



OPEN Neural network based AI model for lung health assessment

Umaisa Hassan¹, Amit Singhal¹ & Gunjan Gupta²✉

Treating pulmonary diseases is pivotal in healthcare since they are the third leading cause of mortality globally. To aid medical experts in diagnosis, various studies have been conducted using artificial intelligence (AI) compatible devices to analyze lung sounds recorded with a stethoscope. In this paper, four datasets have been considered as a combination of two public datasets to assess the performance of the proposed approach. The signals from each dataset undergo a series of pre-processing steps, encompassing normalization, re-sampling, and framing. Thereafter, eight sub-band filters have been taken into account to segregate distinct frequency bands. The sub-band signals are represented using characteristics such as entropy, L^p norm, kurtosis, mean absolute deviation, and standard deviation. This characteristic representation for the signals is then fed to the proposed neural network (NN) for training and classification. The NN architecture consists of three fully connected layers and an output layer for classification. Our proposed approach attains 100% accuracy, specificity, and sensitivity, performing consistently well across all four datasets, which highlights the model's strong generalizability. The proposed architecture is simple, easy to realize, and has a short training time. The classification outcomes obtained through the proposed NN architecture demonstrate its superiority when compared to the existing methods.

Keywords Artificial neural network, Lung sound, Neural network, Pulmonary diseases

Lung auscultation is a commonly employed diagnostic method in pulmonary medicine¹, whereby healthcare practitioners listen to lung sounds at various locations on the anterior and posterior chest walls to evaluate respiratory function^{2,3}. These sounds are indicative of underlying anatomical abnormalities and offer crucial insights for the identification and prognosis of respiratory conditions⁴. A study carried out by the World Health Organization (WHO) estimates that respiratory diseases are responsible for approximately 10 million annual deaths worldwide⁵. Chronic obstructive pulmonary disease (COPD) ranks as the third most prevalent cause of mortality across the globe⁶. As such, lung auscultation represents a non-invasive, low-cost, and widely available diagnostic tool that can assist in identifying abnormal lung sounds and aid in the early detection and management of respiratory disorders⁷.

The timely detection of respiratory diseases is crucial in effectively managing them and limiting their spread. One method used by experts during lung auscultation involves identifying abnormal respiratory sounds such as wheezing, crackling, or stridor, which are indicative of respiratory disorders in the individual. Automated algorithms based on artificial intelligence (AI) could be highly advantageous in identifying respiratory diseases early by screening a larger population than manual screening². In recent years, there has been considerable interest in investigating the automated analysis of lung sounds. Numerous studies have focused on utilizing machine learning and deep learning methods for the classification of respiratory sounds. However, instead of directly predicting respiratory diseases from lung sound recordings, most of these studies have focused on predicting respiratory anomalies, such as identifying the presence of wheezing or crackles in the lung sounds^{8–10}.

Researchers have put forward different approaches to examine respiratory sounds, including the application of techniques such as the mel-frequency cepstrum coefficient (MFCC) in combination with a hidden Markov model¹¹ and short-time Fourier transform (STFT), as well as the use of wavelet transform and support vector machine (SVM)¹². In¹³, the authors proposed a multi-channel model for processing and classifying lung sound abnormalities. This model leveraged the attributes of both Mel spectrogram and empirical mode decomposition (EMD). Harman et al.¹⁴ used artificial neural network (ANN) and k-nearest neighbor (kNN) to classify lung sound signals into three disease classes. Dar et al.¹⁵ introduced an approach called fractional water cycle swarm optimizer-based deep residual network (Fr-WCSO-based DRN) for detecting pulmonary abnormalities using respiratory sound signals. Authors in¹⁶ introduced a lung sound recognition framework utilizing a multi-resolution interleaved network and enhanced time-frequency features. This framework included a diverse dual-branch time-frequency feature extractor (TFE), a feature enhancement module leveraging branch

¹Netaji Subhas University of Technology, Dwarka, Delhi, India. ²Cape Peninsula University of Technology, Cape Town, South Africa. ✉email: guptag@cput.ac.za

attention (FEBA), and a fusion semantic classifier. Authors in¹⁷ investigated the effectiveness of Visibility Graphs representation of LS combined with a deep residual network and named it VGAResNet for COPD classification. Time-frequency representation based neural network for COPD detection is proposed in¹⁸. Further, the authors in¹⁹ investigated various machine learning algorithms applied to features extracted using the Hilbert-Huang transform on multi-channel lung sound signals for COPD detection. To determine the severity of COPD, cuboid and octant-based quantization techniques are employed in²⁰ to identify distinctive abnormalities within the chaos plot and extreme machine learning classifier is employed for classification purpose. In¹⁹, deep belief method is employed to classify lung sound database into different lung ailments. Roy et al.²¹ introduced AsthmaSCENet, a lightweight supervised contrastive embedding learning framework designed to classify asthmatic lung sounds by ensuring a sufficient classification margin between the embeddings of healthy and asthmatic lung sounds. Authors in²² used an inception network to categorize the mel spectrogram images extracted from the lung sound signals into different lung disorders. However, in²³, researchers analyzed the self operational neural network that is driven by triple time frequency feature set to classify lung sound signals into different lung disorders.

Recently, the use of convolutional neural networks (CNNs) for sound classification has gained attention as a way to improve respiratory sound classification^{24–28}. This method involves converting one-dimensional (1D) respiratory sound signals into image data using time-frequency analysis. These data are subsequently used as input for a CNN, which automatically extracts deep features from the respiratory sounds. Finally, a classifier is utilized to automatically classify normal and abnormal respiratory sounds. Prior research has explored the utilization of two forms of respiratory sound conversion images, namely spectrogram, and scalogram, as simultaneous inputs to a CNN^{29–31}. Researchers in³² proposed a CNN network that utilizes feature extraction from MFCCs, Mel spectrograms and chromagrams. However, the computation of scalogram and spectrogram representations can be demanding in terms of computing resources, particularly for large datasets or signals with high dimensions, which can cause longer processing times and increased resource utilization. Authors in³³ extracted deep features through CNNs and subsequently performed classification using SVM. Additionally, an improved CNN combined with linear discriminant analysis (LDA)-random subspace ensembles (RSE) has been proposed as a classification method in³⁴. These studies have shown that utilizing a CNN for classification purposes achieves higher accuracy compared to previous methods that rely on machine learning with manually extracted features from a 1D signal.

Incorporating the time-series characteristics of respiratory sound data solely using a CNN poses a challenge in the classification method. Therefore, there is a need for further improvement in the deep learning model to enhance its performance. Lately, a subsequent amount of research studies have focused on utilizing various machine learning and deep learning techniques to classify respiratory diseases as either normal or pathological (binary)^{35–38}, normal, chronic, or non-chronic (ternary)³⁹, or with multiple disease classes^{39–41}. These studies examined various diseases, achieving good accuracies for binary, ternary, and multi-class classification, respectively.

Most of the existing works consider a single dataset only, for validation of their results. Further, all disease classes have not been considered. Some of the deep learning techniques have performed well but they have a high computational complexity. Moreover, these methods do not guarantee accurate results on all the datasets. Many studies evaluate their models only on a single dataset, limiting their real-world applicability. Without testing on external datasets, it is unclear whether the model is learning generalizable patterns or just adapting to dataset-specific characteristics. This research paper introduces a highly efficient methodology for the identification of lung diseases. We utilized two popular datasets to create four distinct sets of data, which are then used to evaluate the performance of our system. To classify lung diseases, we have considered a feed-forward neural network architecture. We have considered all classes of the datasets and measured various performance metrics, including accuracy, sensitivity, specificity, area under curve (AUC) and ROC. We found that our system performed well on all four datasets, achieving superior performance values as compared to the current existing methods. In addition to this we have used a different dataset to evaluate the cross-dataset performance of the proposed methodology.

The input data is subjected to some pre-processing steps such as re-sampling, framing, normalization and representing the signals in some of their key features, that are evaluated from the frames using fast Fourier transform (FFT) based zero-phase filtering method. These pre-processed signals are then passed to the proposed neural network for training and testing. The innovation of this work lies in the combination and strategic application of these methods to achieve an optimal balance between accuracy, interpretability, and computational efficiency. Compared to existing works that often rely on complex architectures, our method achieves the better performance while remaining lightweight, making it more practical for real-world deployment. The cutoff frequencies for the filters is chosen strategically to achieve the best results and the detailed analysis of parameters used for ANN is provided. The key contributions of this study can be summarized as:

- The pulmonary sound signals undergo a decomposition process into sub-band components achieved through the application of Fourier-based zero-phase filters.
- The classification is performed using a lightweight 4-layer neural network.
- Cross-dataset performance is also evaluated by considering a new dataset for testing.
- The proposed system is highly accurate and efficient, that could potentially lead to an enhanced clinical decision-making.
- Unlike some deep learning-based methods that require extensive hyperparameter tuning, our model generalizes well across different lung disease types without requiring excessive manual intervention.

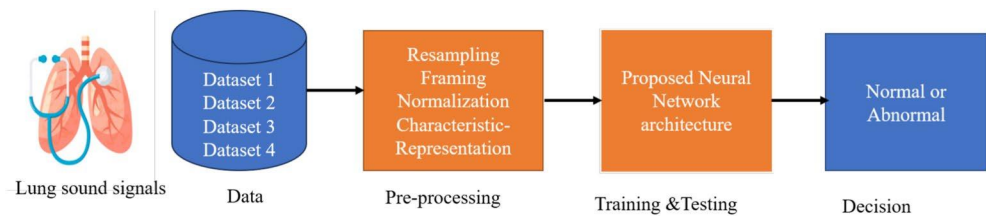


Fig. 1. Block diagram of the proposed methodology.

Type of Disease	Number of signals		
	Dataset1	Dataset2	Dataset3
Normal	35	105	140
Asthma	1	98	99
COPD	793	33	826
Pneumonia	37	15	52
BRON	29	9	38
Heart Failure	–	56	56
Lung Fibrosis	–	15	15
Pleural Effusion	–	6	6
URTI	23	–	23
LRTI	2	–	2

Table 1. Description of datasets used in this study.

The remaining manuscript is structured as follows: Sect. 2 contains details regarding the proposed methodology. Section 3 provides the results on the four datasets, along with the Sect. 4 containing relevant discussions. Section 5 presents the conclusions and future scope for this work.

Methodology

The methodology proposed in this study is illustrated in Fig. 1. The steps followed can be broadly divided into data acquisition, pre-processing, training and testing, and final decision. These steps are discussed in detail in the following sections.

Dataset

In this work we have employed two popularly available datasets, one is famously known as ICBHI 2017 challenge dataset and the other one is KAUH lung sound dataset. Two more datasets are prepared as combinations of these two public datasets. The number of signals used in first three datasets is given in Table 1, which describes the number of signals present in each class.

Dataset 1 The ICBHI 2017 challenge dataset is an openly available dataset of lung sound signals that has been created for the purpose of research and evaluation⁴. The dataset comprises 920 lung sound recordings obtained from 126 individuals with lung diseases, including pneumonia and asthma. The recordings were collected utilizing digital stethoscopes and saved in the WAV format. They have a bit depth of 16 bits and a sampling rate of 4,000 Hz. Each recording can last up to 30 seconds and is labeled with the corresponding diagnosis of the patient. The dataset is divided into two subsets: a training set that contains 689 recordings, and a test set that has 231 recordings. Along with the audio recordings, the dataset contains patient information such as gender, age, and medical history, as well as annotations of lung sound events such as wheezes, crackles, and normal sounds.

Dataset 2 It includes 337 recordings of lung sounds obtained from 112 subjects, consisting of 35 healthy individuals and 77 people with pulmonary diseases⁴². The recordings were collected in a silent environment at King Abdullah University Hospital (KAUH). The lung sounds were acquired using an electronic stethoscope and stored in the WAV file format. The dataset also provides demographic information about the subjects such as their age and gender. Moreover, each recording is annotated with labels indicating the presence of type of disease. Each recording lasts from 5 to 30 seconds. The primary goal of this dataset is to facilitate research and development of automatic lung sound analysis techniques for detecting and diagnosing pulmonary diseases.

Dataset 3 Dataset 1 and Dataset 2 are combined to create the third dataset. It encompasses a total of 1,257 signals, covering conditions such as normal, COPD, pneumonia, BRON, heart failure, URTI, LRTI, lung fibrosis, and pleural effusion.

Dataset 4 This dataset is created by dividing signals from dataset 1 and dataset 2 into 3 classes, i.e., normal, chronic and non-chronic. This dataset includes 140 signals classified as normal, 826 chronic signals, and 291 non-chronic signals.

Dataset 5 The dataset presented in⁴³ comprises 12-channel lung sound recordings from each participant and includes five different severity classes of COPD. In this study, it is utilized as a test dataset to assess cross-dataset performance.

Pre-processing

For pre-processing the signals, we have performed resampling, framing, and normalization. It is required that all signals have a similar sampling rate to extract distinguishable features. Therefore, all the signals are adjusted/re-sampled at a sampling frequency of 4000 Hz. The re-sampled signals then undergo segmentation to make smaller frames having a duration of 3 seconds. It is done to reduce the computational complexity of the analysis steps. Processing large datasets of lung sound signals can be challenging and computationally demanding due to the vast amount of information they contain. Breaking the signals into smaller frames enables more efficient handling and processing of the data. Additionally, breaking the signals into frames helps in accommodating variations in the duration of the different components present in the lung sound signals. These frames are then normalized. Normalizing is a pre-processing technique that aids in mitigating the influence of variations in magnitude and variance across various features or variables. When the input features have different scales or ranges, features with greater magnitudes or variances may exert a more significant impact on the model than features with lower magnitudes or variances. This may result in bias in the model, resulting in sub-optimal performance. Normalization also helps to avoid numerical instability. In this work, we have used min-max normalization, which is a technique used to adjust the values of a signal to a specified range, often ranging from 0 to 1. Mathematically it can be expressed as

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x is the original signal, x_{min} is the minimum value and x_{max} denotes the maximum value. This normalization process ensures that the signal's values are re-scaled to fit within the desired range while preserving their relative proportions. Subsequently, the normalized signal is passed on for further processing and feature extraction.

Signal representation

The normalized frames are analyzed using a bank of zero-phase filters. To enhance signal quality and remove noise or unwanted components from a signal, zero-phase filters are commonly used⁴⁴. These filters do not introduce any delay in the signal, unlike traditional filters that cause phase distortion. They preserve the phase relationship between different frequency components of a signal, which is especially useful for audio signal processing where accurate signal analysis is dependent on phase information. Zero-phase filters are often applied as a pre-processing step in machine learning to improve signal consistency and quality before feeding them to a neural network for classification or training. The frames are partitioned into distinct frequency bands by applying zero-phase filters denoted by

$$\phi_j(m) = \begin{cases} 1, & \text{for } (M_{j-1} + 1) \leq m \leq M_j \\ & \& (N - M_j) \leq m \leq (N - M_{j-1} - 1), \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where $j \in 1, 2, 3, \dots, L + 1$. These filters are identified by cutoffs M_j , where $M_0 = 0$, $M_1 = 0.5N/f_s$, and $M_{L+1} = N/2$, with N and f_s representing the length of the frame and the sampling frequency, respectively. The filtered components are obtained as

$$y_j[n] = \frac{1}{N} \sum_{m=0}^{N-1} F(m) \phi_{j+1}(m) \exp(j2\pi mn/N) \quad (3)$$

for $n = 0, 1, 2, \dots, N - 1$,

with $j \in 1, 2, 3, \dots, L$, where $F(m)$ represents the discrete Fourier transform (DFT) of frame $f[n]$, i.e.,

$$F(m) = \sum_{n=0}^{N-1} f[n] \exp(-j2\pi mn/N) \quad (4)$$

for $m = 0, 1, 2, \dots, N - 1$.

The first filter $\phi_1(m)$ is deployed to remove baseline wander noise⁴⁵ that may occur due to slight patient movements during signal recording. After baseline wander removal, L components are extracted according to equation (3). This filtering approach guarantees that the extracted elements $y_j[n]$ are free from any undesired time delays. The cutoff frequencies for the filters can be chosen using various techniques, like uniform, dyadic, or user-defined. In this study, uniform cutoff frequencies are chosen, with values of 125, 250, 500, 750, 1000, 1250, 1500, and 1750 Hz. To eliminate heart sound noise, which falls within the 0.5 to 150 Hz range, the first zero-phase filter is set with a cutoff frequency of 125 Hz. This ensures the removal of heart sound noise while preserving the lower-frequency components of lung sounds, which span from 100 Hz to 2000 Hz. These values have been chosen since they give equal weightage to all frequencies, resulting in an optimal classification performance.

Each sub-band is represented in terms of its key characteristics such as L^p norm ($p=0.25,0.5$), kurtosis ($Kurt$), mean absolute deviation (MAD), entropy (Ent) and standard deviation (SD). These characteristics are defined as

$$Kurt_j = \sum_{n=0}^{N-1} \left(\frac{y_j[n] - \mu_j}{\sigma_j} \right)^4, \quad (5)$$

$$L_j^p = \left(\sum_{n=0}^{N-1} |y_j[n]|^p \right)^{\frac{1}{p}}, \quad (6)$$

$$MAD_j = \frac{1}{N} \sum_{n=0}^{N-1} |y_j[n] - \mu_j|, \quad (7)$$

$$Ent_j = - \sum_i p_{ij} \log_2(p_{ij}), \quad (8)$$

$$SD_j = \sigma_j = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (y_j[n] - \mu_j)^2} \quad (9)$$

where μ_j and σ_j denote the mean and standard deviation for the signal component $y_j[n]$, respectively, and p_{ij} denote the probabilities for i different sample values of $y_j[n]$. We use a set of eight zero-phase filters to extract frequency components from lung sound signals, and 6 characteristics are computed from each frequency band. This characteristic representation captures important information about the lung sound signals, which aids in the accurate classification and diagnosis of respiratory conditions using the proposed network.

Proposed neural network model

Neural networks (NNs) have emerged as a promising machine learning algorithm for various classification problems. NNs are inspired by the human brain structure and are composed of interconnected neurons that process input data⁴⁶. NNs have seen a rapid and significant rise in their applications within the field of medical research over the past few decades⁴⁷. To apply NN for lung disease detection, the lung sound signals must undergo pre-processing and the signal should be represented in its core attributes. Then, the processed signals are used as input to train the NN on a dataset of lung sound signals labeled with corresponding disease categories. After being trained, the NN can predict the disease category of new, unlabeled lung sound signals. NNs have shown success in this field and are actively being researched and developed for improved accuracy and performance. In this work, we have proposed a simple NN architecture for the training and classification of lung diseases. The architecture of this model is implemented using the Keras application programming interface (API).

The proposed NN architecture is a stack of fully connected layers. Figure 2 depicts the NN architecture, wherein the input comprises the shape 1×40 wherein we are representing the lung sound signals in 8 sub-bands and each band in 5 basic characteristics. The first layer is a dense layer with 64 nodes/neurons. The second layer contains 32 nodes followed by another hidden layer having 16 nodes. In all these layers rectified linear unit

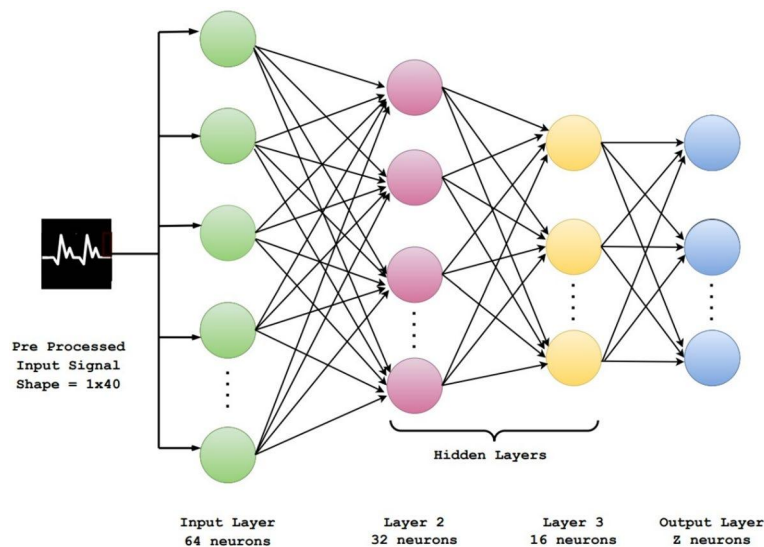


Fig. 2. Proposed NN architecture with four layers.

(ReLU) is used as the activation function. It serves as a straightforward mathematical function that establishes non-linearity to the computations within the network. The definition of the ReLU activation function is given as follows:

$$ReLU(x) = \max(0, x) \tag{10}$$

If the input value x is equal to or greater than zero, ReLU returns the input value. However, if the input value is negative, ReLU outputs zero.

Finally, there is an output layer with z nodes and a sigmoid activation function is used for binary classification to determine if a signal is normal or abnormal. This is because the problem is a binary classification, and the sigmoid outputs a value between 0 and 1, which is interpreted as the probability of the positive class. For multi-class classification, we have used the softmax activation function with z number of nodes, where z is the number of classes (or diseases) in that dataset.

To mitigate the issue of over-fitting, a dropout rate of 0.25 is employed between hidden layers. Table 2 provides the parameters used for the NN architecture, such as learning rate, batch size, optimizer, number of iterations and the dropout rate. The number of epochs used in this network is 200. For optimization, Adam optimizer algorithm is used, along with categorical cross-entropy as a loss function. The model's complexity is carefully balanced to avoid over-fitting, and hyper-parameter tuning is performed to achieve the desired results. Several parameter combinations were experimented and hyperparameter selection was finalized based on the highest validation accuracy while ensuring no overfitting, which was verified using learning curves and loss monitoring.

Evaluation metrics

The effectiveness of the proposed methodology is assessed using a stratified 80-20 train-test split, and the metrics used are mathematically defined as⁴⁸:

$$\text{Accuracy}(Acc) = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Sensitivity}(Sen) = \frac{TP}{TP + FN} \tag{12}$$

$$\text{Specificity}(Spe) = \frac{TN}{TN + FP} \tag{13}$$

where TP is the rate of correctly diagnosing the disease called as true positive value, TN is the true negative rate denoting correct diagnosis of not having disease. FP and FN denote the false detection and miss rates, respectively.

Results

The pre-processing stage of the proposed work is executed using MATLAB 2023a software. Subsequently, the training and testing of the NN model are performed using an Anaconda notebook. The total parameter size in this study is 15,815 and time required to classify one lung sound snippet is 351.5072ms. For training the time increases linearly with sample size, i.e., $O(n)$ while for testing it is constant for each testing sample, i.e., $O(1)$. We have tested our model's performance on both the binary detection of diseases as well as the multi-class classification. For binary classification, normal signals are differentiated from abnormal signals, while for multi-class classification, all diseases are considered separately. In dataset 2, the classes are normal, asthma, BRON, COPD, pneumonia, lung fibrosis, heart failure, and pleural effusion. Dataset 1 contains normal, asthma, COPD, pneumonia, BRON, URTI, and LRTI classes. We create a larger dataset, dataset 3, by merging these two datasets. For dataset 4, three classes are taken into consideration, i.e., normal, chronic and non-chronic.

In this work, we consider a characteristic representation of the signals obtained from the various frequency sub-bands. It is observed that the proposed NN produces the best results for the classification considering all four datasets. However, due to brevity of the paper, we present some results only for Dataset 3, which combines Dataset 1 and Dataset 2. Since all datasets achieve 100% accuracy, the graphs and values are highly similar, making Dataset 3 representative of the overall performance. Several parameter combinations were experimented for the proposed neural network and hyperparameter selection was finalized based on the highest validation accuracy while ensuring no overfitting, which was verified using learning curves and loss monitoring. The ablation study of neural network parameters is provided in Table 3, where the parameters used for Dataset 3 are provided.

Parameter	Value
Learning Rate	0.0001
Number of iterations	200
Batch Size	128
Optimizer	Adam
Dropout	0.25

Table 2. Parameters used for the NN network.

Parameter	Set 1	Set 2	Set 3
Number of neurons	64, 32, 16	128, 64, 32	128, 64, 64, 32, 16
Learning rate	0.0001	0.001	0.0001
Batch size	128	64	256
Optimizer	Adam	SGD	RMSprop
Number of epochs	200	100	300
Training time	6.023 seconds	5.198 seconds	9.254 seconds
Accuracy	100%	93.91%	91.76%

Table 3. Performance comparison considering different sets of parameters for dataset 3.

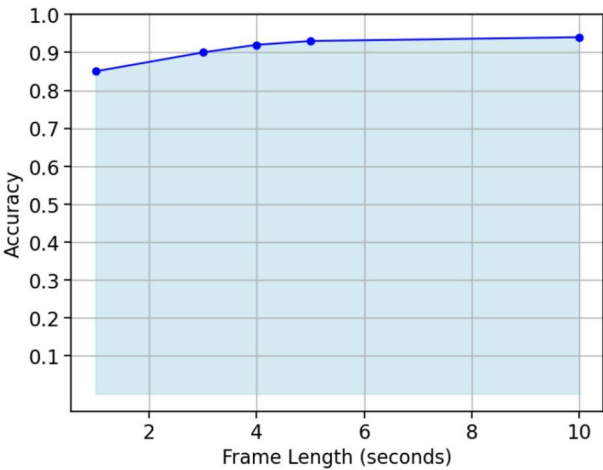


Fig. 3. Ablation study on performance with respect to input signal processing.

The selection of an appropriate frame length plays a crucial role in determining the classification performance of the model. To analyze its impact, we evaluated the model’s accuracy across different frame lengths ranging from 1 to 10 seconds. Figure 3 illustrates the relationship between frame length (x-axis) and classification accuracy (y-axis). As observed, shorter frame lengths (e.g., 1–2 seconds) result in lower accuracy, likely due to insufficient temporal information. However, as the frame length increases, accuracy improves significantly, reaching a plateau around 4–10 seconds. This suggests that a longer frame length provides richer feature representations, leading to enhanced classification performance. Based on these results, a frame length of 4 seconds or more is recommended as it ensures optimal accuracy while maintaining computational efficiency. However, extremely long frames may increase processing time without substantial accuracy gains.

The AUC-ROC metric is used to evaluate the classification performance of the proposed model, providing insight into its robustness and generalization capability. Figure 4 presents the AUC-ROC curve for Dataset 3, where the classifier achieves an AUC of 1.0. This indicates that the model exhibits excellent class discrimination, effectively distinguishing between different lung diseases. The proposed model achieved 100% accuracy, sensitivity, specificity, and F1-score, with 95% confidence intervals of (1.0, 1.0) for all metrics. These results indicate a perfect classification with no misclassifications across all lung sound classes. The confidence intervals confirm that the model’s performance is statistically certain and not influenced by randomness. In this work, t-SNE is applied to visualize the distribution of features extracted from dataset 3. This helps in understanding the clustering patterns, separability of different classes, and the effectiveness of the extracted features in distinguishing between various conditions. Figure 5 depicts the t-SNE visualization of Dataset 3. Each point represents a sample, and different colors indicate different classes.

Now, to provide a comprehensive evaluation of proposed work, we present results across all datasets. Table 4 presents the classification accuracy achieved for binary and multi-class scenarios using the proposed NN for datasets 1, 2 and 3. Further, for dataset 4 we achieved accuracy, specificity and sensitivity of 100%. The confusion matrices for dataset 1, 2, 3 and 4 are provided in Figs. 6, 7, 8 and 9, respectively. All the entries in the confusion matrices are zero, except the diagonal entries. It is clear that the proposed NN classifier produces 100% accuracy, sensitivity, and specificity, in the diagnosis of all the diseases.

In addition to 5-fold cross-validation, which is presented in Table 5, we also employed Leave-One-Subject-Out (LOSO) validation to further assess the generalization capability of the proposed model. LOSO validation is particularly effective in evaluating subject-independent performance, as it ensures that the model is tested on data from a completely unseen subject in each iteration. This approach provides a more realistic estimation of how the model would perform in practical scenarios. The results obtained through LOSO validation are summarized in Table 6, demonstrating the model’s robustness across different subjects.

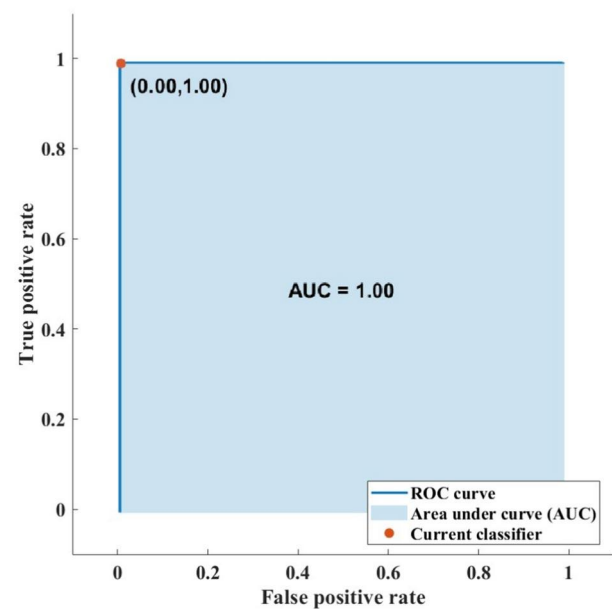


Fig. 4. AUC-ROC curve for dataset 3.

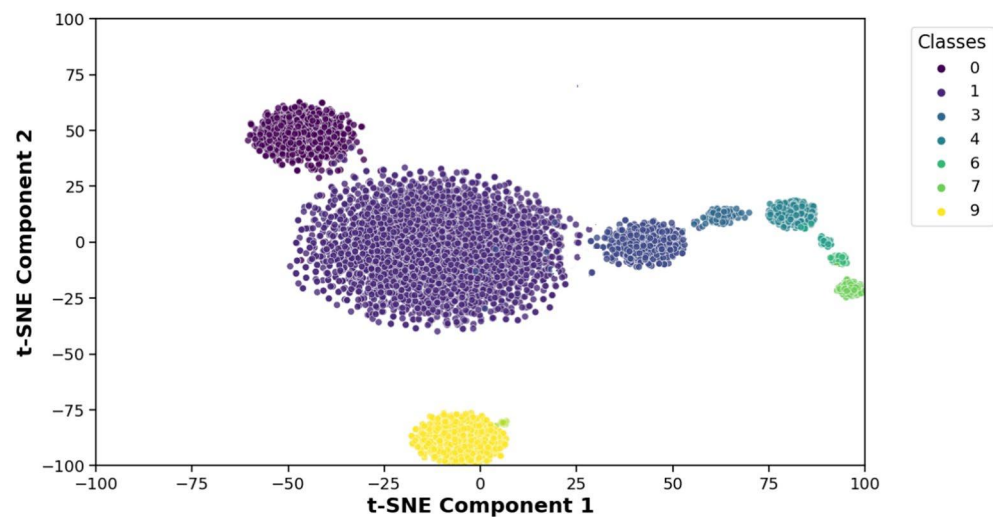


Fig. 5. T-sne plot for dataset 3.

	Dataset1	Dataset2	Dataset3
Normal vs COPD	100.0	100.0	100.0
Normal vs Asthma	100.0	100.0	100.0
Normal vs Pneumonia	100.0	100.0	100.0
Normal vs Heart attack	–	100.0	100.0
Normal vs Pleural effusion	–	100.0	100.0
Normal vs Lung Fibrosis	–	100.0	100.0
Normal vs BRON	100.0	100.0	100.0
Normal vs URTI	100.0	–	100.0
Normal vs LRTI	100.0	–	100.0
Multi-class	100.0	100.0	100.0

Table 4. Performance evaluation of the proposed algorithm in terms of accuracy(%).

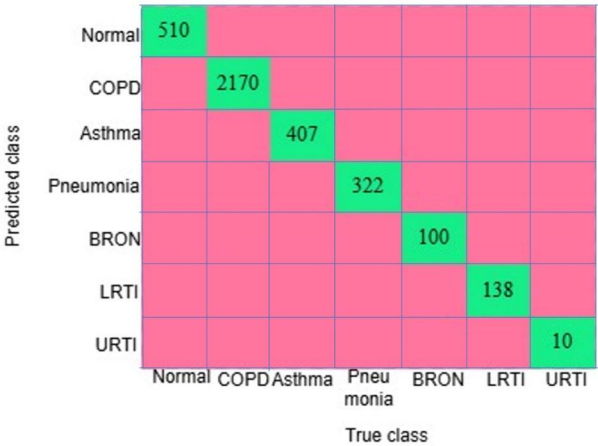


Fig. 6. Confusion matrix obtained for multi-class classification using dataset 1.

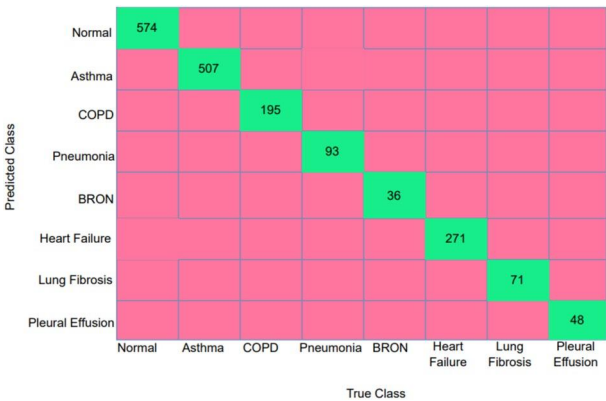


Fig. 7. Confusion matrix obtained for multi-class classification using dataset 2.

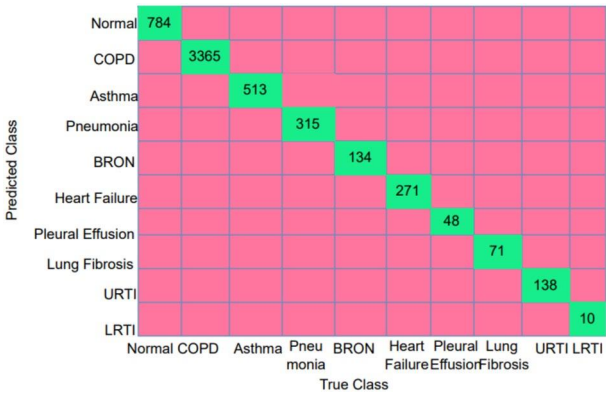


Fig. 8. Confusion matrix obtained for multi-class classification using dataset 3.

To assess the cross-dataset validation performance, Dataset 5 was designated as the evaluation dataset, while the model was trained separately on each of the other datasets. This approach helps in evaluating the model’s generalization ability across different datasets, ensuring its robustness in handling variations in data distribution. Cross-dataset validation is particularly important in real-world applications, where models must perform well on unseen data collected from different sources. The classification accuracy achieved for each training dataset is presented in Table 7, highlighting the effectiveness of the proposed model across diverse datasets.

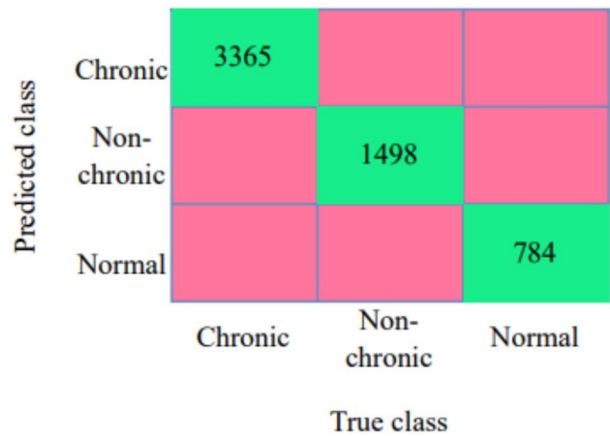


Fig. 9. Confusion matrix obtained for classification of dataset 4.

Fold	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	100	100	100
2	99.9	100	99.8
3	100	100	100
4	100	100	100
5	100	100	100

Table 5. Performance metrics for each fold.

Dataset	Accuracy (%)
Dataset 1	90.02
Dataset 2	92.07
Dataset 3	95.73
Dataset 4	93.63

Table 6. Classification accuracy for different datasets using LOSO.

Dataset	Accuracy (%)
Dataset 1	86.10
Dataset 2	87.42
Dataset 3	89.97
Dataset 4	89.09

Table 7. Cross-dataset validation results using dataset 5 as the evaluation dataset.

Discussion

A comparison of the state-of-the-art methods is provided in Table 8. The performance metrics of existing works are compared with the proposed method. Authors in³³ classified lung sounds from dataset 1, which contained diverse sample frequencies, background sounds, and noise. To convert the lung sound signals to images that could be analyzed using deep learning models, they used the time-frequency analysis and the short-time Fourier transform (STFT). Two deep learning-based methods are employed to classify the lung sounds. Initially, a pre-trained CNN model is utilized for feature extraction, while the classification task is performed using SVM. In the alternate method, the already trained CNN model is fine-tuned for lung sound classification. To determine the accuracy of the proposed methods, ten-fold cross-validation is considered.

Alqudah et al.⁴⁹ assessed three distinct deep learning models applied to both augmented and non-augmented datasets, generated from two distinct datasets resulting in four sub-datasets. The outcomes indicated that augmentation performs better than the non-augmented data. The CNN-LSTM (long short-term memory) model emerges as the best-suited model considering all datasets used since they also used the CNN and LSTM models separately. The authors used a hybrid model in combination with augmentation to achieve the best results

Study	Dataset	Methodology	Acc	Sen	Spe
Demir et al. ³³	1	STFT-CNN-SVM	65.5	–	–
Nguyen et al. ⁵²	1	Imagenet co-tuning	–	82.17	–
Basu et al. ⁴⁰	1	MFCC Features + RNN	95.67		
Tripathy et al. ⁵³	2	Fixed Boundary EWT Domain features	84.76	–	–
Fraiwan et al. ⁵¹	3	Entropy features + Boosted DT	98.27	–	–
Khan et al. ⁵⁰	4	EMD-IMF	96.29	–	–
Alqudah et al. ⁴⁹	1	CNN-LSTM	100	100	100
	2		99.81	99.97	99.97
	3		98.03	96.47	99.65
	4		99.5	98.8	99.9
Roy et al. ²³	1	Pulmo-TS2ONN	97.40	–	–
	2		99.6	–	–
Proposed method	1	NN	100	100	100
	2		100	100	100
	3		100	100	100
	4		100	100	100

Table 8. Comparison of proposed methodology with the current state-of-the-art methods.

and 100% accuracy sensitivity and specificity were attained for dataset 1 only. Authors in⁵⁰ detected chronic and non-chronic classes of lung diseases. The study initially applied empirical mode decomposition (EMD) to the lung sound signals to extract intrinsic mode functions (IMFs). To extract features, a hybrid strategy is employed, taking into account the availability of IMFs across the entire dataset. Ensemble classifiers are used for classification purposes. Fraiwan et al.⁵¹ represented the input data in terms of features like logarithmic energy entropy, Shannon entropy, and spectral entropy based on spectrograms. Discriminant classifiers and Decision trees (DT) are used to create bootstrap aggregation and adaptive boosting ensembles.

Nguyen et al.⁵² employed ResNet models, using them as the foundational architectures for detecting abnormal lung sounds and lung diseases. The pre-trained model’s learned representation is transmitted to the task by employing various techniques, including vanilla fine-tuning, co-tuning stochastic normalization, and an integration of co-tuning and stochastic normalization. To address the issue of class imbalance of the ICBHI dataset, data augmentation is performed in time as well as in time-frequency domains. Additionally, spectrum correction is introduced to address the variations in recording device properties within the dataset 1. Authors in⁵³ employed the empirical wavelet transform (EWT) with fixed boundary points to assess the modes of the lung sound signal. Time-domain features, such as Shannon entropy, as well as frequency-domain features like peak amplitude and peak frequency, are extracted from each mode. Different classifiers are utilized to automatically detect pulmonary diseases using the extracted features from the lung sound signal. The method is evaluated using light gradient boosting machine (LGBM) classifier with five-fold cross-validation. Basu et al.⁴⁰ made use of mel-frequency cepstral coefficients (MFCC) and recurrent neural network (RNN) to detect lung diseases, and attained an accuracy of 95.67% for dataset 1. It is noticeable clearly from Table 4 that the approach employed in this work outperforms the existing techniques for all the datasets. In²³ authors introduced a methodology Pulmo-TS2ONN which employs three self operational neutral scales. The authors employed three databases achieving accuracies of 97.40% for dataset1 and 99.62% for dataset 2. From the comparison, it is evident that existing studies have reported varying levels of accuracy, feature extraction techniques, and classification models used. While these methods have shown promising results, many rely on computationally complex architectures or require extensive preprocessing.

In contrast, our proposed approach achieves 100% accuracy, demonstrating exceptional classification performance while maintaining a lightweight and efficient architecture. This remarkable improvement can be attributed to the effective feature extraction, optimized network design, and robust validation strategy employed in this work.

Conclusions and future scope

In the case of respiratory sounds, the signals are random, non-linear, and highly complex due to the changing lung volume. These characteristics are observable in both individuals without any health issues and those with pathological conditions. However, these properties are more pronounced and noticeable in the lung sounds of individuals with pathological conditions. To aid medical professionals in identifying lung diseases, we introduced a NN, which is designed to be user-friendly in clinical settings. Two publicly available datasets are used independently and merged to create another dataset, while a fourth dataset is created by dividing all signals of dataset 1 and 2 into 3 classes: chronic, non-chronic and normal. The lung sound signals from the datasets are pre-processed by re-sampling to a 4 kHz sampling frequency and then segmenting them into 3-second frames. These frames are processed using FFT based zero-phase filters for decomposition into distinct frequency sub-bands. The key characteristics from these sub-bands are fed to the proposed NN architecture distinguishing between normal and abnormal signals. The proposed NN consists of 4 layers only and yields an accuracy, specificity, and sensitivity value of 100%, demonstrating superior performance compared to existing methods.

In addition to 5-fold cross-validation, we have also employed Leave-One-Subject-Out (LOSO) validation and cross-dataset validation to comprehensively evaluate our model's robustness and generalization capability. The existing methods that achieve the results comparable to our proposed work employ a hybrid complex network and have performed augmentation to achieve the desired results. The proposed NN architecture is simple, easy to realize, and has a short training time, making it practical for clinical implementations. In future, we would explore the possibility of using similar models for detecting other diseases considering different biomedical signals.

Data availability

The data is publicly available at <https://bhichallenge.med.auth.gr/> and <https://data.mendeley.com/datasets/jwy9np4gv/3>

Received: 1 December 2024; Accepted: 27 June 2025

Published online: 12 July 2025

References

- 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.* **25**(8), 2938–2947 (2021).
- Shuvo, S. B., Ali, S. N., Swapnil, S. I., Hasan, T. & Bhuiyan, M. I. H. A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram. *IEEE J. Biomed. Health Inform.* **25**(7), 2595–2603 (2020).
- ALTAN, G., Kutlu, Y., Garbi, Y., Pekmezci, A. Ö. & Nural, S. Multimedia respiratory database (respiratorydatabase@ tr): Auscultation sounds and chest x-rays. *Nat. Eng. Sci.* **2**(3), 59–72 (2017).
- 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: *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18–21 November 2017*, pp. 33–37 (2018). Springer
- Bousquet, J. & Kaltaev, N. *Global Surveillance, Prevention and Control of Chronic Respiratory Diseases: a Comprehensive Approach* (World Health Organization, Geneva, 2007).
- Elvekjaer, M. et al. Physiological abnormalities in patients admitted with acute exacerbation of copd: an observational study with continuous monitoring. *J. Clin. Monit. Comput.* **34**(5), 1051–1060 (2020).
- Rao, A., Huynh, E., Royston, T. J., Kornblith, A. & Roy, S. Acoustic methods for pulmonary diagnosis. *IEEE Rev. Biomed. Eng.* **12**, 221–239 (2018).
- Fernando, T., Sridharan, S., Denman, S., Ghaemmaghami, H. & Fookes, C. Robust and interpretable temporal convolution network for event detection in lung sound recordings. *IEEE J. Biomed. Health Inform.* **26**(7), 2898–2908 (2022).
- 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.* **14**(3), 535–544 (2020).
- Orjuela-Cañón, A.D., Gómez-Cajas, D.F., & Jiménez-Moreno, R. Artificial neural networks for acoustic lung signals classification. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico, November 2–5, 2014. Proceedings* 19, pp. 214–221 (2014). Springer
- Theodoridis, T., Solachidis, V., Vretos, N. & Daras, P. Precision Medicine Powered by pHealth and Connected Health. Springer (2018)
- Serbes, G., Ulukaya, S. & Kahya, Y.P. An automated lung sound preprocessing and classification system based on spectral analysis methods. In: *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18–21 November 2017*, pp. 45–49 (2018). Springer
- Huong, P. T. V., Thinh, L. D., Kien, P. V. & Vu, T. A. Multiple channels model based on mel spectrogram for classifying abnormalities in lung sound. *J. Biomim. Biomater. Biomed. Eng.* **63**, 63–72 (2023).
- Harman, G. Lung sounds ventilation cycle segmentation and classify healthy, asthma and copd. *Fortune J. Health Sci.* **7**, 13–24 (2024).
- Dar, J. A., Srivastava, K. K. & Mishra, A. Lung anomaly detection from respiratory sound database (sound signals). *Comput. Biol. Med.* **164**, 107311 (2023).
- Shi, L., Zhang, J., Yang, B. & Gao, Y. Lung sound recognition method based on multi-resolution interleaved net and time-frequency feature enhancement. *IEEE J. Biomed. Health Inform.* (2023)
- Roy, A., Thakur, A. & Satija, U. Vgaresnet: A unified visibility graph adjacency matrix based residual network for chronic obstructive pulmonary disease detection using lung sounds. *IEEE Sens. Lett.* (2023)
- Roy, A. & Satija, U. A novel multi-head self-organized operational neural network architecture for chronic obstructive pulmonary disease detection using lung sounds (Speech, and Language Processing, IEEE/ACM Transactions on Audio, 2024).
- Altan, G., Kutlu, Y., Pekmezci, A. Ö. & Nural, S. Deep learning with 3d-second order difference plot on respiratory sounds. *Biomed. Signal Process. Control* **45**, 58–69 (2018).
- Altan, G., Kutlu, Y. & Gökçen, A. Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds. *Turk. J. Electr. Eng. Comput. Sci.* **28**(5), 2979–2996 (2020).
- Roy, A. & Satija, U. Asthmascelnet: A lightweight supervised contrastive embedding learning framework for asthma classification using lung sounds. *Entropy* **1282**, 100 (2023).
- Roy, A. & Satija, U. Rdlinet: A novel lightweight inception network for respiratory disease classification using lung sounds. *IEEE Trans. Instrum. Meas.* (2023)
- Roy, A., Satija, U. & Karmakar, S. Pulmo-ts2onn: A novel triple scale self operational neural network for pulmonary disorder detection using respiratory sounds. *IEEE Trans. Instrum. Meas.* **73**, 1–12 (2024).
- Mushtaq, Z., Su, S.-F. & Tran, Q.-V. Spectral images based environmental sound classification using cnn with meaningful data augmentation. *Appl. Acoust.* **172**, 107581 (2021).
- Palanisamy, K., Singhanian, D. & Yao, A. Rethinking cnn models for audio classification. arXiv preprint [arXiv:2007.11154](https://arxiv.org/abs/2007.11154) (2020)
- Demir, F., Abdullah, D. A. & Sengur, A. A new deep cnn model for environmental sound classification. *IEEE Access* **8**, 66529–66537 (2020).
- Bardou, D., Zhang, K. & Ahmad, S. M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **88**, 58–69 (2018).
- Bhatta, L.N., Bhatta, S.M. & Akshay, N. Respiratory analysis–detection of various lung diseases using audio signals. In: *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, pp. 506–511 (2022). IEEE
- Rioul, O. & Vetterli, M. Wavelets and signal processing. *IEEE Signal Process. Mag.* **8**(4), 14–38 (1991).
- Minami, K., Lu, H., Kim, H., Mabu, S., Hirano, Y. & Kido, S. Automatic classification of large-scale respiratory sound dataset based on convolutional neural network. In: *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, pp. 804–807 (2019). IEEE

31. Pham Thi Viet, H., Nguyen Thi Ngoc, H., Tran Anh, V. & Hoang Quang, H. Classification of lung sounds using scalogram representation of sound segments and convolutional neural network. *J. Med. Eng. Technol.* **46**(4), 270–279 (2022).
32. Wanasinghe, T., Bandara, S., Madusanka, S., Meedeniya, D., Bandara, M. & Torre Díez, I. Lung sound classification with multi-feature integration utilizing lightweight cnn model. *IEEE Access* (2024)
33. Demir, F., Sengur, A. & Bajaj, V. Convolutional neural networks based efficient approach for classification of lung diseases. *Health Inf. Sci. Syst.* **8**, 1–8 (2020).
34. Demir, F., Ismael, A. M. & Sengur, A. Classification of lung sounds with cnn model using parallel pooling structure. *IEEE Access* **8**, 105376–105383 (2020).
35. Perna, D. Convolutional neural networks learning from respiratory data. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2109–2113 (2018). IEEE
36. Islam, M. A., Bandyopadhyaya, I., Bhattacharyya, P. & Saha, G. Multichannel lung sound analysis for asthma detection. *Comput. Methods Programs Biomed.* **159**, 111–123 (2018).
37. Hassan, U. & Singhal, A. Automated diagnosis of pulmonary diseases using lung sound signals. *IETE J. Res.*, 1–9 (2023)
38. Lal, K. N. A lung sound recognition model to diagnoses the respiratory diseases by using transfer learning. *Multimed. Tools Appl.* **82**(23), 36615–36631 (2023) (**Springer**).
39. 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* **20**(4), 1214 (2020).
40. Basu, V. & Rana, S. Respiratory diseases recognition through respiratory sound with the help of deep neural network. In: *2020 4th International Conference on Computational Intelligence and Networks (CINE)*, pp. 1–6 (2020). IEEE
41. Hassan, U., Singhal, A. & Chaudhary, P. Lung disease detection using easynet. *Biomed. Signal Process. Control* **91**, 105944 (2024).
42. Fraiwan, M., Fraiwan, L., Khassawneh, B. & Ibnian, A. A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. *Data Brief* **35**, 106913 (2021).
43. Altan, G. & Kutlu, Y. Respiratorydatabase@ tr (copd severity analysis). *Mendeley Data* **1**, 2020 (2020).
44. Mehla, V. K., Singhal, A. & Singh, P. A novel approach for automated alcoholism detection using Fourier decomposition method. *J. Neurosci. Methods* **346**, 108945 (2020).
45. Singhal, A., Singh, P., Fatimah, B. & Pachori, R. B. An efficient removal of power-line interference and baseline wander from ECG signals by employing Fourier decomposition technique. *Biomed. Signal Process Control* **57**, 101741 (2020).
46. Naderirad, I., Saadat, M., Avokh, A. & Mehrparvar, M. Estimation of the basin outflow by wavelet neural network, conjunctive use of wavelet analysis and artificial neural network. *Iran. J. Sci. Technol. Trans. Civ. Eng.*, 1–14 (2023)
47. Hasnul, M. A., Ab. Aziz, N. A. & Abd. Aziz, A. Augmenting ecg data with multiple filters for a better emotion recognition system. *Arab. J. Sci. Eng.*, 1–22 (2023)
48. Hassan, U. & Singhal, A. Convolutional neural network framework for eeg-based adhd diagnosis in children. *Health Inf. Sci. Syst.* **12**(1), 44 (2024).
49. Alqudah, A. M., Qazan, S. & Obeidat, Y. M. Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds. *Soft Comput.* **26**(24), 13405–13429 (2022).
50. Khan, S. I. & Pachori, R. B. Automated classification of lung sound signals based on empirical mode decomposition. *Expert Syst. Appl.* **184**, 115456 (2021).
51. Fraiwan, L. et al. Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. *Biocybern. Biomed. Eng.* **41**(1), 1–14 (2021).
52. Nguyen, T. & Pernkopf, F. Lung sound classification using co-tuning and stochastic normalization. *IEEE Trans. Biomed. Eng.* **69**(9), 2872–2882 (2022).
53. Tripathy, R. K., Dash, S., Rath, A., Panda, G. & Pachori, R. B. Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform. *IEEE Sens. Lett.* **6**(5), 1–4 (2022).

Author contributions

U.H and A.S wrote the main manuscript, U.H,A.S and G.G prepared results, A.S and G.G reviewed the paper.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

The Respiratory Sound database was originally compiled to support the scientific challenge organized at Int. Conf. on Biomedical Health Informatics - ICBHI 2017. The current version of this database is made freely available for research and contains both the public and the private dataset of the ICBHI challenge.

Additional information

Correspondence and requests for materials should be addressed to G.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025