

# Auscultation-Based Pulmonary Disease Detection through Parallel Transformation and Deep Learning

**Khan, R., Khan, S. U., Saeed, U. & Koo, I-S**

Published PDF deposited in Coventry University's Repository

**Original citation:**

Khan, R, Khan, SU, Saeed, U & Koo, I-S 2024, 'Auscultation-Based Pulmonary Disease Detection through Parallel Transformation and Deep Learning', *Bioengineering*, vol. 11, no. 6, 586. <https://doi.org/10.3390/bioengineering11060586>

DOI 10.3390/bioengineering11060586




ESSN 2306-5354

Publisher: MDPI

© 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

## Article

# Auscultation-Based Pulmonary Disease Detection through Parallel Transformation and Deep Learning

Rehan Khan <sup>1</sup>, Shafi Ullah Khan <sup>1</sup>, Umer Saeed <sup>2</sup> and In-Soo Koo <sup>1,\*</sup>

<sup>1</sup> Department of Electrical Electronic and Computer Engineering, University of Ulsan, Ulsan 44610, Republic of Korea; rehan.khan.mte@gmail.com (R.K.); shafiukhan98@gmail.com (S.U.K.)

<sup>2</sup> Research Centre for Intelligent Healthcare, Coventry University, Coventry CV1 5FB, UK; saeedu3@uni.coventry.ac.uk

\* Correspondence: iskoo@ulsan.ac.kr

**Abstract:** Respiratory diseases are among the leading causes of death, with many individuals in a population frequently affected by various types of pulmonary disorders. Early diagnosis and patient monitoring (traditionally involving lung auscultation) are essential for the effective management of respiratory diseases. However, the interpretation of lung sounds is a subjective and labor-intensive process that demands considerable medical expertise, and there is a good chance of misclassification. To address this problem, we propose a hybrid deep learning technique that incorporates signal processing techniques. Parallel transformation is applied to adventitious respiratory sounds, transforming lung sound signals into two distinct time-frequency scalograms: the continuous wavelet transform and the mel spectrogram. Furthermore, parallel convolutional autoencoders are employed to extract features from scalograms, and the resulting latent space features are fused into a hybrid feature pool. Finally, leveraging a long short-term memory model, a feature from the latent space is used as input for classifying various types of respiratory diseases. Our work is evaluated using the ICBHI-2017 lung sound dataset. The experimental findings indicate that our proposed method achieves promising predictive performance, with average values for accuracy, sensitivity, specificity, and F1-score of 94.16%, 89.56%, 99.10%, and 89.56%, respectively, for eight-class respiratory diseases; 79.61%, 78.55%, 92.49%, and 78.67%, respectively, for four-class diseases; and 85.61%, 83.44%, 83.44%, and 84.21%, respectively, for binary-class (normal vs. abnormal) lung sounds.

**Keywords:** respiratory sounds; LSTM; mel spectrogram; convolutional autoencoder; artificial intelligence; continuous wavelet transform; hybrid features; healthcare



**Citation:** Khan, R.; Khan, S.U.; Saeed, U.; Koo, I.-S. Auscultation-Based Pulmonary Disease Detection through Parallel Transformation and Deep Learning. *Bioengineering* **2024**, *11*, 586. <https://doi.org/10.3390/bioengineering11060586>

Academic Editors: Larbi Boubchir and Mihaela Hnatiuc

Received: 18 May 2024

Revised: 5 June 2024

Accepted: 6 June 2024

Published: 8 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Globally, lung diseases are acknowledged as highly fatal and dangerous, affecting millions of people every year. According to the Forum of International Respiratory Societies (FIRS), respiratory disorders cause almost four million fatalities annually and are among the leading causes of morbidity worldwide [1]. Furthermore, the World Health Organization (WHO) reported that after cardiovascular diseases, respiratory diseases are the second largest contributor to the global disease burden; approximately 10 million people lose their lives to respiratory diseases every year [2]. The diagnostic procedures for respiratory diseases primarily involve auscultation, wherein medical specialists listen with a stethoscope to the sounds as air moves in and out of the lungs [3]. Lung auscultation is among the traditional diagnostic techniques employed [4] by medical specialists to assess the status of respiratory diseases. Crackles and wheezes are the two most frequently heard abnormal lung sounds [5]. These sounds are identified based on their frequency, pitch, energy, intensity, and duration. Wheezes are continuous, high-pitched noises typically occurring in the 400–500 Hz range with a duration longer than 100 ms. Wheezes are generally heard in individuals with asthma and chronic obstructive pulmonary disease [6]. Crackles are discontinuous sounds with a pitch ranging between 100 and 2000 Hz. Crackles are generally

heard in patients suffering from heart failure, pneumonia, and bronchitis [7]. Auscultation is cost-effective, easy to apply, and provides essential details about lung conditions and symptoms for a quick diagnosis [8]. However, traditional auscultation with a stethoscope is not infallible, because it depends on the clinician's expertise and auditory sensitivity. Sometimes, during an examination, this leads to misclassification, even when carried out by an expert physician [9]. Research by Salvatore and Nieman [10] revealed that more than half of the pulmonary sounds were incorrectly identified by medical trainee students in the hospital. Since lung sounds are non-stationary, it is challenging to distinguish them through traditional auscultation techniques. Therefore, there is a need to develop a respiratory disease detection system to ensure more efficient clinical diagnoses.

The 2017 public respiratory sound dataset released by the International Conference on Biomedical Health Informatics (ICBHI-2017) [11] has attracted significant interest among research teams developing automated systems for distinguishing lung sounds. Deep learning (DL) and conventional machine learning (ML) have been utilized in studies over the last decade to address the classification task [12–14]. Several attempts have been made to develop algorithms and methods for feature extraction aimed at automatically identifying abnormal lung sounds. Among them, some common feature extraction techniques include spectrograms [15], mel spectrograms [16], wavelet coefficients [17], and the mel-frequency cepstral coefficient (MFCC) [18], as well as a wide range of DL and ML approaches.

Pham et al. [4] extracted various features, including the short-time Fourier transform (STFT) and mel spectrogram. Gairola et al. [19] employed a convolutional neural network (CNN), leveraging mel spectrograms to identify adventitious lung sounds. Bardou et al. [20] utilized the MFCC and traditional ML features (such as local binary patterns) for feature extraction, replacing the CNN model with fully connected layers to train these features, and integrating the output of four CNN models with softmax activation. The authors of [21] optimized an AlexNet pre-trained CNN model, utilizing scalograms to extract a visual representation of the pixel values to accurately detect and classify lung sounds. Tariq et al. [22] developed a model that concatenates three distinct features (a chromagram, the MFCC, and a spectrogram) to classify lung audio samples using ideal CNN models. Similarly, the study in [23] presented various feature extraction techniques to classify different respiratory diseases such as COPD and asthma.

In addition to lung sound analysis, other research has utilized methods such as the wavelet transform and the spectrogram [24], or empirical mode decomposition (EMD) and bandpass filtering for scale selection, as well as processing continuous wavelet transform (CWT)-based scalogram representations with a lightweight CNN for classification of various respiratory diseases. Recent advancements in noninvasive monitoring have led to significant progress in deriving respiratory signals from ECG data, thereby enhancing traditional respiratory sound analysis. Yi and Park [25] demonstrated the derivation of respiratory signals using wavelet transforms directly from the ECG, establishing a foundation for reliable respiratory monitoring without the subject's awareness. O'Brien and Heneghan [26] presented a comparative examination of methods for extracting respiratory signal extraction approaches from the ECG, highlighting the accuracy and robustness of these techniques across various body postures during sleep. Furthermore, Campolo et al. [27] introduced a novel technique employing EMD to derive respiratory signals from the ECG, showcasing its superior performance in accurately reconstructing respiratory waveforms. This approach offers a dual-modality method that enhances diagnostic capabilities by simultaneously analyzing cardiac and respiratory data. In [28], the authors classified electroencephalogram (EEG) signals using CWT and a long short-term memory (LSTM) model, similar to the study in [29], in which a dual scalogram comprising the Stockwell transform and a CWT scalogram was employed for fault diagnosis in centrifugal pumps. Furthermore, recent studies have explored different ML and DL techniques for binary-class (normal vs. abnormal) classification and multi-class classification of respiratory diseases [30].

In order to achieve improved performance for multi-class and binary classification tasks, Nguyen and Pernkopf [31] developed approaches that include sample padding, feature splitting, an ensemble of CNNs, and a focal loss objective. Acharya and Basu [15] introduced a deep hybrid-based CNN and a recurrent neural network (RNN) framework for detecting respiratory sounds utilizing mel spectrograms. Concurrently, Demir et al. [32] identified four different lung sounds by combining deep CNN features with a linear discriminant analysis and random subspace ensemble classifier. Additionally, to resolve imbalances in the training data, Petmezas et al. [33] employed a model combining a CNN with LSTM networks that include the focal loss function. The respiratory sound cycle was transformed into a time-frequency representation and processed using the CNN.

In this study, we propose a hybrid DL technique with signal processing techniques for detecting various lung disorders. We introduce parallel transformation for rich features using a parallel convolutional autoencoder (CAE). Initially, the auscultation recordings undergo preprocessing through segmentation of respiratory cycles, followed by a padding technique to modify the length of each respiratory cycle to a fixed size. The respiratory cycle audio signal is transformed into a time-frequency representation using CWT and a mel spectrogram. Two parallel CAEs extract rich features from scalograms, concatenate features in a hybrid pool, and subsequently feed them into an LSTM model that indicates different respiratory diseases.

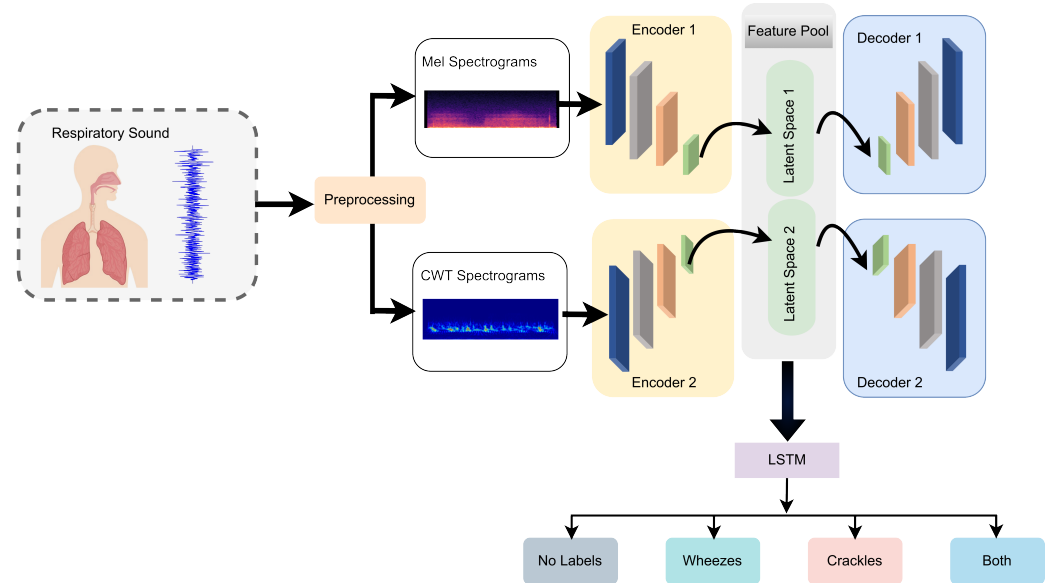
Our principal contributions are as follows:

- (1) We present a novel method that combines deep learning and signal processing for enhanced lung auscultation analysis and classification. This approach addresses the limitations of traditional techniques utilized for lung auscultation.
- (2) This approach utilizes parallel transformation using both CWT and a mel scalogram. A parallel CAE is utilized to extract rich features from the scalograms transformed by CWT and mel at latent spaces.
- (3) A hybrid feature pool is created by fusing the features collected from both the CWT and mel scalograms using CAE latent spaces. These latent spaces provide an extensive and enriched representation of lung sound features, enhancing the analysis and classification approach.
- (4) An LSTM network is employed to classify various lung sounds, leveraging its proficiency in handling time-series data. Lung sounds are sequential, and LSTM is particularly suited to recognizing complex patterns and handling sequential information in time-dependent data.

The rest of this paper is organized as follows. Section 2 provides background information on the dataset and comprehensive details of the proposed model. Section 3 describes the experimentation and the model's performance. Finally, Section 4 summarizes the proposed study along with future expansion and enhancement planned for this work.

## 2. Materials and Methods

The framework of the proposed study for lung sound classification utilizes a hybrid model that combines an autoencoder with a recurrent neural network, specifically the LSTM variant, as illustrated in Figure 1. Initially, all lung sounds are preprocessed to segment the respiratory cycles, ranging from 0.2 s to 16 s, with an average duration of 2.7 s. The respiratory cycles in the dataset are not equal in length, so to address this issue, a padding technique is utilized. Each cycle is preprocessed until the total length equals six seconds. Following this, the cycles are transformed into a dual time-frequency domain using CWT and the mel spectrogram to provide distinct representations of each cycle. Subsequently, these time-frequency spectrogram images are fed into parallel CAEs for feature extraction, and the resulting latent spaces of the parallel CAEs are fused into a hybrid feature pool. Finally, the resulting features from the latent spaces are used as input for the LSTM model to classify various types of respiratory diseases.



**Figure 1.** Framework of the proposed method.

### 2.1. Dataset

In this study, the publicly available ICBHI-2017 respiratory sound dataset from the International Conference on Biomedical Health Informatics [11] was utilized. The dataset was collected at two different hospitals in Greece and Portugal by teams of experts. The data were acquired using digital stethoscopes (the AKG-C417 L Microphone, the 3M Littmann Classic II SE, the 3M Littmann 3200, and the Welch Allyn Meditron Master Elite), which have different sampling frequency ranges of 4, 10, and 44.1 kHz. The dataset comprises annotated respiratory cycle recordings totaling 5.5 h. In total, 920 audio samples were collected at various anatomical locations from 126 individuals [34]. The recordings were obtained from healthy individuals and others with a range of pathological conditions, including seven lung diseases: pneumonia, a lower respiratory tract infection (LRTI), asthma, bronchiectasis, an upper respiratory tract infection (URTI), bronchiolitis, and COPD. Furthermore, all the respiratory cycles were annotated based on the presence of crackles and/or wheezes [35]. Wheezes are a type of abnormal, continuous, high-pitched breathing sound primarily associated with chronic disease. In contrast, crackles are discontinuous lung sounds of shorter duration, heard during both the inspiratory and expiratory phases. The duration is notably shorter in the total respiratory cycle and is mainly associated with non-chronic diseases [36].

### 2.2. Preprocessing and Data Augmentation

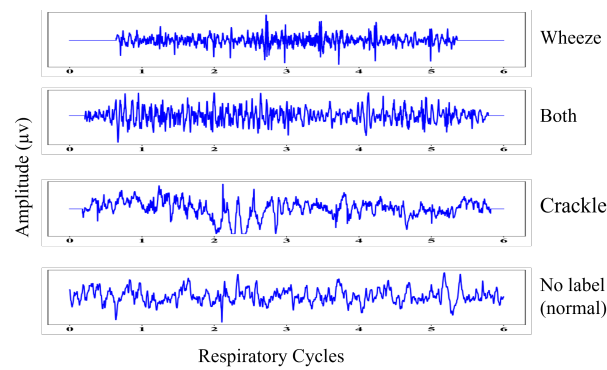
Each respiratory cycle was annotated by an expert as belonging to one of four classes: normal (N), crackle (C), wheeze (W), and both crackle and wheeze (B). The start times, end times, number of crackles, and number of wheezes are shown in Table 1. The source database contains 6898 respiratory cycles, including 1864 with crackles, 886 with wheezes, 3642 with no labels (i.e., from healthy individuals), and 506 with both crackles and wheezes, as shown in Table 2. The duration of the respiratory cycles is not fixed. Although training DL models is possible by utilizing adaptive average pooling, this approach performs poorly in comparison with a fixed-size signal [19]. The length of the audio signals in the dataset varies, so zero padding was employed to achieve a fixed duration of six seconds. Padded samples of respiratory cycles are shown in Figure 2. Data augmentation techniques were utilized to artificially expand the unbalanced dataset by modifying the audio samples, resulting in several modified versions of the dataset, as shown in Table 3.

**Table 1.** Description of the dataset.

Cycle	Start Time (s)	End Time (s)	Crackles	Wheezes
01	1.018	3.411	1	0
02	3.411	5.827	1	0
03	5.827	8.339	1	1
04	8.339	10.923	1	0
05	10.923	13.292	0	1
06	13.292	16.018	1	0
07	16.018	18.482	1	0
08	18.482	19.542	1	0

**Table 2.** Respiratory cycles of lung sounds from ICBHI-2017.

Sound	Original	Augmented
Crackle	1864	4150
Wheeze	886	3544
No label	3642	5666
Both sounds	506	2024
Total	6898	15,384



**Figure 2.** Sample signals of lung sounds.

**Table 3.** Respiratory cycles of lung sounds from ICBHI-2017 for distinct lung diseases.

Disease	Original	Augmented
Asthma	06	24
Bronchiectasis	104	416
Bronchiolitis	160	640
COPD	3642	5746
Healthy	322	1288
LRTI	32	128
Pneumonia	285	1140
URTI	243	972
Total	6898	10,354

The time-domain audio data augmentation approaches employed to enlarge our audio samples were as follows:

- (1) **Time Stretching:** This technique involves either increasing or decreasing the sample speed by specific factors [37]. In this work, we augmented the minority class by stretching the respiratory audio signals along with their temporal variations at a stretching rate of 1.2. The length of an audio signal was adjusted based on this rate, calculated by multiplying the original length of the audio by the stretching rate. The time stretching method is useful for modifying the audio’s temporal properties



without altering its pitch, which makes it effective for enhancing datasets that have a low representation of particular class samples. A more balanced dataset with improved temporal diversity in the audio signals is the anticipated outcome.

- (2) **Pitch Shifting:** This technique involves modifying the lung sound signal by increasing or decreasing the pitch while keeping the audio signal duration constant. In [38], the significance of the pitch-shifting process was investigated for CNN-based sound classification. To enlarge minority classes in the dataset, we employed pitch shifting by randomly shifting the audio signals along the time axis by a maximum percentage value of 0.2. Pitch shifting changes the audio's frequency content, adding additional variances that enhance the model's ability to generalize. An expanded range of pitch-modified samples is anticipated, which will strengthen and balance the training dataset.
- (3) **Adding Noise:** To further augment the dataset, noise was added to the recordings to increase the sample sizes of minority classes. Noise was introduced from within the function, and a noise vector was generated using a Gaussian distribution of zero mean and unit variance with a length matching the input audio signal. By scaling this noise vector by a factor of 0.005, the amplitude of the noise could be controlled to achieve the desired augmentation. The scaled noise vector was then added element-wise to the original signals, resulting in augmented samples. Using this noise-adding technique effectively enhances the model's resistance to noise and other fluctuations found in real-world recordings.

### 2.3. Transformation of Lung Sounds

During preprocessing, time-frequency analysis is performed to transform the audio sample into a parallel scalogram. Instead of directly feeding the audio signals into the classification model, we first transform them into a spectrogram from the time-series domain to the time-frequency domain. Transformation is a crucial technique for transforming the audio lung samples into a time-frequency domain, specifically into parallel spectrograms. STFT is applied to the time-domain signal,  $S(\tau)$ , to compute the spectrogram using Equations (1) and (2), where  $t$  denotes the time localization and  $W(\tau - t)$  is the window function that cuts and filters the signal [22]. The angular frequency is denoted by  $\omega$ , and  $j$  is the imaginary unit, defined as the square root of  $-1$ . This process facilitates detailed analysis of lung sound signals, providing a comprehensive feature set for the subsequent classification task.

$$\text{Spectrogram}(t, \omega) = |\text{STFT}(t, \omega)|^2 \quad (1)$$

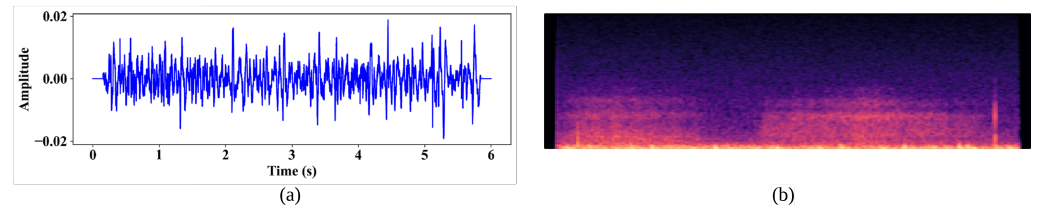
$$\text{STFT}(t, \omega) = \int_{-\infty}^{\infty} S(\tau) \cdot W(\tau - t) \cdot e^{-j\omega\tau} d\tau \quad (2)$$

#### 2.3.1. Mel Scalogram

The mel spectrogram, the human auditory system, and scientific research on speech processing are the sources of inspiration for the mel scale. The human ear is more sensitive to variations in lower frequencies than in higher ones and perceives loudness on a logarithmic scale as opposed to a linear one. Transforming the lung sound sample using STFT converts the signal from the time domain to the frequency domain at a sampled frequency of 4000 Hz. A two-dimensional (2D) image is generated, where columns represent time (windows) and rows represent frequencies in the mel scale. Each value in the image corresponds to the signal's log amplitude for a specific frequency and set of time windows. The time domain is transformed into the frequency domain via STFT. Then, the frequency is mapped to the mel scale and the color dimension to the amplitude [39]. Equation (3) is used for calculating the mel scale, where  $f$  represents the frequency:

$$M = 2595 \log(1 + f/700) \quad (3)$$

We obtain the log mel spectrogram after computing the logarithm values to condense the dynamic range. The mel spectrogram provides an intricate representation of the power spectrum, showing the energy distribution across frequencies over time. The log value of the energy is expanded in the time domain to generate the mel spectrogram. Figure 3a, b, respectively, illustrate the respiratory cycle sound signal and the mel spectrogram of a respiratory cycle.



**Figure 3.** (a) The respiratory audio signal, and (b) the mel spectrogram respiratory audio signal.

### 2.3.2. CWT Scalogram

CWT is an effective technique for signal processing and is used for analyzing non-stationary signals, including audio signals. Within the context of respiratory sound analysis, CWT provides a robust method for extracting relevant features that capture variations in frequency content over time. Respiratory sounds are recorded using various stethoscopes, producing non-stationary signals. The wavelet transform preserves temporal resolution and computationally analyzes non-stationary signals by decomposing them into different frequency components. The wavelet transform utilizes fundamental operations known as wavelets, enabling simultaneous analysis in both frequency and time domains. The mathematical expression for the wavelet transform is shown in Equation (4):

$$WT(s, t) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} f(\tau) \psi^* \left( \frac{\tau - t}{s} \right) d\tau \tag{4}$$

where  $f(\tau)$  represents the time-frequency domain of the input signal,  $\psi^*(\cdot)$  is the conjugate of the wavelet function scaled by factor  $s$ , and the translation factor correlating with the time adjustments is denoted by  $t$ , where the scale factor  $s > 0$ . The multi-resolution capabilities of CWT are particularly advantageous for deciphering time-frequency signals since various physiological events may manifest at various scales. The mathematical representation of the CWT details the relationship between the wavelet  $\psi(t)$  and function  $\psi(t)$ , as shown in Equation (5).

$$CWT_f(s, t) = \int_{-\infty}^{\infty} f(\tau) \psi_{s,t}^*(\tau) d\tau \tag{5}$$

A complex Morlet wavelet is employed as a mother wavelet,  $\psi(t)$ , in Equation (6):

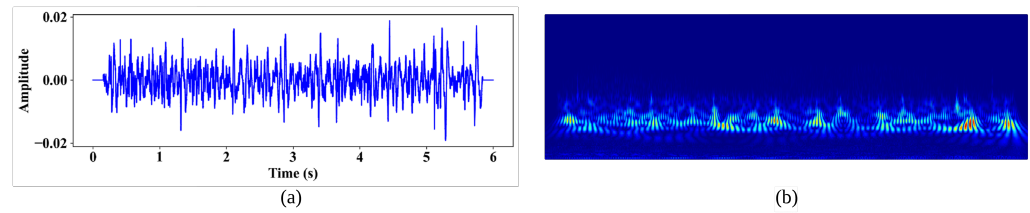
$$\psi_{s,t}(\tau) = \frac{1}{\sqrt{|s|}} \psi \left( \frac{\tau - t}{s} \right) d\tau \tag{6}$$

where  $s$  is the scaling factor and  $t$  is the translation factor that adjusts the function in time, determining whether it is stretched or compressed, depending on whether  $s > 1$  or  $0 < s < 1$ . The normalizing term,  $\frac{1}{\sqrt{|s|}}$ , ensures that the wavelet energy remains constant across all scales. Equations (4) and (5) are utilized to translate and scale the original mother wavelet,  $\tau$ , for analysis of a signal at various frequencies and time positions. CWT converts the lung audio signal into images using the Morlet wavelet.

In fact, a complex sinusoid with Gaussian windows forms the complex Morlet wavelet, and the wavelet transform’s best time localization is achieved via its second-order exponential decay. Moreover, the complex Morlet wavelet function is particularly suited to capturing coherence between harmonic frequencies, providing information on both amplitude and phase. CWT with the Morlet wavelet as the mother wavelet allows for the



extraction of detailed images of the lung sound wave spectrum, demonstrating temporal resolutions. Figure 4a and b, respectively, show the respiratory cycle audio signal and a CWT image of the respiratory cycle.



**Figure 4.** (a) The respiratory audio signal, and (b) the CWT spectrogram of the respiratory audio signal.

### 2.4. Convolutional Autoencoders

CAEs have garnered significant attention in recent years owing to their ability to learn hierarchical representations of data, particularly in image processing tasks. Initially introduced by Theis et al. [40] and Ballé et al. [41], CAEs are specialized neural networks designed to encode and decode spatially hierarchical inputs such as images. CAEs use convolutional layers to leverage spatial locations in data, making them particularly adept at processing images. The primary goal of a CAE is to approximate an identity function while abiding by particular limitations, such as hidden layers having a certain number of neurons. In a CAE, the encoder functions as a funnel, mapping the input,  $x \in \mathbb{R}^n$ , to a latent space. The input consists of  $n$  feature maps,  $x \in \mathbb{R}^{n \times l \times l}$ , originating from the first layer, where each feature map covers  $l \times l$  pixels, and the output layer contains  $m$  feature maps involving convolutional kernels. The dimensions of the convolutional kernel are  $d \times d$  with  $d \leq l$ .

The process begins by encoding the input image, which is segmented into  $d \times d$  pixel patches, labeled as  $x_i$ , where  $i = 1, 2, 3, \dots, p$ . For each patch, the input image is extracted, and convolution operations are carried out using weight  $w_j$  of the  $j^{\text{th}}$  convolution kernel, resulting in neuron values  $o_{ij}$  for  $j = 1, 2, 3, \dots, m$  in the output layer:

$$o_{ij} = f(x_i) = \sigma(w_j \cdot x_i + b) \tag{7}$$

The nonlinear activation function is represented by the symbol  $\sigma$ . In this study, the rectified linear unit (ReLU) activation function is used:

$$ReLU(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{8}$$

After convolutional decoder output  $o_{ij}$  is processed through encoding,  $x_i$  is reconstructed using  $o_{ij}$  to obtain  $\hat{x}_i$ :

$$x_i = f'(o_{ij}) = \varphi(w_i \cdot o_{ij} + \hat{b}) \tag{9}$$

The CAE layer is optimized through the iterative refinement of weights and errors using stochastic gradient descent. These optimized parameters are used to create the feature maps. For every instance,  $\hat{x}_i$  is formed after convolutional encoding and decoding. The reconstruction process involves patches  $p$ , each with a size of  $d \times d$ , and the mean square error between the reconstructed patch,  $\hat{x}_i$ , and the original input picture patch,  $x_i$ , where  $i = 1, 2, 3, \dots, p$ . Equation (10) presents the cost function in its unique form, while Equation (11) elaborates on the reconstruction error [42].

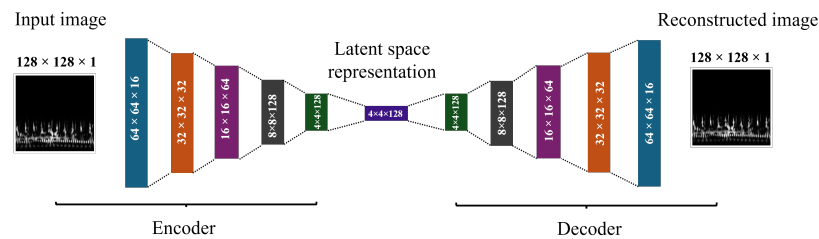
$$JC(\theta) = \frac{1}{p} \sum_{i=1}^p L[x_i, \hat{x}_i] \tag{10}$$

$$LC[x_i, \hat{x}_i] = \|x_i - \hat{x}_i\|^2 = \|x_i - \varphi(\sigma(x_i))\|^2 \tag{11}$$

In the proposed model, optimization techniques like backpropagation are used to minimize loss when a multiple-layer encoder and decoder are employed to train a CAE. The classification model takes the scalogram as input and passes it through a series of convolutional layers. These layers isolate key features by gradually reducing the dimensionality of the image. Following encoding, the model undergoes a decoding stage to reconstruct the image into its original state. Table 4 outlines the CAE layers, along with their respective input and output dimensions, as applied in this study. The efficiency of the CAE is shown by high peak signal-to-noise ratio values for CWT and mel at 56.66 dB and 71.01 dB, respectively, indicating accurate image reconstruction. The layers of the working architecture are depicted in Figure 5.

**Table 4.** The convolutional autoencoder layers.

Encoder		Decoder	
Layer	Output	Layer	Output
Conv2D	128 × 128 × 1	Dense	2048
Conv2D	64 × 64 × 16	Reshape	4 × 4 × 128
Conv2D	32 × 32 × 32	Conv2DTranspose	8 × 8 × 128
Conv2D	16 × 16 × 64	Conv2DTranspose	16 × 16 × 64
Conv2D	8 × 8 × 128	Conv2DTranspose	32 × 32 × 32
Conv2D	4 × 4 × 128	Conv2DTranspose	64 × 64 × 16
Flatten	2048	Conv2DTranspose	128 × 128 × 1
Trainable parameters: 147,842			
Non-trainable parameters: 0			
Total parameters: 147,842			



**Figure 5.** The convolutional autoencoder model architecture.

### 2.5. Long Short-Term Memory

LSTM was first proposed in 1997 by Hochreiter and Schmid Huber, and the improved RNN model has gained substantial interest for time-series data owing to its specialized cellular architecture [43].

Typically, an LSTM architecture consists of an input gate, an output gate, a forget gate, and a memory cell. The forget gate initially determines which informational segments the cell states should discard, and it is expressed mathematically as follows:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \tag{12}$$

where  $x_t$  is the current input;  $h_{t-1}$  is the previous hidden layer output;  $W$  and  $b$  represent the weight matrix and bias, respectively; and  $\sigma$  is the sigmoid activation. The input gate subsequently controls the retention of data in the cell state by dividing them into two parts, determining which data need to be updated, and configuring the updated state. The following are the mathematical expressions:

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{13}$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (14)$$

The output gate plays a crucial role in deciding the final output. The segments of the cell state for the output are determined by the sigmoid function, followed by pointwise multiplication with the output of the tanh function:

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (15)$$

$$h_t = o_t \times \tanh(C_t) \quad (16)$$

In the field of biomedicine, LSTMs have shown the capability to recognize time-based patterns, which is particularly useful for the diagnosis of respiratory diseases characterized by detailed time-based patterns. In this study, the LSTM model employs 64 units to process the time-based patterns of the respiratory sound data. Subsequently, the data pass through a dense layer that utilizes a softmax activation function to categorize the LSTM output into specific categories. The model is optimized for best categorization results using the categorical cross-entropy loss function and the “Adam” optimizer. The LSTM architecture employed in this research is detailed in Table 5.

**Table 5.** The LSTM architecture.

Layer Type	Output Shape	Parameters
lstm (LSTM)	(None, 64)	147,712
dense 4 (dense)	(None, 2)	130
Total parameters: 147,842		
Trainable parameters: 147,842		
Non-trainable parameters: 0		

### 3. Results and Discussion

In this study, a publicly available dataset of respiratory sounds was chosen to evaluate the performance of the proposed framework [11]. The proposed DL framework was developed and implemented in Python 3.9.18, leveraging TensorFlow 2.15.0 as the foundation for the Keras library. All experiments were conducted using a desktop computer with an AMD Ryzen 9 5900X 12-Core 3.70 GHz CPU, 64 GB of RAM, and an NVIDIA GeForce RTX 3080 GPU with 64 GB of memory. The respiratory sound dataset encompasses four sub-tasks, which include a binary-class problem distinguishing between normal (N) and abnormal (Ab) samples. Three-class and four-class tasks categorize respiratory cycles into one of four classes (W, C, N, and B). Eight-class categorization is also performed, where classifications include healthy samples and seven distinct lung diseases: pneumonia, LRTI, asthma, bronchiectasis, URTI, bronchiolitis, and COPD. The dataset was split, allocating 80% for training and 20% for testing. After evaluating the proposed model for the binary-class problem, the experimentation was extended to the three-class, four-class, and eight-class problems. We used several metrics to evaluate the performance of respiratory sound classification: accuracy, F1-score, precision, and sensitivity. These metrics collectively provide a nuanced view of the model’s ability to correctly identify and differentiate between the various respiratory diseases. In the classification framework, true positive (TP) is when an instance was accurately identified as positive, and true negative (TN) means an instance was accurately identified as negative. A false positive (FP) is an instance incorrectly identified as positive, and a false negative (FN) is a positive instance incorrectly labeled as negative. The following equations are used to calculate these metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$\text{F1-score} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \tag{18}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{19}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{20}$$

In this study, the proposed model, based on a hybrid approach involving digital signal processing and DL, was evaluated using various classification tasks to assess its effectiveness in distinguishing various respiratory diseases. This evaluation was conducted across multiple classification tasks ranging from simple binary classification problems to three-class, four-class, and eight-class problems. The following are the specific scenarios for each classification problem:

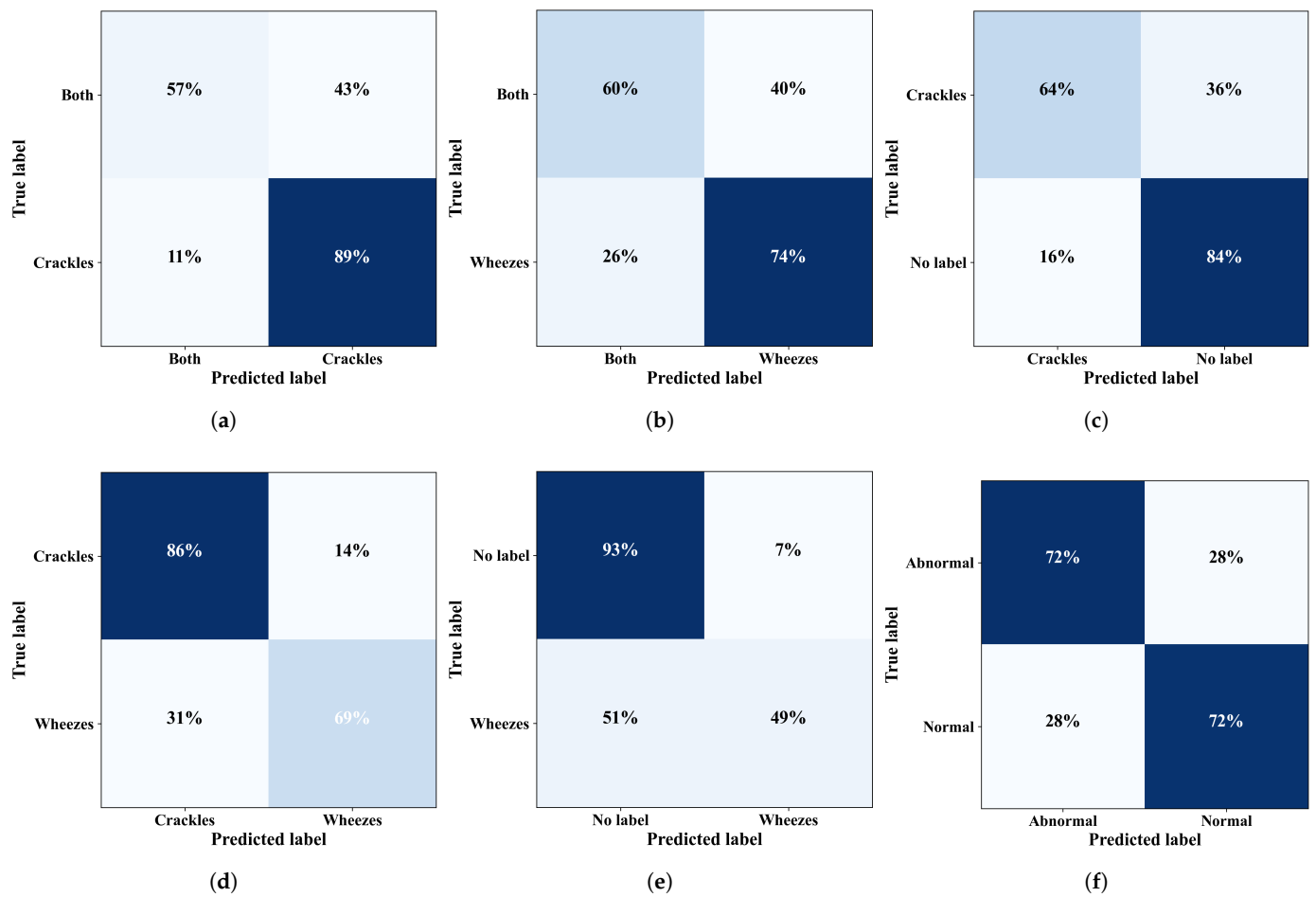
- Binary-class problems: N-Ab, C-W, B-C, B-W, C-N, C-W, and W-N.
- Three-class problems: B-C-W, and C-N-W.
- Four-class problems: C, W, N, and B.
- Eight-class problems: Healthy (H), pneumonia (P), LRTI (L), asthma (A), bronchiectasis (B1), URTI (U), bronchiolitis (B2), and COPD (C).

### 3.1. Binary Classification

In the binary classification problems, the proposed model demonstrated remarkable accuracy in identifying crucial respiratory sounds. Our model exhibited remarkable performance in differentiating between C, W, N, and B. Several experiments were conducted on both the official and the augmented datasets to validate the effectiveness of our proposed model. For the N-Ab problem, our model achieved an average accuracy of 85.61%, an F1-score of 84.21%, a precision of 85.36%, and a sensitivity of 83.44%. Similarly, for the B-C problem, the results were 94.41%, 93.65%, 93.57%, and 93.74% for accuracy, F1-score, precision, and sensitivity, respectively. For the C-W problem, the results were 93.57%, 93.51%, 93.50%, and 93.53%, respectively. The results for the remaining binary-class problems are shown in Table 6. Figure 6 depicts the confusion matrices, showing the predicted versus the true labels for different binary-class problems.

**Table 6.** Performance metrics for binary-class problems.

Class	Accuracy	F1-Score	Precision	Sensitivity
Non-augmented dataset				
C-W	≈81%	≈78%	≈78%	≈78%
B-C	≈81%	≈68%	≈69%	≈67%
B-W	≈69%	≈67%	≈67%	≈68%
C-N	≈78%	≈75%	≈75%	≈74%
W-N	≈85%	≈73%	≈76%	≈72%
N-Ab	≈67%	≈67%	≈67%	≈67%
Augmented dataset				
C-W	≈94%	≈94%	≈94%	≈94%
B-C	≈94%	≈94%	≈94%	≈94%
B-W	≈94%	≈94%	≈93%	≈94%
C-N	≈86%	≈88%	≈86%	≈85%
W-N	≈90%	≈89%	≈89%	≈99%
N-Ab	≈86%	≈84%	≈85%	≈84%



**Figure 6.** Confusion matrices for binary-class problems: (a) B-C, (b) B-W, (c) C-N, (d) C-W, (e) N-W, and (f) N-Ab.

### 3.2. Three-Class Classification

After achieving promising results for the binary-class problems, we extended our evaluation to three-class classification problems. We further examined and compared the internal relationships and variations for the B-C-W and C-N-W problems. On the official and augmented datasets, our proposed model achieved an average accuracy of 89.45%, an F1-score of 88.41%, a precision of 88.68%, and a sensitivity of 88.16% for the B-C-W problem. For the C-N-W problem, the results were 82.04%, 82.15%, 81.94%, and 82.41%, respectively, as shown in Table 7. Figure 7 presents the confusion matrices for the B-C-W and C-N-W three-class problems.

**Table 7.** Performance metrics for three-class problems.

Class	Accuracy	F1-Score	Precision	Sensitivity
Non-Augmented dataset				
B-C-W	≈71%	≈63%	≈64%	≈62%
C-N-W	≈67%	≈60%	≈61%	≈53%
Augmented dataset				
B-C-W	≈90%	≈89%	≈89%	≈88%
C-N-W	≈82%	≈82%	≈82%	≈83%

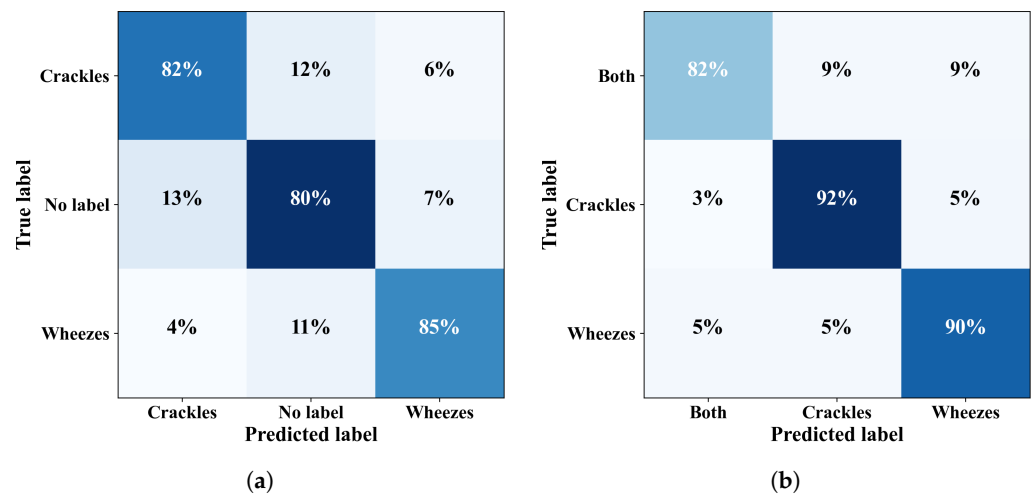


Figure 7. Confusion matrices for three-class problems: (a) C-N-W, and (b) B-C-W.

### 3.3. Four-Class Classification

To evaluate the model’s ability to identify four-class respiratory sound problems, both datasets were used to compare the C, W, N, and B categories. The model demonstrated promising performance across all scenarios, as illustrated by the confusion matrix in Figure 8a. The proposed model achieved an average accuracy of 79.61%, an F1-score of 78.67%, a precision of 78.86%, and a sensitivity of 89.56% on the augmented dataset, as shown in Table 8.

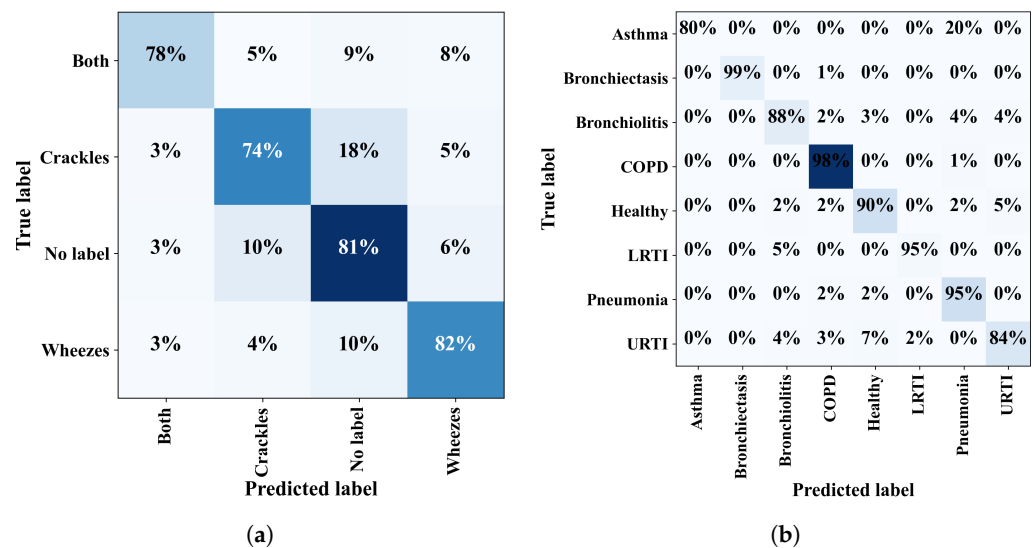


Figure 8. Confusion matrices for (a) four-class problems, and (b) eight-class problems.

Table 8. Performance metrics for four-class problems.

Class	Dataset	Accuracy	F1-Score	Precision	Sensitivity
Four-Class	Non-augmented	≈64%	≈54%	≈55%	≈53%
Four-Class	Augmented	≈80%	≈79%	≈79%	≈79%

### 3.4. Eight-Class Classification

Finally, the evaluation of the proposed framework for eight-class problems included healthy samples and seven distinct lung diseases (P, L, A, B1, U, B2, and C), as shown in Table 9. The confusion matrix in Figure 8b illustrates that the model yielded an overall



accuracy of 94.16%, a sensitivity of 89.56%, an F1-score of 89.56%, and a precision of 89.87%. In summary, these findings demonstrate the proposed model’s robust and reliable performance across various respiratory sound classification scenarios. including binary-class, three-class, four-class, and eight-class problems, even on unbalanced datasets.

**Table 9.** Performance metrics for eight-class problems.

Class	Dataset	Accuracy	F1-Score	Precision	Sensitivity
Eight-Class	Non-augmented	≈93%	≈61%	≈61%	≈63%
Eight-Class	Augmented	≈94%	≈90%	≈90%	≈90%

### 3.5. Discussion

We proposed a novel approach to evaluating various adventitious lung sounds by employing a hybrid model that combines parallel CAEs and an LSTM network. The model’s performance was evaluated across multiple classification problems: binary-class, three-class, and four-class problems, as well as eight-class problems involving healthy samples and seven distinct diseases. In this study, lung sound signals were not directly fed into the classification model—all lung sound signals were transformed into the frequency domain as spectrograms. For feature extraction, dual CWT and mel transformations were fed into parallel CAEs, and the features extracted from CAE latent spaces were concatenated to create a hybrid feature pool. This parallel transformation allows for more precise extraction of rich features, while fusion improves data classification by efficiently capturing diverse signal characteristics. The sequential nature of LSTM is utilized for the classification of various diseases. To assess the impact of hybrid features from the CAE latent space features from both CWT and the mel spectrogram, we conducted an ablation study using an eight-class classification framework. The results of training the LSTM network with various feature sets are shown in Table 10. When solely CAE latent space features of CWT were used, the LSTM model achieved an average accuracy of 78.50%, an F1-score of 82.14% , a precision of 85.34%, and a sensitivity of 80.42%. In contrast, training with only latent space features from the mel spectrogram resulted in an average accuracy of 90.83%, an F1-score of 85.7%, a precision of 88.31%, and a sensitivity of 84.59%. However, the model’s performance significantly improved when combining both the CAE latent space features, with the accuracy rising to 94.69%, F1-score to 90.69%, precision to 91.89%, and sensitivity to 89.78%. This shows that the fusion of both CAE latent space features significantly improves the LSTM network’s capacity to classify and detect various respiratory disorders in multi-class problems. Table 11 illustrates the overall performance of our proposed model in multiple-class tasks using a publicly available respiratory disease dataset

**Table 10.** Ablation experiment.

Features	Accuracy %	F1-Score %	Precision %	Sensitivity %
CWT latent space features	87.78	82.14	85.34	80.42
Mel latent space features	90.83	85.72	88.31	84.59
Combined features	94.69	90.68	91.89	89.78

**Table 11.** Comparison between the proposed model and already established works.

Study	Class	Method	Performance %			
			Accuracy	Sensitivity	Specificity	F1-Score
Demir et al. [32]	4	CNN, LDA	≈71	≈61	≈86	≈65
Lie et al. [44]	4	ARCS-NET	-	≈41	≈67	≈57
	2		≈80	≈81	≈80	-
Petmezas et al. [33]	4	CNN-LSTM	≈76	≈53	≈85	≈69
Demir et al. [45]	4	VGG-16, SVM	≈66	≈53	≈83	≈55

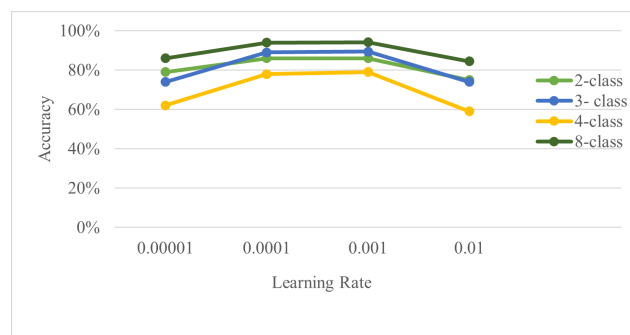
Table 11. Cont.

Study	Class	Method	Performance %			
			Acc.	Sen.	Spe.	F1-Score
Ma et al. [46]	4	Bi-ResNet	≈53	≈31	≈69	≈50
	2		-	≈48	≈69	≈49
Chambres et al. [47]	4	HMM, NLSp	≈50	≈21	≈78	≈50
	2		-	≈33	≈78	≈56
Rocha et al. [48]	4	LDA	≈61	≈52	≈66	≈59
Acharya and Basu [15]	4	CNN, RNN	-	≈49	≈84	≈67
Mang et al. [49]	4	Cochleagram, CNN	≈63	≈53	≈69	≈61
Wanasinghe et al. [50]	6	Mel, MFCC, CNN	≈93	≈92	≈98	≈93
Choi et al. [51]	10	Mel, CNN	≈90	-	-	-
Li et al. [52]	4	TQWT, STFT	-	≈37	≈72	≈54
	2		-	≈52	≈72	≈62
<b>Proposed Model</b>						
ICBHI dataset	A-B1-C-U-L-B-P-H	CAE, LSTM	≈94	≈90	≈99	≈90
	C-W-N-B		≈80	≈79	≈93	≈79
	B-C-W		≈90	≈88	≈95	≈89
	C-N-W		≈82	≈83	≈91	≈82
	N-AB		≈86	≈84	≈84	≈85
	C-W		≈94	≈95	≈94	≈94
SJTU dataset	B-C	≈95	≈94	≈94	≈94	
	C-F-N-R-S-W-B	CAE, LSTM	≈82	≈39	≈92	≈41
	N-AB	≈84	≈75	≈75	≈76	

The overall accuracy, sensitivity, specificity, and F1-score for the eight-class problems were 94.16%, 89.56%, 99.10%, and 89.5%, respectively. Similarly, for the four-class problems, the overall results were 79.61%, 78.55%, 92.49%, and 78.67%, respectively, and for the three-class problems, the overall results were 89.45%, 88.16%, 94.54%, and 88.41%, respectively. Meanwhile, for the binary-class problems, the overall results for normal vs. abnormal were 85.61%, 83.44%, 83.44%, and 84.21% for accuracy, sensitivity, specificity, and F1-score, respectively, and for crackles and wheezes, they were 84.21%, 93.57%, 93.53%, and 93.15%, respectively. To further validate the robustness of our framework, we also conducted experiments using another public dataset, the SJTU Paediatric dataset [53], for various respiratory diseases, including healthy samples and seven distinct lung diseases: coarse crackle (C), fine crackle (F), rhonchi (R), stridor (S), wheeze (W), and both wheeze and crackle (B). The results, presented in Table 11, demonstrate that our findings are not only applicable to a single dataset but also generalize well across different datasets. This additional validation underscores the generalizability of our model, reinforcing its effectiveness on diverse datasets. The variations in the error rates are associated with the imbalanced nature of the dataset, where some classes are over-represented, influencing the model's learning bias. Furthermore, the inherent acoustic similarities across various respiratory disorders make it more complex for the model to correctly identify the lung sound. For example, high-pitched sounds like crackles and wheezes provide a special problem since their slight acoustic variances are hidden behind similar spectral sequences.

Several experiments were performed to optimize the proposed model. Specifically, performance was evaluated while varying the learning rate and the number of epochs. Figure 9 shows the classification accuracies across different learning rates ranging from 0.00001 to 0.01 over 200 epochs. The results indicate that for the binary-class problems, the accuracy remained high as the learning rate increased from 0.00001 to 0.001. For the three-class problems, a slight decline in accuracy was observed as the learning rate increased. For the four-class problems, increasing the learning rate noticeably reduced the model's accuracy after the initial increase, and for the eight-class problems, increasing the learning rate to 0.001 gradually increased the accuracy. Figure 9 indicates that a learning rate of

0.001 over 200 epochs achieved the highest scores across all classification problems. The hybrid approach, combining DL with digital signal processing techniques such as parallel CAEs and dual scalograms, achieved promising results, even on imbalanced datasets.



**Figure 9.** Impact of the learning rate on accuracy.

#### 4. Conclusions and Future Work

Our study introduced an advanced, intelligent, lung sound recognition framework for detecting respiratory diseases. We applied dual transformation using mel scalograms and continuous wavelet transform to generate detailed time-frequency scalograms. Parallel convolutional autoencoders were trained to extract essential features from CWT and mel samples. This framework integrates parallel convolutional autoencoders and an LSTM network, reducing the possibility of misclassifying significant features while extracting rich features. The features extracted from both latent spaces are concatenated into a hybrid feature pool and processed through the LSTM model, addressing multiple-class problems. We evaluated our method on the ICBHI 2017 dataset, and the experimental results showed that our proposed model achieved promising results across multiple classification problems. For eight-class problems involving healthy samples and seven distinct lung diseases (asthma, bronchiectasis, bronchiolitis, COPD, LRTI, pneumonia, and URTI), the proposed model achieved an average accuracy of 94.16%, an average sensitivity of 89.56%, an average specificity of 99.10%, and an average F1-score of 89.56%. For the four-class problems, including crackles, wheezes, no label, and both crackles and wheezes, the model achieved an average accuracy of 79.61%, an average sensitivity of 78.55%, an average specificity of 92.49%, and an average F1-score of 78.67%. The results for the three-class problems were an average accuracy of 89.45%, an average sensitivity of 88.16%, an average specificity of 94.54%, and an average F1-score of 88.41%. Finally, for the normal vs. abnormal binary-class problems, the model achieved an average accuracy of 85.61%, an average sensitivity of 83.44%, an average specificity of 83.44%, and an average F1-score of 84.21%, outperforming all other research. In future work, we will deploy the proposed framework in a clinical setting. Additionally, we plan to enhance the robustness of the framework by increasing the number of sound samples through the integration of multiple datasets.

**Author Contributions:** Conceptualization, R.K. and S.U.K.; methodology, R.K.; software, R.K.; validation, R.K., S.U.K., U.S. and I.-S.K.; formal analysis, R.K. and U.S.; investigation, R.K.; resources, R.K.; data curation, R.K.; writing—original draft preparation, R.K.; writing—review and editing, S.U.K., U.S. and I.-S.K.; visualization, R.K.; supervision, S.U.K., U.S. and I.-S.K.; project administration, I.-S.K.; funding acquisition, I.-S.K. All authors have read and agreed to the published version of this manuscript.

**Funding:** This study was supported by the Regional Innovation Strategy (RIS) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-003).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets that support the findings of this study are openly available and are as follows. Rocha, B.; Filos, D.; Mendes, L.; Vogiatzis, I.; Perantoni, E.; Kaimakamis, E.; Natsiavas, P.; Oliveira, A.; Jácome, C.; Marques, A.; 507 et al. "A respiratory sound database for the development of automated classification" (Accessed on 31 January 2024).

**Conflicts of Interest:** The authors declare they have no conflicts of interest related to the publication of this paper.

## References

- Marciniuk, D.; Schraufnagel, D.; Ferkol, T.; Fong, K.; Joos, G.; Varela, V. Forum of International Respiratory Societies. In *The Global Impact of Respiratory Disease*, 2nd ed.; European Respiratory Society: Sheffield, UK, 2017.
- Cruz, A.A. *Global Surveillance, Prevention and Control of Chronic Respiratory Diseases: A Comprehensive Approach*; World Health Organization: Geneva, Switzerland, 2007.
- Sarkar, M.; Madabhavi, I.; Niranjana, N.; Dogra, M. Auscultation of the respiratory system. *Ann. Thorac. Med.* **2015**, *10*, 158–168. [[CrossRef](#)]
- Pham, L.; Phan, H.; Palaniappan, R.; Mertins, A.; McLoughlin, I. CNN-MoE based framework for classification of respiratory anomalies and lung disease detection. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2938–2947. [[CrossRef](#)]
- Rocha, V.; Melo, C.; Marques, A. Computerized respiratory sound analysis in people with dementia: A first-step towards diagnosis and monitoring of respiratory conditions. *Physiol. Meas.* **2016**, *37*, 2079. [[CrossRef](#)]
- Pramono, R.X.A.; Imtiaz, S.A.; Rodriguez-Villegas, E. Evaluation of features for classification of wheezes and normal respiratory sounds. *PLoS ONE* **2019**, *14*, e0213659. [[CrossRef](#)]
- Sengupta, N.; Sahidullah, M.; Saha, G. Lung sound classification using cepstral-based statistical features. *Comput. Biol. Med.* **2016**, *75*, 118–129. [[CrossRef](#)]
- İçer, S.; Genç, Ş. Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digit. Signal Process.* **2014**, *28*, 18–27. [[CrossRef](#)]
- Shi, L.; Du, K.; Zhang, C.; Ma, H.; Yan, W. Lung sound recognition algorithm based on vggish-bigru. *IEEE Access* **2019**, *7*, 139438–139449. [[CrossRef](#)]
- Mangione, S.; Nieman, L.Z. Pulmonary auscultatory skills during training in internal medicine and family practice. *Am. J. Respir. Crit. Care Med.* **1999**, *159*, 1119–1124. [[CrossRef](#)]
- Rocha, B.; Filos, D.; Mendes, L.; Vogiatzis, I.; Perantoni, E.; Kaimakamis, E.; Natsiavas, P.; Oliveira, A.; Jácome, C.; Marques, A.; et al. A respiratory sound database for the development of automated classification. In *Proceedings of the Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017*, Thessaloniki, Greece, 18–21 November 2017; Springer: Berlin/Heidelberg, Germany, 2018; pp. 33–37.
- Saeed, U.; Shah, S.Y.; Alotaibi, A.A.; Althobaiti, T.; Ramzan, N.; Abbasi, Q.H.; Shah, S.A. Portable UWB RADAR sensing system for transforming subtle chest movement into actionable micro-doppler signatures to extract respiratory rate exploiting ResNet algorithm. *IEEE Sens. J.* **2021**, *21*, 23518–23526. [[CrossRef](#)]
- Saeed, U.; Shah, S.Y.; Ahmad, J.; Imran, M.A.; Abbasi, Q.H.; Shah, S.A. Machine learning empowered COVID-19 patient monitoring using non-contact sensing: An extensive review. *J. Pharm. Anal.* **2022**, *12*, 193–204. [[CrossRef](#)]
- Saeed, U.; Shah, S.Y.; Zahid, A.; Ahmad, J.; Imran, M.A.; Abbasi, Q.H.; Shah, S.A. Wireless channel modelling for identifying six types of respiratory patterns with sdr sensing and deep multilayer perceptron. *IEEE Sens. J.* **2021**, *21*, 20833–20840. [[CrossRef](#)]
- Acharya, J.; Basu, A. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE Trans. Biomed. Circuits Syst.* **2020**, *14*, 535–544. [[CrossRef](#)]
- Acharya, J.; Basu, A.; Ser, W. Feature extraction techniques for low-power ambulatory wheeze detection wearables. In *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju Island, Republic of Korea, 11–15 July 2017; pp. 4574–4577.
- Bahoura, M. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Comput. Biol. Med.* **2009**, *39*, 824–843. [[CrossRef](#)]
- Lin, B.S.; Lin, B.S. Automatic wheezing detection using speech recognition technique. *J. Med. Biol. Eng.* **2016**, *36*, 545–554. [[CrossRef](#)]
- Gairola, S.; Tom, F.; Kwatra, N.; Jain, M. Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting. In *Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Guadalajara, Mexico, 1–5 November 2021; pp. 527–530.
- Bardou, D.; Zhang, K.; Ahmad, S.M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **2018**, *88*, 58–69. [[CrossRef](#)]
- Jayalakshmy, S.; Sudha, G.F. Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks. *Artif. Intell. Med.* **2020**, *103*, 101809. [[CrossRef](#)]
- Tariq, Z.; Shah, S.K.; Lee, Y. Feature-based fusion using CNN for lung and heart sound classification. *Sensors* **2022**, *22*, 1521. [[CrossRef](#)]

23. Dubey, R.; M Bodade, R. A review of classification techniques based on neural networks for pulmonary obstructive diseases. In Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA); 2019. Available online: [https://www.academia.edu/79686469/A\\_Review\\_of\\_Classification\\_Techniques\\_Based\\_on\\_Neural\\_Networks\\_for\\_Pulmonary\\_Obstructive\\_Diseases](https://www.academia.edu/79686469/A_Review_of_Classification_Techniques_Based_on_Neural_Networks_for_Pulmonary_Obstructive_Diseases) (accessed on 18 May 2024).
24. Kaplun, D.; Voznesensky, A.; Romanov, S.; Andreev, V.; Butusov, D. Classification of hydroacoustic signals based on harmonic wavelets and a deep learning artificial intelligence system. *Appl. Sci.* **2020**, *10*, 3097. [[CrossRef](#)]
25. Yi, W.; Park, K. Derivation of respiration from ECG measured without subject's awareness using wavelet transform. In Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society [Engineering in Medicine and Biology, Houston, TX, USA, 23–26 October 2002; Volume 1, pp. 130–131.
26. O'Brien, C.; Heneghan, C. A comparison of algorithms for estimation of a respiratory signal from the surface electrocardiogram. *Comput. Biol. Med.* **2007**, *37*, 305–314. [[CrossRef](#)]
27. Campolo, M.; Labate, D.; La Foresta, F.; Morabito, F.; Lay-Ekuakille, A.; Vergallo, P. ECG-derived respiratory signal using empirical mode decomposition. In Proceedings of the 2011 IEEE International Symposium on Medical Measurements and Applications, Bari, Italy, 30–31 May 2011; pp. 399–403.
28. Khan, S.U.; Jan, S.U.; Koo, I. Robust Epileptic Seizure Detection Using Long Short-Term Memory and Feature Fusion of Compressed Time–Frequency EEG Images. *Sensors* **2023**, *23*, 9572. [[CrossRef](#)]
29. Zaman, W.; Ahmad, Z.; Kim, J.M. Fault Diagnosis in Centrifugal Pumps: A Dual-Scalogram Approach with Convolution Autoencoder and Artificial Neural Network. *Sensors* **2024**, *24*, 851. [[CrossRef](#)]
30. García-Ordás, M.T.; Benítez-Andrades, J.A.; García-Rodríguez, I.; Benavides, C.; Alaiz-Moretón, H. Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. *Sensors* **2020**, *20*, 1214. [[CrossRef](#)]
31. Nguyen, T.; Pernkopf, F. Lung sound classification using snapshot ensemble of convolutional neural networks. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 760–763.
32. Demir, F.; Ismael, A.M.; Sengur, A. Classification of lung sounds with CNN model using parallel pooling structure. *IEEE Access* **2020**, *8*, 105376–105383. [[CrossRef](#)]
33. Petmezas, G.; Cheimariotis, G.A.; Stefanopoulos, L.; Rocha, B.; Paiva, R.P.; Katsaggelos, A.K.; Maglaveras, N. Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. *Sensors* **2022**, *22*, 1232. [[CrossRef](#)]
34. Chen, H.; Yuan, X.; Pei, Z.; Li, M.; Li, J. Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access* **2019**, *7*, 32845–32852. [[CrossRef](#)]
35. Zhang, K.; Wang, X.; Han, F.; Zhao, H. The detection of crackles based on mathematical morphology in spectrogram analysis. *Technol. Health Care* **2015**, *23*, S489–S494. [[CrossRef](#)]
36. Pramono, R.X.A.; Bowyer, S.; Rodriguez-Villegas, E. Automatic adventitious respiratory sound analysis: A systematic review. *PLoS ONE* **2017**, *12*, e0177926. [[CrossRef](#)]
37. Wei, S.; Xu, K.; Wang, D.; Liao, F.; Wang, H.; Kong, Q. Sample mixed-based data augmentation for domestic audio tagging. *arXiv* **2018**, arXiv:1808.03883.
38. Salamon, J.; Bello, J.P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [[CrossRef](#)]
39. Srivastava, A.; Jain, S.; Miranda, R.; Patil, S.; Pandya, S.; Kotecha, K. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. *PeerJ Comput. Sci.* **2021**, *7*, e369. [[CrossRef](#)]
40. Theis, L.; Shi, W.; Cunningham, A.; Huszár, F. Lossy image compression with compressive autoencoders. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
41. Ballé, J.; Laparra, V.; Simoncelli, E.P. End-to-end optimized image compression. *arXiv* **2016**, arXiv:1611.01704.
42. Chen, M.; Shi, X.; Zhang, Y.; Wu, D.; Guizani, M. Deep feature learning for medical image analysis with convolutional autoencoder neural network. *IEEE Trans. Big Data* **2017**, *7*, 750–758. [[CrossRef](#)]
43. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [[CrossRef](#)]
44. Xu, L.; Cheng, J.; Liu, J.; Kuang, H.; Wu, F.; Wang, J. Arsc-net: Adventitious respiratory sound classification network using parallel paths with channel-spatial attention. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 1125–1130.
45. Demir, F.; Sengur, A.; Bajaj, V. Convolutional neural networks based efficient approach for classification of lung diseases. *Health Inf. Sci. Syst.* **2019**, *8*, 4. [[CrossRef](#)]
46. Ma, Y.; Xu, X.; Yu, Q.; Zhang, Y.; Li, Y.; Zhao, J.; Wang, G. Lungbrn: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm. In Proceedings of the 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 17–19 October 2019; pp. 1–4.
47. Chambres, G.; Hanna, P.; Desainte-Catherine, M. Automatic detection of patient with respiratory diseases using lung sound analysis. In Proceedings of the 2018 International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, France, 4–6 September 2018; pp. 1–6.



48. Rocha, B.M.; Pessoa, D.; Marques, A.; Carvalho, P.; Paiva, R.P. Automatic classification of adventitious respiratory sounds: A (un)solved problem? *Sensors* **2020**, *21*, 57. [[CrossRef](#)]
49. Mang, L.D.; Cañadas-Quesada, F.J.; Carabias-Orti, J.J.; Combarro, E.F.; Ranilla, J. Cochleogram-based adventitious sounds classification using convolutional neural networks. *Biomed. Signal Process. Control* **2023**, *82*, 104555. [[CrossRef](#)]
50. Wanasinghe, T.; Bandara, S.; Madusanka, S.; Meedeniya, D.; Bandara, M.; de la Torre Díez, I. Lung Sound Classification with Multi-Feature Integration Utilizing Lightweight CNN Model. *IEEE Access* **2024**, *12*, 21262–21276. [[CrossRef](#)]
51. Choi, Y.; Lee, H. Interpretation of lung disease classification with light attention connected module. *Biomed. Signal Process. Control* **2023**, *84*, 104695. [[CrossRef](#)]
52. Li, J.; Yuan, J.; Wang, H.; Liu, S.; Guo, Q.; Ma, Y.; Li, Y.; Zhao, L.; Wang, G. LungAttn: Advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram. *Physiol. Meas.* **2021**, *42*, 105006. [[CrossRef](#)]
53. Zhang, Q.; Zhang, J.; Yuan, J.; Huang, H.; Zhang, Y.; Zhang, B.; Lv, G.; Lin, S.; Wang, N.; Liu, X.; et al. Sprsound: Open-source sjtu paediatric respiratory sound database. *IEEE Trans. Biomed. Circuits Syst.* **2022**, *16*, 867–881. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.