estimation for feature extraction. Simultaneously, SVM-based classifiers have been applied to lung sound classification research, with more satisfactory classification results. For example, Abbasi et al. [14] used SVMs to distinguish between normal and abnormal lung sounds and achieved a classification performance better than feed-forward neural networks (NN) and probabilistic NN. The research on classification methods is gradually improving as the research on lung sounds deepens, and machine learning-based lung sound classification methods are gradually moving towards deep learning.

B. DEEP LEARNING-BASED CLASSIFICATION METHODS

Deep learning has recently attracted much attention due to its unrivaled success in various applications, including clinical diagnosis and biomedical engineering. Studying the intrinsic connection between lung sounds and lung diseases can provide an important basis for diagnosing acute lung diseases in the clinic. Meanwhile, Jan Feiba and colleagues [15] have elaborated on the classification of pathological lung sounds and the sources of noise interference of lung sound signals in the lung sound monitoring system. They have highlighted the current existence of the short-time Fourier transform(STFT), wavelet analysis identification, and higher-order spectral analysis feature extraction methods for analyzing and identifying pathological lung sounds. The Breath Sounds Database was originally compiled to support the scientific challenges of the International Database Organization (IDO) [16], but it is now difficult to access pure breath sounds. Among them, the compilation of the ICBHI2017 challenge dataset further contributes to the study of respiratory diseases, especially with the current scarcity of medical data, allowing the intrinsic connection between lung sounds and lung disease illnesses to be better explored. Different lung sounds in the ICBHI2017 dataset are collected in different ways, leading to an imbalanced data situation, and most deep learning models require massive amounts of data. Sangmin et al. [17] proposed a simple patch-mixing augmented learning method to identify mixed features in latent space, achieving state-of-the-art performance on the ICBHI dataset with a 4.08% improvement over the previous leading score. The lung sound data processing using deep NN frequently requires huge data. A set of new techniques using device-specific fine-tuning, connection-based enhancement, blank region cropping, and smart filling are proposed [18], making more efficient use of small data. This achieves a state-of-the-art performance improvement of 2.2% over the latest results for four-class classification on the ICBHI dataset.

Li et al. [19] proposed an automated unscheduled lung sound detection method that combines augmented convolution into ResNet blocks to improve lung sound classification accuracy. Based on this, a feature extraction algorithm was proposed using a two-tone Q-factor wavelet transform combined with a triple short time-distance Fourier transform, and multi-channel spectrograms were obtained as feature inputs. This algorithm outperforms the traditional state-of-the-art method for official segmentation of the ICBHI by 1.69%. A two-channel convolutional structure was proposed to extract the Log-Mel spectrogram for feature extraction and fusion to address the problem of indeterminate speech classification by NN models [20]. A bilinear Bi-ResNet model [21] was used for simultaneous training and learning of key features required in the recognition process to

differentiate between different types of indeterminate lung sounds, with a score improvement of 4%. Shuvo et al. [22] proposed a lightweight CNN architecture for respiratory disorders using hybrid-based lung sound feature Classification.

Chen et al. [23] proposed a breath sound three-classification method based on optimized S-transform and deep residual network, aiming to achieve accurate classification of breath sounds. The main objective is to improve the classification accuracy and performance of breath sounds by combining the methods of S-transform and deep residual network. Phettom et al. [24] proposed an automatic abnormal lung sound identification method based on time-frequency analysis and CNN. They used traditional spectral analysis techniques to extract time-frequency features of lung sound signals. By combining the information of time-frequency features and the powerful representation capability of deep learning models, the method can accurately determine the abnormal nature of lung sounds. Tsai et al. [25] introduced a deep neural network-based model for separating heart and lung sounds. This novel deep learning architecture combines CNN and recurrent neural networks (RNN) to learn the complex representation of heart and lung sounds. The proposed model utilizes the spatial information captured by CNN and the temporal dependencies captured by RNN, potentially improving the accuracy of diagnostic systems in the field of respiratory and cardiac medicine.

III. DATA PREPROCESSING AND FEATURE EXTRACTION

A. ICBHI 2017 DATASET

The ICBHI 2017 dataset is a large publicly available database that provides official splitting and assessment methods [16]. The dataset was derived from 126 study participants, recorded for 5.5 h, and contained 6898 respiratory cycles. The expiratory cycles included, 364 'normal,' 1864 'crackle,' 886 'wheeze,' and 506 'crackle plus wheeze.' However, these breath sound recordings were collected from various hospitals using four different devices, the duration of the recordings ranged from 10s to 90 s, and the data were uneven. Meanwhile, the use of different devices leads to the presence of varying levels of noise in the collected lung sounds. Therefore, preprocessing and data enhancement were performed in this study to address the above issues.

B. PREPROCESSING TECHNOLOGY

The ICBHI 2017 dataset contains two types of abnormal breath sounds: crackle and wheeze. Crackle a low-pitched sound with a duration shorter than 20 ms that frequently occurs in multiple consecutive and brief occurrences. This lung sound is frequently associated with bronchiectasis and chronic bronchitis. Wheeze are additional sounds of respiration that are high-pitched, musical in character, and have a duration of more than 250 ms. This lung sound is frequently associated with obstructive lung diseases, such as bronchial asthma and cystic fibrosis.

The frequency of these two abnormal respiratory sounds in the dataset is above 1,000 to 2,500Hz [26], while the frequency range of normal respiratory sounds is mostly between 60 to 600Hz. Therefore, it is necessary to resample all respiratory signals to a uniform sampling frequency. Here, it was chosen to resample the respiratory signals to 4,000Hz to ensure the consistency of data processing. Additionally, in order to mitigate the influence of different environmental noises, a third-order Butterworth high-pass filter with a maximum attenuation of 2dB in the passband was used to preserve the frequency band of 50 to 2,000Hz [19], [27]. This approach removes low-frequency and high-frequency noise while retaining the main frequency components of the respiratory signals. Finally, the authors normalized the input signals to a unified standard. This involved applying the same scaling process to all respiratory signals, resulting in their numerical range falling within a standard range of 0 to 1, to facilitate better handling and comparison of the data in subsequent data analysis and model training.

C. FEATURE EXTRACTION

STFT and wavelet transform (wavelet) are commonly used for feature extraction [28], [29]. This method can convert raw time-domain audio into time-frequency or multi-scale representations for subsequent classification or signal-processing tasks. Combining SFTF and wavelet can obtain a richer feature representation of audio signals and improve its characterization and discriminative performance to play a critical role in different audio signal processing.

By combining the use of STFT and wavelet, we can fully leverage their respective advantages and obtain a more comprehensive feature representation. STFT provides high frequency resolution and good capability for capturing local features, but the window length of STFT affects the trade-off between frequency and time resolution. A shorter window length can provide higher time resolution but lower frequency resolution, while a longer window length is the opposite. This means that when using STFT, we need to balance between time and frequency resolution. On the other hand, wavelet analysis can provide multi-resolution analysis of signals at different frequencies by using wavelet functions of different scales. Therefore, for signals with varying frequency characteristics, wavelet analysis can provide good time and frequency multi-resolution properties. Therefore, the combination of STFT and wavelet can better capture the time and frequency features of signals and improve the model's detection ability for events. Accurate and reliable feature extraction is crucial in the field of medical applications for diagnosis, monitoring, and treatment. Applying STFT and wavelet transform to medical data, respiratory and lung sounds can provide more accurate and comprehensive feature representation. Through this approach, a deep analysis of respiratory and lung sounds in medical data can be conducted, bringing more accurate and effective solutions to medical research and clinical practice.

Specifically, the method of combining STFT and wavelet transform can be achieved through the following steps:

(1) Perform STFT analysis on the signal: STFT divides the signal into windows of different frequencies and calculates the spectrum of each window to extract the frequency information of the signal.

(2) Apply wavelet transform to the signal: Applying wavelet transform to the signal to take advantage of the time and frequency multi-resolution properties of wavelet transform. This can extract the time and frequency characteristics of the signal, thereby obtaining a more comprehensive feature representation.

(3) Feature fusion and model training: Fuse the features obtained from STFT and wavelet transform and use them for model training. The fused features can more comprehensively describe the time and frequency characteristics of the signal, thereby achieving more accurate and effective signal analysis and event detection in the field of medical applications.

1) SHORT TIME FOURIER TRANSFORM (STFT)

STFT is a commonly used time-frequency analysis technique applicable in speech recognition and audio processing. It can decompose the signal in both time and frequency domains and extract the time-frequency features of the signal for subsequent classification and recognition tasks. The primary idea of the method is to divide the signal into several small segments not exceeding the window length and perform the Fourier Transform (FT) on each of the small segments separately to obtain information on each small segment in the frequency domain. For the non-stationary characteristics of lung sound signals, STFT can establish a local relationship between the time domain and frequency domain, thereby reducing the requirement for the stationarity of lung sound signals. For the lung sounds in the dataset, a window length of 20ms and a step size of 10ms are used on each segment, and a Hanning window is applied to capture the frequency domain information within a short time. Figure 1 shows the waveform of the right channel for different lung sound types, as well as the spectrogram obtained after the short-time Fourier transform. Figure 1 provides a good observation of the time-domain waveform of the original lung sound signal and the corresponding frequency domain information at each moment. The extracted spectrogram also effectively reflects the spatial distribution of energy in the lung sound signal, indicating that different energy distributions represent variations in signal characteristics, thus providing strong support for the analysis of signal characteristics.

The STFT formula is as follows:

$$X(m,k) = \sum_{n=0}^{N-1} x(n)\omega(n-m)e^{-j2\pi nk/K} \tag{1}$$

where, $x(n)$ is the time-frequency sample of the original signal and $X(m,k)$ is the frequency domain result after STFT. $\omega(n-m)$ is the window function, using Hann window. N is the window length, m is the window position, and K is the resolution of the frequency domain, which is usually the same as the one in the frequency domain or a few times more to improve the resolution of the frequency domain.

2) WAVELET TRANSFORMS

Wavelet transform is a transform method that decomposes the signal by scale, effectively extracting signal characteristics from multiple scales while reducing the influence of non-identical frequency band noise in lung tone. STFT calculates the signal based on a fixed window size. Thus, combining the information characteristics of different frequency components is possible, causing some interference. However, the wavelet transform uses wavelet basis functions, such as db8, to select appropriate scales according to different features of the signal, thereby improving the time-frequency resolution. In this case, the choice of wavelet base also indirectly affects the analysis results of lung sound signals; thus, choosing a suitable wavelet base for feature extraction is necessary. However, wavelet analysis is poorly adaptive to lung sound signals, primarily because it is unsuitable for local analysis. Combining the two can more fully utilize the advantages of both types of methods and improve the model recognition accuracy.

The wavelet transform is an inner product of the original image and the scale function. However, the wavelet basis function in the wavelet transform was generated by scaling and translating the same fundamental wavelet function. The author employed discrete wavelet transform (DWT) and selected appropriate wavelet basis functions based on different signal characteristics. The main steps of this transformation include:

(1) defining wavelet basis functions: selecting suitable wavelet basis functions, which are crucial for signal decomposition and reconstruction;

(2) multi-level decomposition: performing multi-level decomposition of the signal into multiple components for a more detailed analysis;

(3) calculating detail coefficients and approximation coefficients: computing detail coefficients and approximation coefficients for each component to obtain a set of wavelet components and detail components, reflecting the high-frequency and low-frequency components of the signal, respectively;

(4) reconstructing the original signal: combining the wavelet components and detail components through inverse wavelet transform for signal analysis and processing. Figure 2 shows the waveform of the original lung sound right channel and the spectrogram obtained after wavelet transform. These spectrograms reflect the frequency domain characteristics of lung sound signals, facilitating a deeper analysis and understanding of the lung sound signals.

Given an input signal $x(n)$ of length N, DWT decomposes it into a set of low-frequency coefficients $c_a(n)$ and a set of high frequency coefficients $c_d(n)$ of length $N/2$, as follows:

$$x(n) = c_a(n) + c_d(n) \tag{2}$$

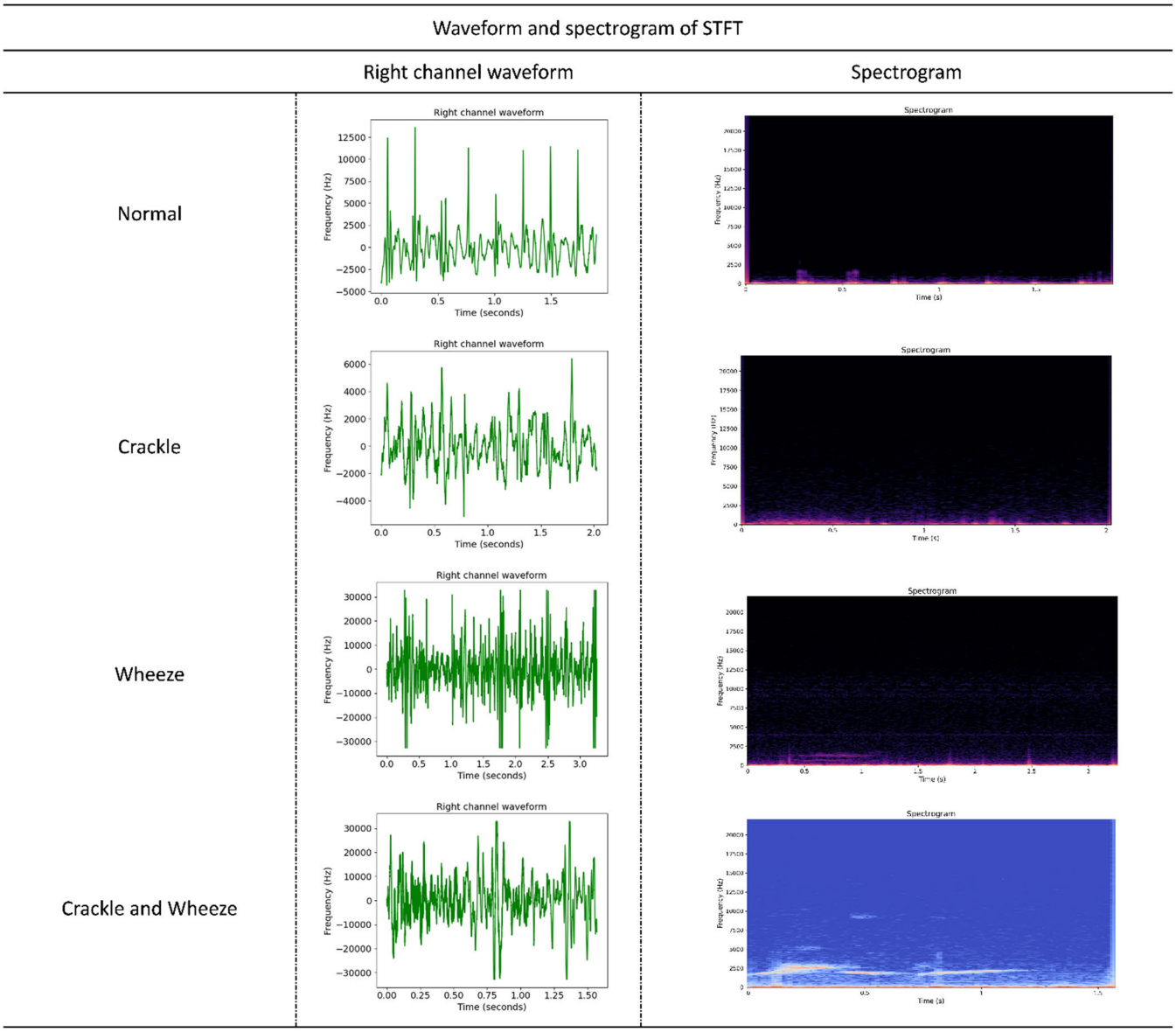

FIGURE 1. Visualization of STFT raw lung sound left and right channel waveforms and spectrograms.

The low-frequency coefficients c_a of layers 2 to 4 were saved, and the high-frequency coefficients c_d of layers 1 to 5 were horizontally spliced. Finally, all the coefficients were horizontally spliced together as the output of the wavelet transform to facilitate the analysis and processing of the signal.

D. MIXED DATA ENHANCEMENT AND DISTRIBUTION STRATEGY

1) MIXED DATA ENHANCEMENT

These data are considered unbalanced in the ICBHI 2017 dataset. This creates a common problem in medical classification tasks [30], namely that the probability of obtaining anomalous samples is too small, and such a problem causes the classification model to ignore a few samples or produce overfitting, leading to classification errors. A data-independent mixed data augmentation approach was utilized to achieve a linear transformation of decision boundaries from class to class [31]. Mixed enhancement is a data enhancement method based on domain risk minimization, using linear interpolation to obtain new sample data, attempting to make discrete sample points continuous [32]. This mixed approach increases the diversity of samples, smoother the transition between decision boundaries of different categories, reduces the misidentification of some samples, increases the model's robustness and stability during training, and improves the model's robustness. Therefore, its data generation method is as follows:

$$(x_n, y_n) = \lambda(x_i.y_i) + (1 - \lambda)(x_j.y_j) \tag{3}$$

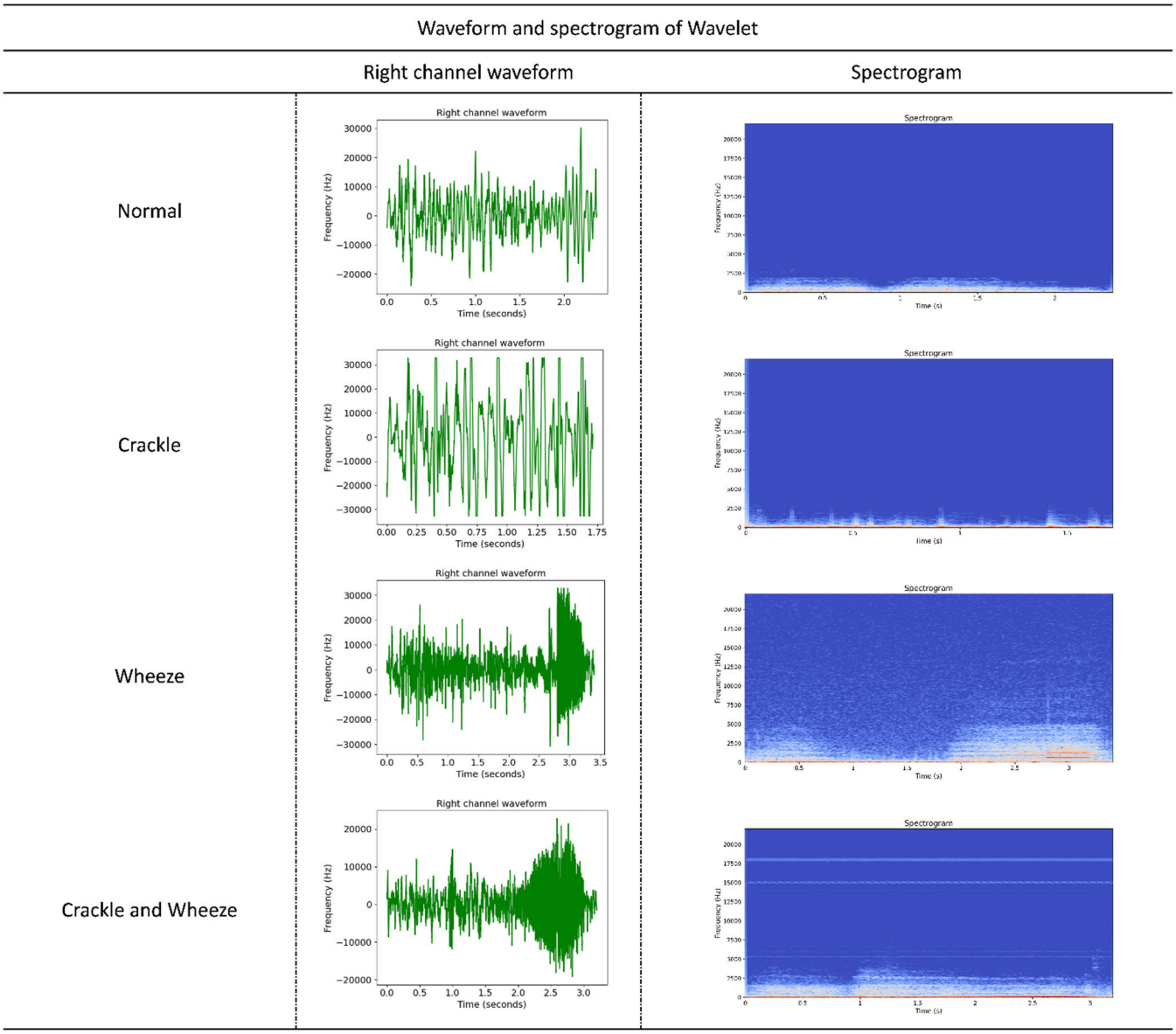

FIGURE 2. Visualization of Wavelet raw lung sound left and right channel waveforms and spectrograms.

where x_i and x_j are the two training samples, x_n are the generated training samples, y_i, y_j, y_n and are the corresponding sample labels. The mixing ratio λ is sampled from the beta distribution, ranging between 0 and 1.

In this study, a mixed data enhancement-based approach [31] was used to process the data for the lung sound characteristics on the ICBHI217 dataset (Figure 3). The mixed data augmentation method proposed in this paper can be achieved through the following steps:

(1) Data selection and grouping: Firstly, samples of crackle cycles, wheeze cycles, and normal cycles are selected from the ICBHI217 dataset and grouped into different categories.

(2) Data mixing process: Crackle cycles are combined with normal cycles to increase the number of crackle cycles. Similarly, wheeze cycles are combined with normal cycles to increase the number of wheeze cycles. Additionally, by combining crackle cycles and wheeze cycles, samples containing both types of abnormal breath sounds can be obtained.

(3) Data labeling and integration: The mixed data processed is labeled to differentiate between different types of cycles, ensuring the integrity and accuracy of the label information. These mixed processed data are then integrated into the original dataset to form the enhanced dataset.

Through this method, the quantity of crackle and wheeze cycles can be increased, thereby improving the representation of these two types of abnormal breath sounds and assisting in enhancing the performance and generalization ability of the model. Furthermore, it provides a more diverse and enriched data sample, aiding in a more in-depth analysis and understanding of lung sound characteristics. Table 2 clearly

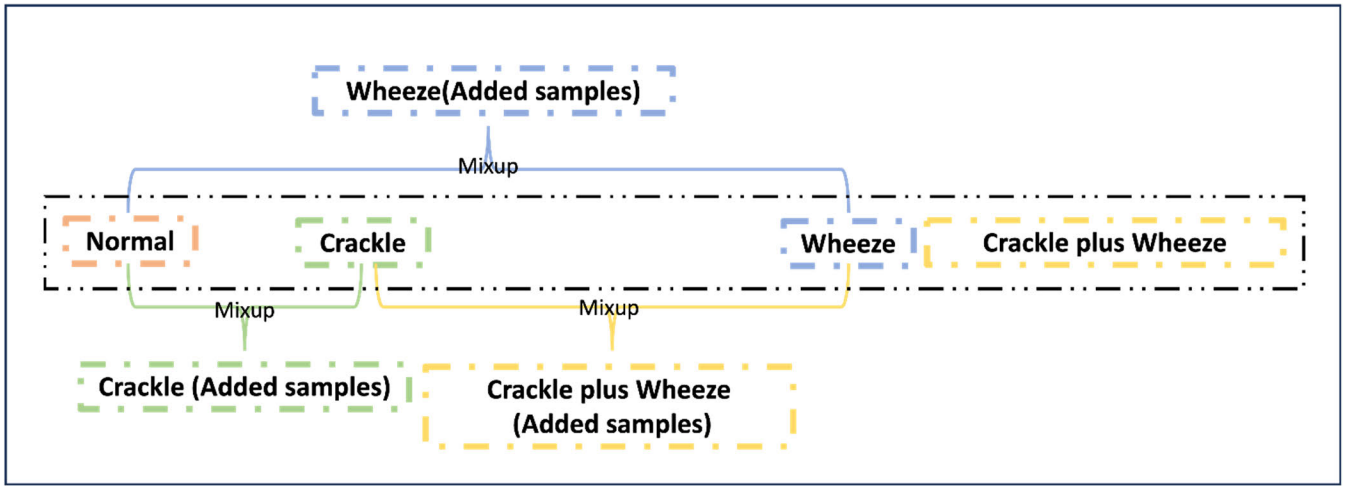

FIGURE 3. Mixed data enhancement.

TABLE 2. Number of lung sound samples before and after data enhancement.

Signal Type	Before data enhancement(cycles)	Before data enhancement(cycles)
Normal	3642 cycles	7265 cycles
Crackle	1864 cycles	6164 cycles
Wheeze	886 cycles	4025 cycles
Crackle and Wheeze	506 cycles	3305 cycles

demonstrates the number of subjects and the number of respiratory cycles analyzed for each type of lung sound sample before and after data enhancement.

2) DISTRIBUTION STRATEGY

In order to avoid data leakage and ensure the validity of model evaluation, we adopted a data allocation strategy. In this study, we are particularly concerned with ensuring that the same subject data does not appear in both the training and test sets when using data enhancement techniques. The following are the specific steps we took:

(1) Subject-level segmentation: we ensure that when dividing the dataset into training and test sets, each subject's data is completely contained in one set. This means that all respiratory cycle data for each subject is either fully assigned to the training set or fully assigned to the test set. This strategy ensures that the model is not exposed to overlapping data from the same subjects during training and test.

(2) Independence of data augmentation: when augmenting the data, the training set and test set are augmented separately. This means that even though the augmented data may appear in both sets, the original and augmented data do not span the two sets.

IV. IMPROVEMENT OF Bi-ResNet

A. IMPROVED Bi-ResNet NN MODEL

The three-channel spectrograms after STFT and wavelet transforms were fed into the convolutional layer separately (Figure 4). Afterward, they were fed into the linear ResNet I and ResNet II modules with down-sampling by maximum pooling and then into the ResNet II layer by matmul, and into the respiratory sounds via the ResNet [33], GroupNorm [34], ReLU [35], global average pooling [35], and the two fully-connected layers. Finally, the respiratory sounds were classified into four categories.

The bilinear ResNet model architecture adds a bilinear pooling layer to the ResNet model. Bilinear pooling can achieve better image extraction of high-level features of the model, improve the model's generalization ability, and to some extent, increase its robustness compared to the traditional CNN model without increasing its complexity.

Based on dual ResNet, the researchers enrich the feature information by adding directly connected edges to the model and trained by residual network after feature fusion, which allows the model to better capture audio features and make better use of the existing features (Figure 4). Simultaneously, this ResNet module can effectively improve the accuracy and robustness of feature extraction as well as avoid the gradient vanishing problem [33], thereby improving the accuracy of speech recognition and audio classification. At the end of the NN architecture, dropout [36] was applied to the two fully connected layers, thus effectively avoiding the fitting problem.

B. MODEL OPTIMISER AND LOSS FUNCTION

The researchers used a stochastic gradient descent (SGD) optimization algorithm to improve the accuracy and avoid the problem of model overfitting. Although not every

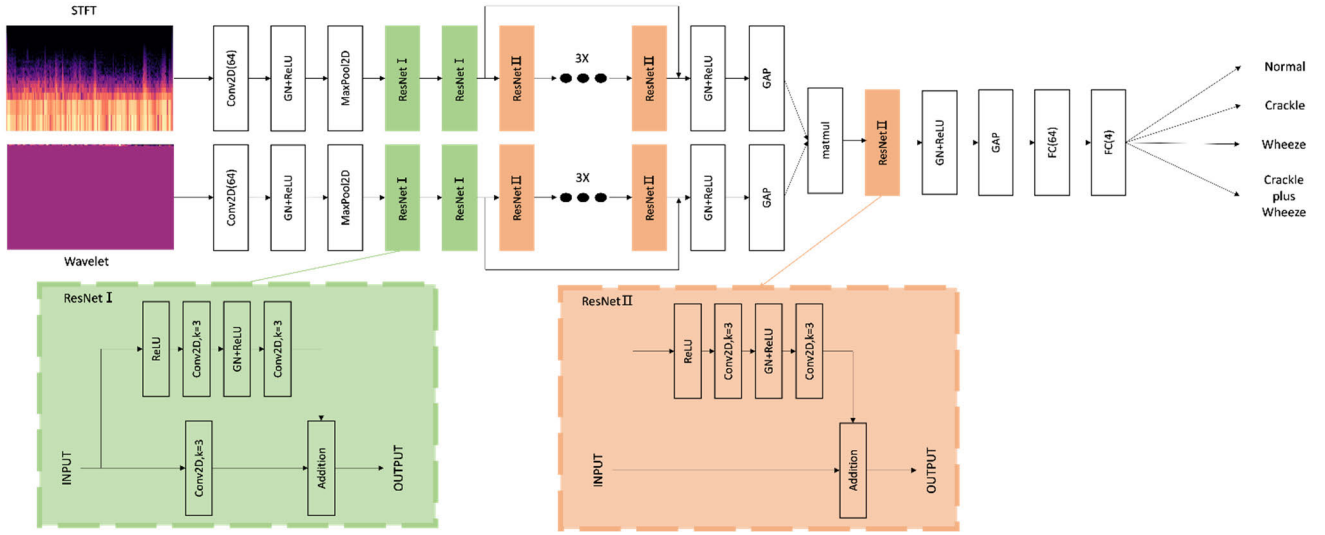

FIGURE 4. Improvement of the Bi-ResNet model.

iteration of SGD for the optimization problem produces a loss function in the direction of the global optimum, the direction of the result is in the vicinity of the global optimal solution.

This study used a loss function to solve the multi-label multi-classification problem. This function compares the predicted values of the inputs with the true values and calculates the difference between the two, allowing the network to gradually improve its ability to predict the target variable more accurately. This loss function can calculate the loss function and gradient values in a forward pass together, which can greatly improve the efficiency in training the network structure.

C. EVALUATION METHODOLOGY

This study used the 60/40 division dataset splitting method and the following evaluation method to validate the effectiveness of the data processing [37].

$$S_e = \frac{P_c + P_w + P_b}{N_c + N_w + N_b} \tag{4}$$

$$S_p = \frac{P_n}{N_n} \tag{5}$$

$$Score = \frac{S_e + S_p}{2} \tag{6}$$

where, P_c, P_w, P_b and P_n are the number of respiratory cycles correctly predicted by the four types of lung sounds, and N_c, N_w, N_b and N_n are the total number of instances in each type of sub-lung sound cycle, respectively. Sensitivity (S_e) measures the proportion of lung disease samples detected by the lung sound classification system, while specificity (S_p) measures the proportion of healthy samples correctly identified as healthy by the lung sound classification system. These metrics are important measures of the accuracy and reliability of automated lung sound classification systems.

V. EXPERIMENTAL SETUP AND RESULT

A. EXPERIMENTAL SETUP

In this study, the training and test sets were divided according to the official division of the ICBHI2017 dataset, and the NN classification model was implemented in Python 3.10 using pytorch and evaluated on a CPU based on a MacOS system with 16GB of RAM and an Apple M2 chip. For the network structure in this study, the learning rate was set to 10e-6, the batch size was 64, the epoch was 100, and the learning rate decays exponentially every 20 epochs. These parameters are reasonable settings chosen after a combination of several experiments to obtain better convergence and performance during training. Dropout rates of 0.2, 0.3, and 0.4 were applied to the network structure in this paper. A higher dropout rate can enhance the model's robustness, reduce the likelihood of overfitting, but may also lead to underfitting. Thus, the best dropout rate was selected to train the model in the experiments. Additionally, a weight decay coefficient of 0.06 was set to prevent overfitting and keep the weights at a small value to prevent gradient explosion. This parameter constrained the complexity of the network by constraining the weights, avoiding the model being overly complex and having poor generalization ability. In this paper, a series of experiments were conducted to evaluate the effectiveness of the proposed network structure and data augmentation techniques. By comparing with state-of-the-art methods, the authors gained a better understanding of the model's performance on this task, enabling the assessment of the method's strengths and weaknesses. Additionally, the authors analyzed the impact of parameter selection on experimental results to better understand the behavior and performance of the model.

B. EXPERIMENTAL RESULT

1) EFFECTIVENESS OF WAVELET BASE SELECTION

In this study, the data from the lung sound dataset is a non-stationary signal, and the researchers are trying to better

TABLE 3. Comparison of wavelet base selection results.

Wavelet base	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Coif6	30.42%	18.90%	83.64%	51.27%
Coif12	60.53%	98.43%	1.83%	50.13%
db8	77.81%	61.99%	90.10%	71.05%

TABLE 4. Comparison of results before and after MIXED data enhancement.

Data Enhancement	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Before data enhancement	43.68%	40.16%	51.82%	45.99%
After data enhancement	77.81%	61.99%	90.10%	71.05%

identify the abnormal lung sound data. As this study uses DWT for feature extraction of lung sound data, the wavelet transform is an inner product operation of the original image with wavelet basis and scale functions. Different wavelet basis functions have different frequency and time characteristics and can be applied to different signals. Thus, the choice of wavelet basis greatly impacts the subsequent lung sound classification results.

According to the characteristics and requirements of the signal, it is necessary to choose a suitable wavelet basis function for wavelet transform, and the common wavelet basis functions include Daubechies, Symlets, and Coiflets. Daubechies wavelet basis function is the most used, suitable for smooth and non-smooth signals at lower frequencies. Daubechies is more compact and is suitable for high frequency signals. Finally, Coiflets wavelet basis functions are used for non-smooth signals and have better time localization properties.

For the choice of wavelet basis functions, we conducted experiments to compare the effectiveness of Coiflets and Daubechies wavelet bases in abnormal lung sound classification, considering their performance and computational capabilities in discrete wavelet transform. First, we noticed that Daubechies wavelet bases have orthogonality, which means they can decompose the signal into mutually orthogonal subspaces. This orthogonality gives Daubechies wavelet bases an advantage in signal analysis and feature extraction. Orthogonality provides better frequency resolution and temporal information, allowing for more accurate capture of details and features in abnormal lung sound signals. In contrast, Coiflets wavelet bases are slightly inferior in terms of orthogonality. Secondly, Daubechies wavelet bases exhibit excellent computational capabilities in discrete wavelet transform. They have high computational efficiency and stability, better preserving the energy and shape characteristics of the signal. This is important for the classification and feature extraction of abnormal lung sound signals as they often contain useful information and subtle vibration patterns. Based on our experimental results as shown in Table 3, it can be seen that for the task of lung sound classification, the use of the Daubechies wavelet basis (db8) achieved the highest accuracy, 77.81%. Additionally, db8 also demonstrated relatively high sensitivity (61.99%) and specificity (90.10%). On the one hand, it can be seen from the table that the sensitivity of Coif12 is as high as 98.43%, while the sensitivity of db8 is significantly lower than that of Coif12. The sensitivity indicates the ability of the classification model to correctly identify the normal lung sound samples, and therefore there is still a certain shortcoming in the case of db8 in relation to Coif12 in terms of not being able to capture the normal lung sound samples well enough to lead to omissions. However, all things considered, it can be concluded that for the lung sound classification task, the use of the db8 wavelet basis is most appropriate as it achieves better performance in terms of accuracy, sensitivity and specificity. This will help to improve the accuracy and reliability of medical data analysis and can improve the accuracy and performance of classification models.

2) EFFECTIVENESS OF MIXED DATA ENHANCEMENT

To address the issue of class imbalance in the ICBHI2017 dataset, we employed data augmentation techniques. Data augmentation is a method of increasing the quantity of minority class samples by transforming, expanding, and synthesizing them. By increasing the number of samples in the minority classes, we are able to improve the accuracy and sensitivity of the classification model on these classes. We generated new samples that are similar to the original data but slightly different. This way, during the training of the classification model, we can use both the original data and the augmented data, allowing the classifier to better learn the characteristics and patterns of the minority classes.

Based on our experimental results shown in Table 4, the accuracy and F1 scores of the model are significantly improved after applying the data enhancement technique. Data enhancement also significantly improved the sensitivity and specificity of the model, which is very appropriate and effective in lung sound classification tasks. This indicates a significant impact in improving the model performance,

TABLE 5. Results of experiments comparing methods related to lung sounds.

Model	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
LungAttn[14]	N/A	36.36%	71.44%	53.09%
Bi-ResNet[15]	52.79%	31.12%	69.20%	50.16%
ResNet50(Co-Tuning, Log-Mel)[17]	N/A	37.24%	79.34%	50.58%
CNN-DNN(Log-Mel)[38]	N/A	30.00%	69.00%	46.00%
C-DNN+Autoencoder (Gammatonegram) [39]	N/A	30.00%	69.00%	42.00%
CNN-MoE[40]	N/A	26.00%	68.00%	47.00%
Ours	77.81%	61.99%	90.10%	71.05%

N/A: Not mentioned in the reference papers.

TABLE 6. Results of ablation experiments.

Model	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Ours(No directly connected edge - no ResNet II-6)	41.96%	38.19%	53.66%	45.92%
Ours(directly connected edge – no ResNet II-6)	57.86%	92.52%	4.82%	48.67%
Ours	77.81%	61.99%	90.10%	71.05%

as well as being effective in improving the model's classification performance in a category imbalanced dataset.

Although we increased the number of samples for normal lung sounds, crackle sounds, wheeze sounds, and the combination of crackle and wheeze sounds by 3623, 4300, 3139, and 2799, respectively, through the mixed data enhancement technique, the amount of data is still small for classification model training, which may lead to the model not being able to learn the features of the data adequately, which may result in a certain degree of recognition bias, and problems such as failure of the training process to converge, which affects model training effectiveness and performance. The problem affects the training effect and performance of the model.

3) EFFECTIVENESS OF THE PROPOSED CLASSIFICATION MODEL

The research findings in this paper contribute significantly to the early diagnosis and treatment of lung diseases. Through improved models and deep learning techniques, this study efficiently identifies abnormal lung sounds and achieves significant performance improvement in classification tasks. This provides accurate auxiliary information for doctors during the early diagnosis stage, helping them promptly adopt treatment measures and enhance the therapeutic effects and survival rates of patients.

To demonstrate the effectiveness of the proposed model, a comparison was made with recent convolutional neural network models for lung sound classification, such as LungBRN, ResNet50, and Bi-ResNet. The results are shown in Table 5. It is clear from the table that the proposed model in this paper has achieved significant improvements compared to other lung sound classification models. Firstly, the accuracy of the model in this paper reaches 77.81%, which is a 25.02% improvement over previous models. Additionally, as an important metric for evaluating the performance of classification models in machine learning, the F1 score is as high as 71.05%, demonstrating excellent performance of the proposed classification model in handling the imbalanced ICBHI2017 dataset. It is worth noting that the proposed method has achieved significant improvements in accuracy, sensitivity, and F1 score, further proving the superior performance of the model in the task of lung sound classification. The significant enhancement in sensitivity, as clearly seen in the table, demonstrates the good performance of the improved model in identifying abnormal lung sounds, indirectly proving the important clinical significance of the model in the early diagnosis and analysis of pulmonary diseases. This indicates that the improved model in this paper exhibits good performance in recognizing abnormal lung sounds and provides more accurate classification results for doctors. This is essential information for doctors in the decision-making process, helping them make more accurate diagnoses and treatments for lung diseases.

4) ABLATION EXPERIMENT

To evaluate the effectiveness of the proposed improved ResNet model in this paper, we conducted a comparison of the quantities of the original ResNet model and ResNet II model, considering multiple evaluation metrics including F1 score, accuracy, sensitivity, and specificity.

Specifically, we experimentally compared the performance of the improved model with the original model and the ResNet II model in different tasks. According to the results in Table 6, the performance of the improved model under different settings can be observed.

We note that shortcut connections are also important for raw feature extraction. These shortcuts act as information shortcuts and help the model to better utilize the original

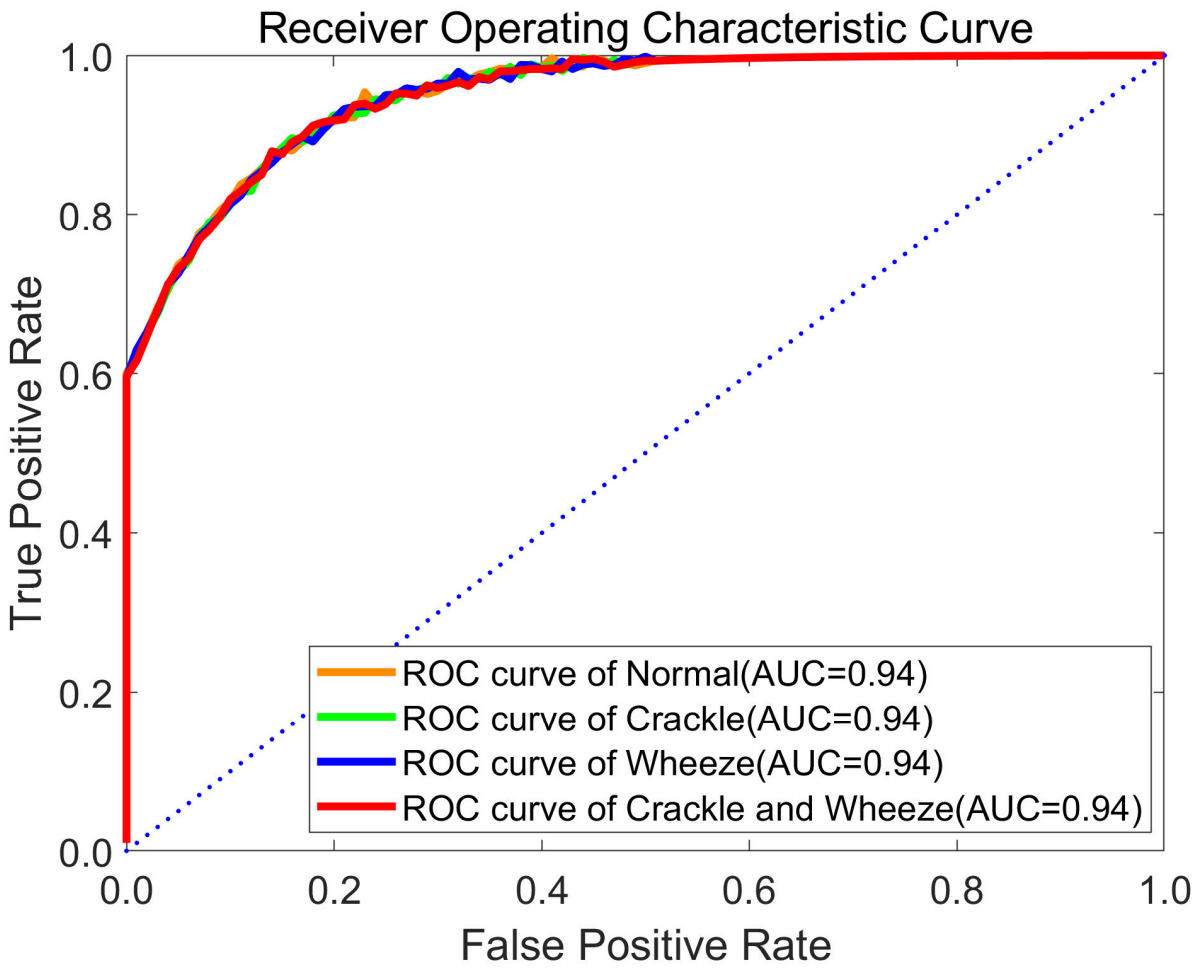

FIGURE 5. The receiver operating characteristic (ROC) curve.

features and avoid information loss and ambiguity. With the introduction of shortcut connections, the model can more accurately extract and utilize the key information in the original data, thus improving the accuracy and sensitivity of classification. Observing the change from the first row to the second row through Table 6, we can see that when the model uses the directly connected edge (directly connected edge), the accuracy increases from 41.96% to 57.86%, which is a significant improvement. However, this improvement is also accompanied by a significant decrease in specificity from 53.66% to 4.82%, and the model has a high misdiagnosis rate in identifying negative samples, which may lead to the model generating more false positives in practical applications. This indicates that although the model has improved in some aspects, there is a significant decrease in the performance in terms of specificity. However, there was a greater improvement in sensitivity, allowing better identification of normal lung sound samples and helping physicians to rule out cases of non-chronic respiratory disease. This method of feature fusion using directly connected edges allowed the model to capture features from different layers more comprehensively and achieved significant improvements in all metrics. The improved model was able to make better use of the valid information in the data by using features processed in the first step in conjunction with those processed in the subsequent step. This feature fusion approach allowed the model to capture features at different levels more comprehensively and achieved significant improvements in various metrics.

Additionally, we also observed that the residual modules play a crucial role in accurate classification. The residual modules introduce skip connections, allowing information to be directly propagated in the network, avoiding the issues of gradient disappearance and model degradation. This architectural design enables the model to learn the features and patterns in the data at a deeper level, thereby enhancing its ability to extract and represent key information from lung sounds. As a result, it improves the classification performance. As observed in Table 6, although the overall improvement in classification is somewhat, the sensitivity decreases to 61.99%, indicating that there is a certain bias situation in our model in identifying the correct lung sound samples.

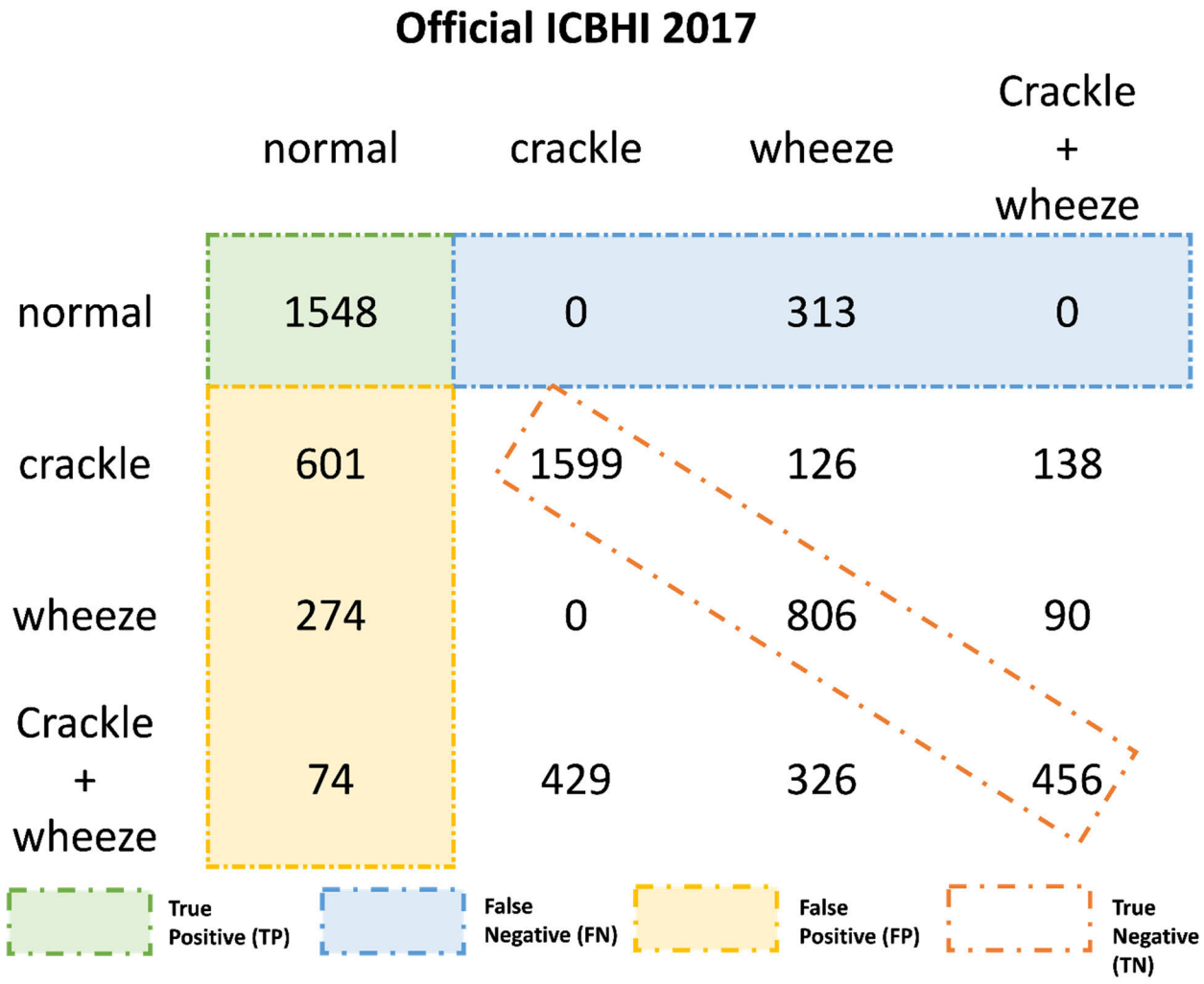

FIGURE 6. Confusion matrix based on the official ICBHI 2017.

Overall, when the model used both shortcut connected edges and ResNet II-6, the accuracy was further improved to 77.81%, sensitivity to 61.99%, specificity to 90.10%, and F1 Score to 71.05%.

In summary, the improved ResNet model proposed in this paper, through techniques such as feature fusion, residual modules, and shortcut connections, can better utilize the effective information in the data, improving the performance and results of classification tasks. Experimental results demonstrate that fully utilizing the extracted feature information can enhance the accuracy of the classification model. The proposed improved model can better utilize information and achieve good classification results, which is of significant for the further development of fields such as lung sound classification.

5) SEGMENTATION PERFORMANCE

According to the results in Table 7, we can clearly see the comparison between the division method based on the official ICBHI2017 dataset and the results of the author's randomly divided 80/20 dataset. From the table, it can be observed that without the official strict division, the 80/20 division for the identification of abnormal lung sounds is significantly inferior to the official division. Specifically, when using the official division (60/40 division), the accuracy reached 77.81%, while without data augmentation, the accuracy was 43.68%. This indicates a clear advantage of the official division for the identification of abnormal lung sounds.

On the other hand, when using the 80/20 division, even with data augmentation, the accuracy was only 42.69%, and without data augmentation, the accuracy was 38.12%. This further confirms the author's point that the 80/20 division is clearly inferior to the official division in the identification of abnormal lung sounds.

Therefore, based on the results in Table 7, it can be concluded that the requirements of the official division have a certain advantage in the identification of abnormal lung sounds in the ICBHI2017 dataset. This viewpoint has been validated through experimental results.

TABLE 7. Official division and 80/20 division experimental results.

Method of division	Data Enhancement	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Official division (60/40 division)	Yes	77.81%	61.99%	90.10%	71.05%
	No	43.68%	40.16%	51.82%	45.99%
80/20 division	Yes	42.69%	12.26%	92.31%	52.29%
	No	38.12%	9.43%	94.44%	51.94%

Figure 5 shows the Receiver Operating Characteristic (ROC) curve, from which it can be clearly seen that the model has good classification performance. However, it is still challenging to accurately distinguish between ''normal'' sounds, ''crackle'', ''wheeze'', and the combination of crackle and wheeze. Figure 6 presents the confusion matrix of the training and testing results obtained from official partition. It can be observed from the figure that the proposed model can better classify lung sounds to some extent. By accurately classifying different types of lung sounds, doctors can have a better understanding of the patient's lung condition, thereby providing more accurate and effective diagnosis and treatment plans. Additionally, the classification results of the model can help doctors quickly identify abnormal cases from a large amount of lung sound data, improving the sensitivity and early detection ability of diseases. Furthermore, the automated classification of the model can reduce the workload on doctors, improve the efficiency of medical institutions, and enhance the consistency of diagnoses.

VI. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. This paper proposes an improved method for abnormal lung sound classification, aiming to enhance the accuracy of classification. The main contributions of this paper can be summarized as follows: (1) Feature extraction and fusion: In this paper, the short-time Fourier transform and wavelet transform are mainly used for feature extraction, and the original feature information is fully utilized by the direct connecting edges introduced by the model, and feature fusion is applied to correlate the data after the extracted features are processed in parallel. These feature representations can more accurately characterize the abnormal lung sounds and provide useful inputs for subsequent classification models. (2) Data augmentation: To increase the number and diversity of samples, the paper adopts a mixed data augmentation method to generate new samples of crackles and wheezes in the respiratory cycle. By increasing the diversity of samples, the accuracy of the deep model classification can be improved, enabling the model to better adapt to different sample conditions. (3) Classification with deep models: In the ResNet-based classification model, the paper introduces shortcut connections and residual networks. Shortcut connections allow the model to fully utilize the initial feature information, avoiding information loss and blurring. Residual networks can better extract and fuse features, enhancing the model's understanding of sample information. Meanwhile, the paper processes two independent multidimensional features in parallel and fully integrates them, improving the performance and effectiveness of the classification model. Through experiments on the ICBHI2017 lung sound dataset, the effectiveness of the proposed method is verified, and compared with other lung sound classification models, the proposed method in this paper exhibits better performance. However, there is still room for improvement in the model proposed in this paper. In future research, we will further explore the effectiveness of other classification modules, simplify model parameters to reduce model complexity, and improve the classification efficiency of the model. Specifically, we will investigate which classification module can better adapt to the task of abnormal lung sound classification and optimize model parameters to balance accuracy and model complexity. The method proposed in this paper not only has practical value in the field of lung sound signal processing, but also can bring new technical ideas and solutions to the field of image processing.

In summary, the proposed method for abnormal lung sound classification in this paper effectively improves the accuracy of classification through techniques such as feature extraction, data augmentation, and classification with deep models. However, although these techniques help to improve classification performance, there are still some challenges and limitations. Feature extraction may be limited by feature selection, resulting in the model not being able to fully mine potential information in the data. In addition, data augmentation may introduce noise or fake data, affecting the generalization ability of the model. For classification depth models, their training and tuning require a lot of computational resources and time, and small data samples also affect model training to some extent. Future research will continue to refine the model and explore additional optimization methods to further enhance the performance and effectiveness of classification. In addition, further improvements can be made to the time-frequency analysis methods, such as optimizing the selection of window functions and window length. Moreover, exploring methods to integrate multimodal information, such as combining chest imaging data and clinical features, can provide more comprehensive information to assist in the classification of abnormal lung sounds. Finally,

although our preliminary study yielded promising results on the ICBHI2017 dataset, we recognize that in order to fully assess the clinical value of the classifier, it is essential to conduct external validation, which will provide important information about the performance of the model in different patient populations and healthcare settings. Due to resource and data access constraints, we have not yet performed external validation on new subject data from different medical centres. However, we see this as an important direction for future work and hope to implement this validation process in future work.

We plan to open source our algorithms and model code on the GitHub platform. You can also contact us through the email 731443570@qq.com.

REFERENCES

[1] B. B. Ken and G. G. Brusselle, "Chronic obstructive pulmonary disease," in *Mucosal Immunology*, 4th ed., J. Mestecky and W. Strober, Eds. New York, NY, USA: Academic, 2015, pp. 1857–1866.

[2] P. Śliwiński and K. Puchalski, "Chronic obstructive pulmonary disease in the awareness of Polish society. Report from the public opinion survey by the Polish Respiratory Society and TNS Polska," *Adv. Respiratory Med.*, vol. 83, no. 1, pp. 1–14, Jan. 2015, doi: 10.5603/piap.2015.0001.

[3] N. Zhong, C. Wang, W. Yao, P. Chen, J. Kang, S. Huang, B. Chen, C. Wang, D. Ni, Y. Zhou, S. Liu, X. Wang, D. Wang, J. Lu, J. Zheng, and P. Ran, "Prevalence of chronic obstructive pulmonary disease in China: A large, population-based survey," *Amer. J. Respir. Crit. Care Med.*, vol. 176, no. 5, pp. 753–760, 2007.

[4] X. Zhang, B. Yu, T. He, and P. Wang, "Status and determinants of health services utilization among elderly migrants in China," *Glob Health Res. Policy*, vol. 3, no. 8, p. 8, 2018, doi: 10.1186/s41256-018-0064-0.

[5] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," in *Proc. Int. Conf. Content-Based Multimedia Indexing (CBMI)*, La Rochelle, France, Sep. 2018, pp. 1–6, doi: 10.1109/CBMI.2018.8516489.

[6] N. Jakovljevic and T. Loncar-Turukalo, "Hidden Markov Model Based Respiratory Sound Classification," in *Proc. Precision Medicine Powered By PHealth and Connected Health*, vol. 66. Singapore, 2018, pp. 39–43.

[7] 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, doi: 10.1109/TBCAS.2020.2981172.

[8] Y. Ma, X. Xu, and Y. Li, "Lungrn+nl: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation," in *Proc. Interspeech*, 2020, pp. 2902–2906. [Online]. Available: https://api.semanticscholar.org/CorpusID:226206978,

[9] M. Bahoura and C. Pelletier, "New parameters for respiratory sound classification," in *Proc. Can. Conf. Electr. Comput. Eng. Toward Caring Humane Technol. (CCECE)*, vol. 3, Montreal, QC, Canada, May 2003, pp. 1457–1460, doi: 10.1109/ccece.2003.1226178.

[10] V. Jindal, V. Agarwal, and S. Kalaivani, "Respiratory sound analysis for detection of pulmonary diseases," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Kolkata, India, Dec. 2018, pp. 293–296, doi: 10.1109/ASPCON.2018.8748284.

[11] M. Bahoura and C. Pelletier, "Respiratory sounds classification using Gaussian mixture models," in *Proc. Can. Conf. Electr. Comput. Eng.*, Niagara Falls, ON, Canada, May 2004, pp. 1309–1312, doi: 10.1109/CCECE.2004.1349639.

[12] 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.*, Taipei, Taiwan, Apr. 2009, pp. 517–520, doi: 10.1109/ICASSP.2009.4959634.

[13] 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)*, Prague, Czech Republic, May 2011, pp. 693–696, doi: 10.1109/ICASSP.2011.5946498.

[14] S. Abbasi, R. Derakhshanfar, A. Abbasi, and Y. Sarbaz, "Classification of normal and abnormal lung sounds using neural network and support vector machines," in *Proc. 21st Iranian Conf. Electr. Eng. (ICEE)*, Mashhad, Iran, May 2013, pp. 1–4, doi: 10.1109/IranianCEE.2013.6599555.

[15] F. B. Jang, J. Yin, and Q. H. He, "Analysis and recognition method for pathological lung sound signal," *Chin. J. Med. Phys.*, vol. 33, no. 7, pp. 739–742, 2016.

[16] B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho, and N. Maglaveras, "A respiratory sound database for the development of automated classification," in *Proc. Precision Medicine Powered by PHealth and Connected Health*, vol. 66. Singapore, Nov. 2017, pp. 33.37, doi: 10.1007/978-981-10-7419-6_6.

[17] B. Sangmin, K. June-Woo, C. Won-Yang, B. Hyerim, S. Soyoun, L. Byungjo, H. Changwan, T. Kyongpil, K. Sungnyun, and Y. Se-Young, "Patch-mix contrastive learning with audio spectrogram transformer on respiratory sound classification," *Audio Speech Process.*, vol. abs/2305, Mar. 2023, Art. no. 14032v2.

[18] S. Gairola, F. Tom, N. Kwatra, and M. Jain, "RespireNet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 527–530, doi: 10.1109/EMBC46164.2021.9630091.

[19] J. Li, J. Yuan, H. Wang, S. Liu, Q. Guo, Y. Ma, Y. Li, L. Zhao, and G. Wang, "LungAttn: Advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram," *Physiol. Meas.*, vol. 42, no. 10, Oct. 2021, Art. no. 105006.

[20] Q. Y. Zhang, and Y. K. Wang, "Speech classification model based on improved Inception network," *J. Comput. Appl.*, vol. 43, no. 3, pp. 90–915, 2023.

[21] Y. Ma, X. Xu, Q. Yu, Y. Zhang, Y. Li, J. Zhao, and G. Wang, "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-ResNet deep learning algorithm," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Nara, Japan, Oct. 2019, pp. 1–4, doi: 10.1109/BIOCAS.2019.8919021.

[22] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2595–2603, Jul. 2021, doi: 10.1109/JBHI.2020.3048006.

[23] H. Chen, X. Yuan, Z. Pei, M. Li, and J. Li, "Triple-classification of respiratory sounds using optimized S-transform and deep residual networks," *IEEE Access*, vol. 7, pp. 32845–32852, 2019, doi: 10.1109/ACCESS.2019.2903859.

[24] R. Phettom, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic identification of abnormal lung sounds using time-frequency analysis and convolutional neural network," in *Proc. 15th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Chiang Mai, Thailand, Oct. 2023, pp. 1–6, doi: 10.1109/icitee59582.2023.10317776.

[25] T.-S. Tsai, C.-C. Hsu, C.-L. Hsieh, and Y.-J. Liu, "Separation of heart and lung sounds by a deep network-based model," in *Proc. IEEE 6th Int. Conf. Knowl. Innov. Invention (ICKII)*, Sapporo, Japan, Aug. 2023, pp. 631–633, doi: 10.1109/ickii58656.2023.10332660.

[26] A. Bohadana, G. Izbicki, and S. Kraman, "Fundamentals of lung auscultation," *New England J. Med.*, vol. 370, no. 8, pp. 744–751, 2014.

[27] G. Serbes, S. Ulukaya, and Y. Kahya, "An automated lung sound preprocessing and classification system based onspectral analysis methods," in *Precision Medicine Powered by PHealth and Connected Health*. Singapore: Springer, 2018, pp. 45–49.

[28] T. Okubo, N. Nakamura, M. Yamashita, and S. Matsunaga, "Classification of healthy subjects and patients with pulmonary emphysema using continuous respiratory sounds," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 70–73.

[29] X. H. Kok, S. Anas Imtiaz, and E. Rodriguez-Villegas, "A novel method for automatic identification of respiratory disease from acoustic recordings," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 2589–2592, doi: 10.1109/EMBC.2019.8857154.

[30] T. Ko, V. Peddinti, D. Povey, et al., "Audio augmentation for speech recognition," in *Proc. Interspeech*, 2015, p. 711. [Online]. Available: https://api.semanticscholar.org/CorpusID:7360763

[31] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.

[32] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[34] Y. Wu and K. He, "Group normalization," 2018, *arXiv:1803.08494*.
[35] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
[36] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[37] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. P. Paiva, I. Chouvarda, and P. de Carvalho, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Meas.*, vol. 40, no. 3, Mar. 2019, Art. no. 035001, doi: 10.1088/1361-6579/ab03ea.
[38] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, and S. Kido, "Automatic classification of large-scale respiratory sound dataset based on convolutional neural network," in *Proc. 19th Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2019, pp. 804–807, doi: 10.23919/ICCAS47443.2019.8971689.
[39] T. Nguyen and F. Pernkopf, "Lung sound classification using cotuning and stochastic normalization," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 9, pp. 2872–2882, Sep. 2022, doi: 10.1109/TBME.2022.3156293.
[40] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, "CNN-MoE based framework for classification of respiratory anomalies and lung disease detection," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 2938–2947, Aug. 2021, doi: 10.1109/JBHI.2021.3064237.

CHENWEN WU received the B.S. degree in mechanical manufacturing from Gansu University of Technology, Lanzhou, China. He was a Visiting Scholar with the Department of Automation, Tsinghua University, Beijing, China. His current research interests include computer networks, medical data analysis, and data mining.

NA YE was born in Shanpingba, Chongqing, China, in 2000. She received the bachelor's degree from Changjiang Teachers College, China, and the master's degree from Lanzhou Jiaotong University, China. Her research interest includes data mining.

JIALIN JIANG was born in Deyang, Sichuan, China, in 1999. She received the B.S. degree from Chengdu University of Information Engineering, China, and the M.S. degree from Lanzhou Jiaotong University, China. Her research interest includes data mining.