IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1733∼1745
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1733-1745

ResNet based deep learning approach for chronic obstructive pulmonary disease prediction using lung sound analysis

Babitha Sudhakar Ullal[1], Veena Kalludi Narasimhaiah[1], Rithul Kamesh[2]
[1]School of Electronics and Communication Engineering, REVA University, Bengaluru, India
[2]Department of Electronics and Communication Engineering, PES University, Bengaluru, India

Article Info

Article history:

Received Aug 21, 2025
Revised Jan 15, 2026
Accepted Feb 6, 2026

Keywords:

Audio signal processing
Chronic obstructive pulmonary disease
Convolutional neural network
Long short-term memory
Residual networks

ABSTRACT

Chronic obstructive pulmonary disease (COPD) affects around 300-400 million people worldwide representing a critical healthcare challenge that requires early detection for effective intervention. This work introduces chronic lung analysis via audio signal prediction (CLASP), a novel framework achieving 97.90% accuracy in predicting COPD automatically through respiratory audio signal analysis. This method integrates advanced signal processing and deep learning architectures, comparing long short-term memory (LSTM), convolutional neural networks (CNN), and residual networks (ResNet) models for optimal performance. The ResNet architecture exhibits superior diagnostic capability with precision of 98.72%, recall of 96.86%, and 0.9937 area under the curve (AUC), as compared to existing methods by significant margins. These results establish a new benchmark for noninvasive COPD detection, thus enabling practical deployment in clinical settings thereby dramatically improving the patient outcomes by early detection and also reduce healthcare costs.

Corresponding Author:

Babitha Sudhakar Ullal
School of Electronics and Communication Engineering, REVA University
Bengaluru, India
Email: babitharoshan91@gmail.com

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is one of the critical challenges in present-day respiratory medicine, globally affecting around 384 million people, and is estimated to become the world's third most common cause of death by 2030 [1], [2]. This disease is progressive in nature [3], and is characterized by persistent problems of respiration and airflow limitation [4], [5]. It demands early detection techniques that can identify patients before irreversible lung damage occurs [6], [7]. Spirometry and clinical evaluation are the current diagnostic approaches used which often do not detect COPD in its early stages [8] where medical interventions could be the most effective to improve patient outcome, creating a demanding need for more sensitive and accessible screening methods.

Machine learning approaches in detection of various respiratory diseases has exhibited remarkable promise [9]–[13] by the application of audio signal analysis using the vast temporal and spectral information present in breath sounds [14], [15]. Recent developments in deep learning architectures, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and residual networks (ResNets), have shown exceptional capability in pattern recognition tasks of biomedical signals. These technologies provide the

necessary computational foundation to implement automated diagnostic systems that can analyse complex respiratory audio patterns with high accuracy while maintaining the noninvasive nature.

The specific challenge addressed in this work is the implementation of an automated system to distinguish COPD-related respiratory patterns from other normal or abnormal breathing patterns with acceptable levels of accuracy for clinical deployment. Traditional machine learning approaches have achieved modest success in this area, with accuracies typically varying between 82.5% and 93% [16], [17]. But these performance levels fall short of the required standards for clinical deployment.

The proposed approach introduces chronic lung analysis via audio signal prediction (CLASP), a complete framework combining the advanced signal processing techniques with cutting-edge deep learning models giving unparalleled accuracy in automated detection of COPD. The key innovations in the work are as follows. First, development of an optimized audio preprocessing pipeline incorporating Mel-frequency cepstral coefficients (MFCC) along with first derivative, delta and second derivative, delta-delta features to enable enhanced temporal pattern capture. Second, comparative evaluation of three distinct deep learning models—long short-term memory (LSTM), CNN, and ResNet. Third, implementation of threshold optimization techniques to enhance its clinical utility.

2. RELATED WORK

Lee *et al.* [16] designed a model that uses thermal imaging to capture respiratory patterns, focusing on four features considered as primary: total volume of respiration, average expiration distance, average inspiration distance, and respiration rate. Later, Z-score normalization was applied to these features and combined them through weighted summation to generate a composite score used in classification later. The accuracy of the model is 82.5%. The model's high recall indicates the ability to identify individuals with COPD more accurately minimizing false negatives.

Siddiqui *et al.* [17] explored the use of a non-invasive method - ultra-wideband (UWB) radar to differentiate COPD patients from individuals who are healthy. Data was collected from 70 subjects (35 each of COPD condition and healthy controls). The researchers extracted data of respiration and incorporated further features such as age of patient, smoking history and gender to enhance the performance accuracy. Several machine learning classifiers including support vector machine (SVM), naïve Bayes (NB), adaptive boosting (AdaBoost), k-nearest neighbor (KNN), random forest (RF) and deep learning models like CNN and LSTM networks were employed. Among these, highest accuracy of 93% was achieved by LSTM model, demonstrating the potential of combination of UWB radar and machine learning as a non-invasive and effective method for COPD detection.

Abineza *et al.* [18] utilized time-stamped electronic health records from COPD patients to develop an LSTM model to predict subsequent exacerbation by analyzing symptoms, patterns, and arterial oxygen saturation levels over time. Experiment was done with varying time windows, ranging from one to six prior days, to forecast the likelihood of an exacerbation on the following day. The model demonstrated optimal performance when using a one-day time window, achieving testing accuracy of 85%, training accuracy of 87%, and area under the curve (AUC) of 0.83. These results were obtained from a dataset comprising 54 patients, which is a small number. Reliance on saturation of peripheral oxygen (SPO2) is the only clinical variable used as factors influencing COPD exacerbations.

Mei *et al.* [19] introduces DeepSpiro, a deep learning novel architecture designed to enhance detection as well as early prediction of COPD using spirogram data. DeepSpiro comprises of four primary components: SpiroSmoother, SpiroEncoder, SpiroExplainer, and SpiroPredictor. The model resulted in 0.8328 value of AUC distinguishing individuals with COPD from those without. In early prediction tasks, DeepSpiro effectively differentiated between the groups of low-risk and high-risk observing substantial differences in future COPD development. This underscores the model's potential in forecasting the progression of COPD over long-term. The model accuracy depends on the quality of spirogram data.

Yin *et al.* [20] uses fractional-order dynamics and deep learning techniques to predict COPD. Thorax breathing effort, respiratory rate, and oxygen saturation levels were extracted to obtain fractional dynamic signatures to train a deep neural network (DNN). The model accuracy was 94.01% when trained on the WestRo Porti COPD dataset and tested on the WestRo COPD dataset, and 90.13% accuracy in the reverse scenario. However, the relatively small number of unique patients may limit the model's generalizability to broader populations.

Bairagi and Kanwade [21] used a non-invasive technique by analyzing surface electromyography (sEMG) signals from the sternomastoid muscle, a primary respiratory muscle aiming to overcome limitations of traditional spirometry. EMG signals were examined across time domain, frequency domain and time-frequency domains. An algorithm that detects onset based on slope was employed to identify muscle activation periods, enhancing the accuracy of feature extraction. Features were extracted using continuous wavelet transform (CWT) at single-frequency of 7, 8, and 10 Hz facilitating COPD classification based on grades of severity, By employing this technique, classification accuracy of 85.89% across different COPD severity grades is acheived.

Kumar *et al.* [22] presents an innovative approach in diagnosing COPD by integrating images scanned by computed tomography (CT) with audio samples of lung to enhance the diagnostic process for COPD. Features are extracted from scan images and audio samples, including texture, histogram intensity, Gaussian scale space, chroma, and MFCCs. To assess severity level of the patient by performing early classification, the extracted features are fed into the ensemble learning technique. The proposed framework achieved accuracies of 97.50% for fusion technique based early diagnosis, 98% for early diagnosis using the CT diagnostic model, 95.30% for early diagnosis utilizing the cough sample model. These high accuracies have contribution not from the audio signals alone, but also from CT images.

Ullah *et al.* [23] employed dataset from Kaggle (respiratory sound database) and chest wall lung sounds. To ensure uniformity along with duration of fixed-length, raw signals were resampled to 4 kHz and later zero-padded. Later segmentation was performed to prepare the data for feature extraction using MFCC (13 features) and short term Fourier transform (STFT) (1,000 features). SVM, artificial neural network (ANN), KNN, RF, and decision tree (DT) machine learning algorithms were employed. The models were trained for 70% data and validated on 30%. The combination STFT+MFCC-ANN yielded best accuracy.

Nunavath *et al.* [24] explores deep learning architectures to predict exacerbation in COPD patients. The authors employ LSTM to analyze patient data (only 94) and identify patterns that indicate the likelihood of future exacerbation. The deep neural network performed better than traditional machine learning approaches. in predicting COPD exacerbation. The LSTM model showed 92.86% accuracy.

Jenefa *et al.* [25] presents a novel approach to predicting COPD by integrating both CNN and LSTM. This approach leverages the strengths of both architectures: spatial features extraction using CNN and temporal dependencies in sequential data is captured using LSTM. The model efficiently captured complex COPD patterns, leading to more effective predictions. The method identified early stages of COPD with accuracy greater than 95%. However, these approaches often struggle in capturing long-term temporal patterns in the audio signals, which are critical for accurate COPD diagnosis. CLASP builds upon these works using LSTM model, a CNN model and a ResNet model to capture long-term and short-term patterns in respiratory audio signals and compare their performance.

3. METHOD

The CLASP framework uses a systematic approach to detect COPD through the analysis of respiratory audio signal. The methodology includes three primary components. First, computational signal processing techniques for audio preprocessing and feature extraction. Second, experimental deep learning architectures for pattern recognition. Third, comprehensive evaluation protocols for clinical validation. The proposed methodology utilizes a publicly available dataset available in Kaggle at https://www.kaggle.com/code/mariammagdy22/pulmonary-diseases-detection-system/input.

3.1. Computational techniques

Audio signals are processed using a sampling frequency of 22,050 Hz with windowed segmentation. MFCC extraction is done by including 13 static coefficients augmented with delta and delta-delta features for capture of temporal dynamics. The foundation for MFCC computation involves STFT analysis, Mel-scale filterbank application, and processing of discrete cosine transform (DCT). The proposed implementation uses librosa library with optimized parameters validated through preliminary testing on the International Conference on Biomedical Health Informatics (ICBHI) dataset [15]. This approach gives robust feature representations that capture spectral characteristics along with temporal variations required for respiratory pattern analysis. Spectrogram grid showcasing respiratory audio signals used in the CLASP framework is shown in Figure 1.

3.2. Experimental techniques

Three different deep learning architectures were implemented and methodically compared: LSTM networks with attention mechanisms, CNN with global pooling strategies, and ResNet with skip connections. Each architecture was designed explicitly for the 39-dimensional MFCC feature vectors, with carefully considering temporal dependencies in respiratory audio signals. Training protocols used stratified data splitting with 80-20 train-test split, ensuring balanced representation of COPD and non-COPD classes. Adam optimizer is used for model optimization with 0.001 as learning rate. To prevent overfitting, early stopping and strategies to reduce learning rate are used which also maximizes convergence stability.

3.3. Error analysis and validation

The evaluation protocols addressed both random and systematic error sources through metric assessment including accuracy, precision, specificity, F1-score, recall, and area under the curve (AUC). Threshold optimization techniques are used to enhance clinical utility, prioritizing sensitivity for screening applications in medical field. Strategies for cross-validation and analysis of confusion matrix provides detailed insights into model performance characteristics. Training time analysis quantified computational efficiency trade-offs and statistical significance testing confirmed the robustness of performance differences across architectures. CLASP pipeline architecture with three deep learning models (CNN, ResNet, and LSTM with attention) is shown in Figure 2.

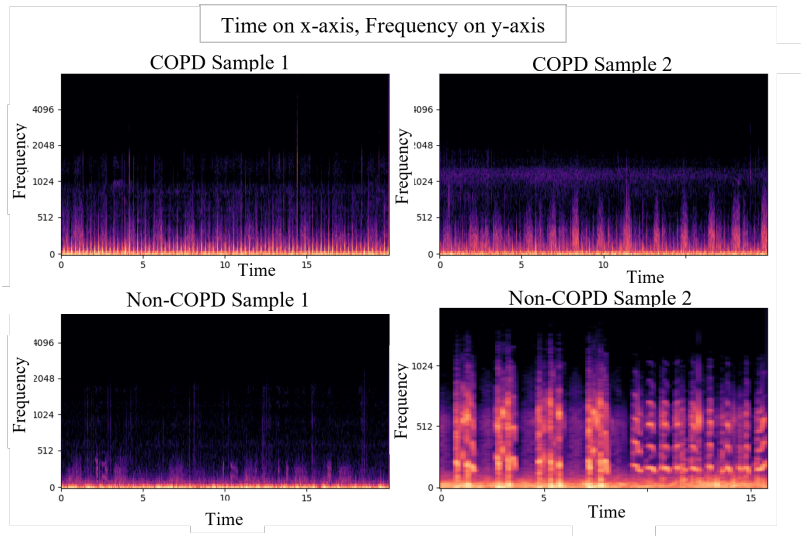

Figure 1. Spectrogram grid used in the CLASP framework

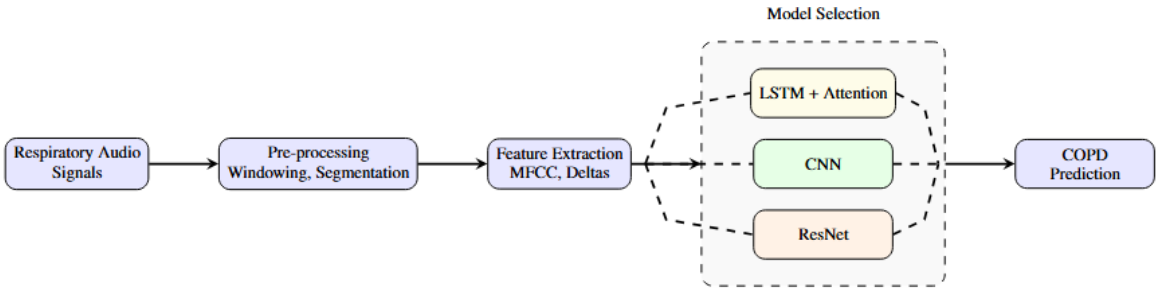

Figure 2. CLASP pipeline architecture

4. SYSTEM ARCHITECTURE AND EXPERIMENTAL SET-UP

The CLASP framework consists of three main components optimized for respiratory audio signal analysis: pre-processing, feature extraction, and neural network-based prediction. Each component is

designed to work seamlessly with the other, forming a comprehensive analysis pipeline. The audio signal pre-processing, MFCC computation, dynamic feature computation and the three models (LSTM, CNN, and ResNet architecture) are discussed in detail in the following sections.

4.1. Audio signal pre-processing

The pre-processing phase converts raw audio signals into a format that is suitable for analysis through the following steps:

i) Sampling frequency (f_s): audio signals are sampled at 22,050 Hz for high resolution.
ii) Window size (w_s): a window size of 20 ms corresponds to 441 samples.
iii) Step size (s_s): a step size of 10 ms ensures a 50% overlap between consecutive windows.
iv) Fixed number of windows: each segment has 10 consecutive windows with total of 110 ms per segment.

To segment the signals for analysis, we apply a windowing function. Mathematically, the process of segmentation is expressed as (1).

$$x_i[n] = x[n + is_s], \quad 0 \leq n < N_w \tag{1}$$

Where $x_i[n]$ represents the i-th segment, and N_w is the window size in samples. At the segment boundaries spectral leakage is minimized by using Hamming window function to taper each segment. This function is given as (2).

$$w[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N_w - 1}\right), \quad 0 \leq n < N_w \tag{2}$$

The trade-off between temporal precision and frequency resolution was balanced by suitable choice of window function and overlap which was decided by extensive experimentation. The Hamming window, in particular, was chosen for its superior frequency resolution compared to rectangular windows, while maintaining good temporal resolution.

4.2. Feature extraction

This section details MFCC computation and dynamic feature computation. STFT, Mel filterbank application, DCT, and the feature composition are involved in MFCC computation, whereas dynamic feature computation includes delta (rate of change of the cepstral coefficients) and delta-delta (rate of change of delta features).

4.2.1. Mel-frequency cepstral coefficients computation

The MFCC computation process to capture different aspects of the audio signal is as follows:

i) STFT: STFT converts each windowed signal into its frequency representation as (3).

$$X_i[k] = \sum_{n=0}^{N_w - 1} y_i[n] e^{-j2\pi kn/N} \tag{3}$$

Where $y_i[n]$ represents the windowed signal and $X_i[k]$ gives its frequency components.

ii) Mel filterbank application: now, a set of Mel-scale filters is applied to the STFT magnitudes. This step maps the frequency components to Mel-scale, which better represents human auditory perception as (4).

$$S_i[m] = \sum_{k=0}^{N/2} |X_i[k]|^2 H_m[k], \quad 0 \leq m < M \tag{4}$$

Where $H_m[k]$ represents the m-th Mel-filter response, M is the total number of Mel-filters in the Mel filterbank, $S_i[m]$ is Mel-filtered spectral energy for the m-th Mel-filter at time frame i.

iii) DCT: now, MFCC is computed using the DCT as (5).

$$c_i[n] = \sum_{m=0}^{M-1} \log(S_i[m]) \cos\left(\frac{\pi n(m + \frac{1}{2})}{M}\right) \qquad 0 \leq n < 13 \tag{5}$$

Where $c_i[n]$ is the n-th MFCC for the i-th time frame.

iv) Feature composition: the final feature vector comprises 13 static MFCCs, delta coefficients, and 13 delta-delta coefficients.

4.2.2. Dynamic feature computation

Temporal variations in the audio signal are captured by computing dynamic features (deltas) from the MFCCs, as defined in (6). These features represent the rate of change of the cepstral coefficients and indicate how the spectral characteristics of the audio change over time.

$$c\Delta f_t = \frac{\sum_{\theta=1}^{\Theta} \theta(f_{t+\theta} - f_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{6}$$

Where $\Theta = 3$ defines computation window width and θ represents our time lag, f_t is cepstral coefficient at time frame t, and Δf_t is first-order derivative of f_t. Similarly, rate of change of delta features, $\Delta^2 f_t$ is (7).

$$\Delta^2 f_t = \frac{\sum_{\theta=1}^{\Theta} \theta(\Delta f_{t+\theta} - \Delta f_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{7}$$

4.3. Model architectures

4.3.1. Long short-term memory-based architecture

The LSTM model processes sequential MFCC features using bidirectional long short-term memory (BiLSTM) layers followed by an attention mechanism to identify the most important segments in the respiratory audio signals. Each of the two BiLSTM layers are followed by batch normalization and dropout layers. Attention mechanism helps to focus on important parts of the sequence. The final layers include dense layers and sigmoid activation for binary classification. Adam optimizer with learning rate of 0.001 is used to train the model. LSTM model architecture with attention mechanism is shown in Figure 3.

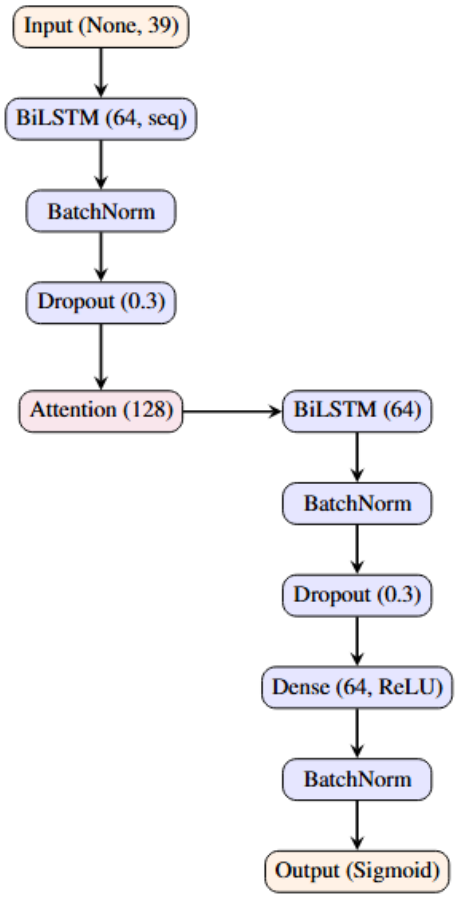

Figure 3. LSTM model architecture with attention mechanism

4.3.2. Convolutional neural network-based architecture

The CNN model reshapes sequential MFCC features into an image-like format and applies convolutional blocks to extract spatial and temporal features. This approach treats the time-frequency representation of audio as a 2D image to leverage the pattern recognition strength of CNNs. The architecture consists of three convolutional blocks, each followed by batch normalization, max pooling, and dropout layers to prevent overfitting. For binary classification, the final layer consists of global average pooling, dense layers, and a sigmoid activation. This model is trained using Adam optimizer and a learning rate of 0.001 is used. The architecture of CNN model is shown in Figure 4.

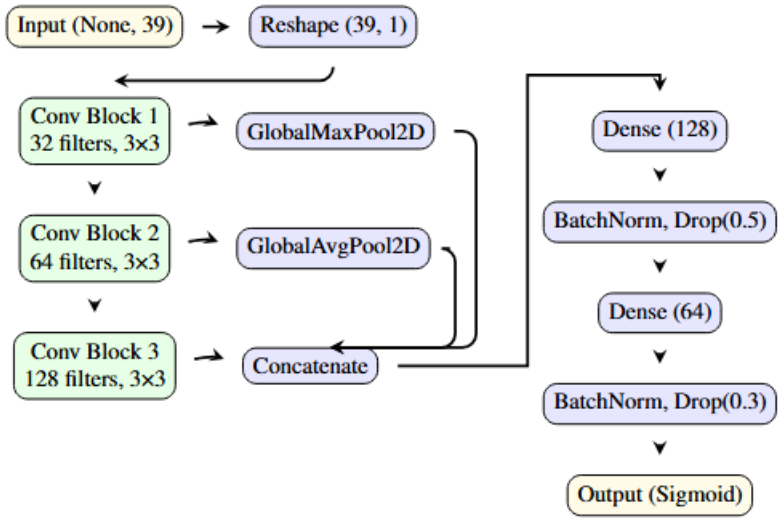

Figure 4. CNN model architecture

4.3.3. ResNet-based architecture

The ResNet model uses residual connections to enable deeper networks to learn complex feature representations while avoiding vanishing gradients, effectively capturing subtle respiratory audio patterns that distinguish COPD from other conditions. The architecture begins with convolutional layer and is followed by three stages of residual blocks. Each block includes two convolutional layers, batch normalization followed by rectified linear unit (ReLU) activation, along with a shortcut connection. The model concludes with global average pooling, dense layers, and a sigmoid activation. The skip connections in residual blocks allow the gradient to flow more easily during backpropagation, reducing the vanishing gradient problem. ResNet architecture is shown in Figure 5 and ResNet residual block structure is shown in Figure 6, where solid red arrow shows the projection pathway when dimensions change (1×1 convolution), the dashed red arrow shows the direct skip connection when dimensions match. The model training is done using Adam optimizer and a learning rate of 0.001 is used and can be trained with focal loss to address class imbalance.

4.4. Training and evaluation

4.4.1. Dataset

The dataset is taken from Kaggle (respiratory sound database), which contains 920 audio recordings from 126 subjects. It includes samples of healthy individuals and patients with various respiratory conditions, including COPD. This dataset was augmented with signal transformations, resulting in a total of 800 recordings for COPD and Non-COPD, with a 80-20 split between training and testing sets.

4.4.2. Training methodology

This section outlines the training methodology of the CLASP framework. The discussion is organized into three main components. These components are described as follows.

i) Class imbalance bandling: the CLASP framework implements three complementary strategies to address class imbalance, a common issue to address in medical datasets. Data augmentation is done using time domain augmentation, feature domain augmentation, and sample mixing. Also, balanced training sets and focal loss with false negative weighting are discussed in the following sections.

– Data augmentation: the audio processor component applies targeted augmentation to increase minority class representation: time-domain augmentation: time stretching/compression (±10%), pitch shifting (±2 semitones); feature-domain augmentation: spectral masking, random feature scaling, additive Gaussian noise; and sample mixing: linear combination of similar-class samples with random weights.
– Balanced training sets: for each training epoch, the framework dynamically creates balanced mini-batches: minority class samples are upsampled to match majority class frequency; create_balanced_subset function, which ensures equal class representation while maintaining diversity within classes; and random state initialization ensures reproducibility across training runs.
– Focal loss with false negative weighting: a specialized loss function FL(p_t) is employed to prioritize correct classification of COPD cases as in (8).

$$\text{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{8}$$

Where p_t is the probability of estimation of the model for the correct class, α_t is factor for balancing, and γ is focusing parameter. For COPD detection specifically, we further modify the loss function to assign a higher penalty to false negatives as in (9).

$$\text{FNFL}(y,\hat{y}) = \begin{cases} \text{FL}(\hat{y}) & \text{if } y = 0 \\ \text{FL}(\hat{y}) \cdot \text{fn\_weight} & \text{if } y = 1 \text{ and } \hat{y} < 0.5 \end{cases} \tag{9}$$

Where fn_weight is set to 3.0 by default, effectively tripling the loss for missed COPD cases.

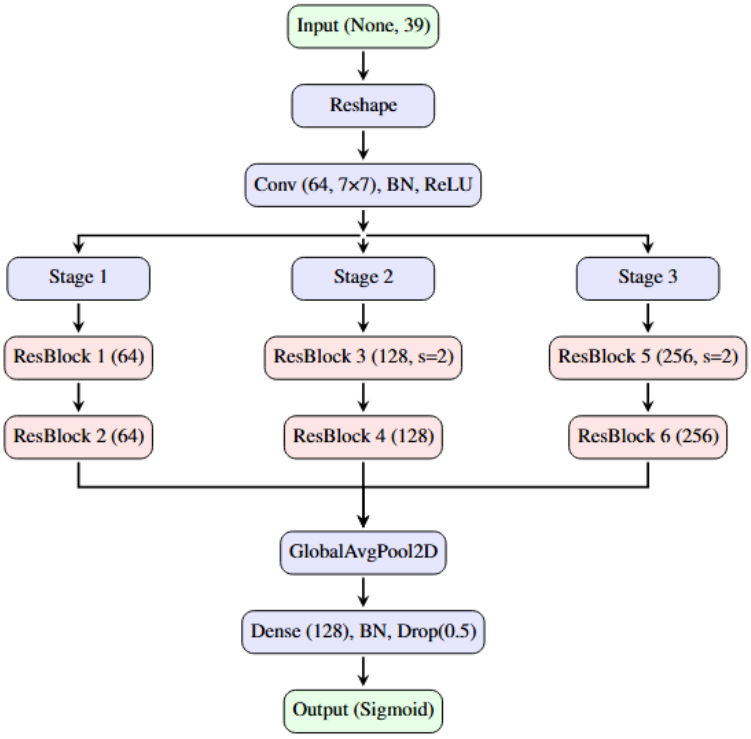

Figure 5. ResNet architecture

ii) Threshold optimization: CLASP uses a sensitivity-weighted optimization to find the optimal classification threshold instead of standard value of 0.5 allowing for fine-tuning the specificity - sensitivity trade-off. Comparison of classification thresholds for each model is shown in Figure 7. The sensitivity–specificity threshold selection method is as follows:
– Multiple candidate thresholds are evaluated on validation data.
– For each threshold, a weighted score is computed as in (10).

$$\text{Score}(\theta) = \frac{\text{sensitivity}(\theta)\cdot w + \text{specificity}(\theta)}{w+1} \tag{10}$$

Where w is the sensitivity weight (default =2.0).

– The threshold maximizing this weighted score is selected as optimal.

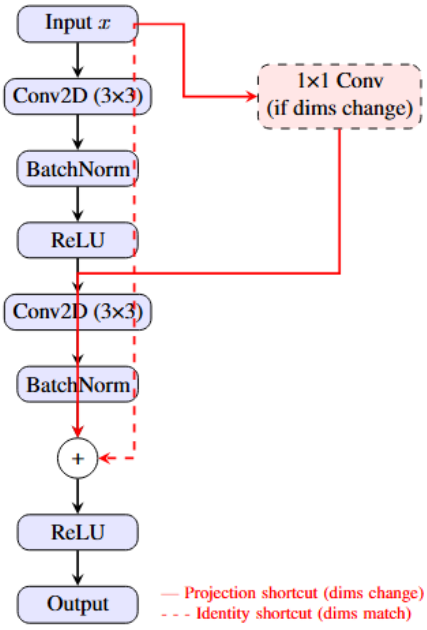

Figure 6. ResNet residual block structure

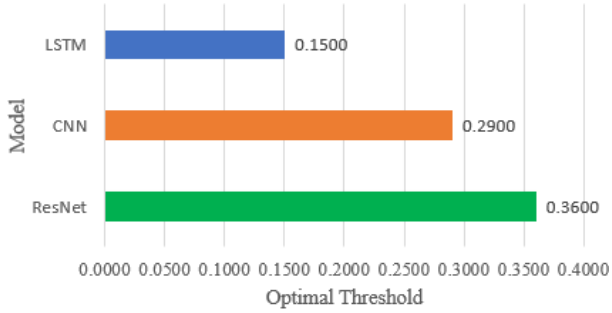

Figure 7. Comparison of classification thresholds

iii) Training configuration: the models are trained with the configurations as follows:

– Optimizer: Adam with learning rates 0.001.
– Batch normalization: stabilizes the training when applied after each major layer.
– Dropout: strategic application with increasing rates deeper in the network (20% to 50%).
– Early stopping: based on validation loss with patience of 10-15 epochs.
– Learning rate reduction: factor of 0.2 when validation loss plateaus.
– Epochs: up to 50 for LSTM and ResNet,40 for CNN (typically terminating earlier due to early stopping).
– Validation split: 15% of training data. These configuration choices were opted by conducting extensive experimentation and specifically tuned for respiratory audio classification.

5. RESULTS AND DISCUSSION

The proposed CLASP framework shows clear improvements over existing approaches and is validated through performance evaluations. A systematic comparison of three deep learning architectures reveals optimal

performance characteristics, underscoring its effectiveness. The following section discusses precision recall curve comparison, receiver operating characteristic (ROC) curve, performance metric comparison, training time comparison, and state-of-the-art comparison.

5.1. Model performance comparison

The precision-recall curve plots for the three models discussed is shown in Figure 8. The graphs show ResNet with a highest average precision (AP) of 0.994, followed by LSTM model with AP of 0.991 and CNN model with AP of 0.968. The ROC curve for LSTM, CNN, and ResNet models are as in Figure 9 which again shows ResNet model with highest AUC of 0.994, outperforming LSTM with AUC of 0.988 and CNN with AUC of 0.958. The performance comparison of the three models using accuracy, recall, precision, F1-score, specificity, and AUC as metrics is in Figure 10 and Table 1. The ResNet architecture demonstrates exceptional diagnostic capability with accuracy of 97.90%, significantly exceeding the performance of existing approaches which typically achieve 82% to 93% accuracy. Precision of 98.72% indicates exceptional specificity in positive COPD identification, while the recall of 96.86% ensures minimal missed cases—a critical consideration for clinical screening applications where false negatives have severe consequences for patient outcomes. The training time comparison of all the 3 models is as in Figure 11 with ResNet taking maximum training time of 32.2 seconds while 11.3 seconds for CNN and 10.9 seconds for LSTM. But considering the other metrics (as in Table 1), this delay is far more negligible given the substantial accuracy improvements and the non-real-time nature of diagnostic screening applications. Also, a comparison between the proposed ResNet architecture and existing approaches that use only audio signals as input to their model and achieve accuracies above 90% is shown in Figure 12. The accuracy of 93% is the work of [17], 92.86% accuracy is the work of [24], and 95% is the work executed by [25].

COPD prediction model evaluation: the confusion matrix values for LSTM, CNN, and ResNet models provide valuable insights into the classification performance of each model and is as follows:

- LSTM: true negatives: 171, false positives: 3, false negatives: 7, and true positives: 152.
- CNN: true negatives: 170, false positives: 4, false negatives: 20, and true positives: 139.
- ResNet: true negatives: 172, false positives: 2, false negatives: 5, and true positives: 154.

During validation the optimal threshold value obtained is 0.15 for LSTM, 0.29 for CNN and 0.36 for ResNet. The learning curves for all three models showed an appropriate convergence behavior, with validation metrics closely tracking training metrics, suggesting good generalization without significant overfitting. The ResNet model in particular demonstrated excellent stability in both loss and accuracy metrics during the training process. The 0.9937 AUC score indicates near-perfect discriminative capability, while the balanced precision-recall characteristics provide optimal trade-offs for clinical deployment. Based on this comprehensive evaluation, the ResNet architecture is recommended for deployment in the CLASP system, as it provides best diagnostic accuracy and reliability for clinical applications, with 96.86% recall and 98.85% specificity which is important for clinical use to minimize missed cases, despite its computational requirements.

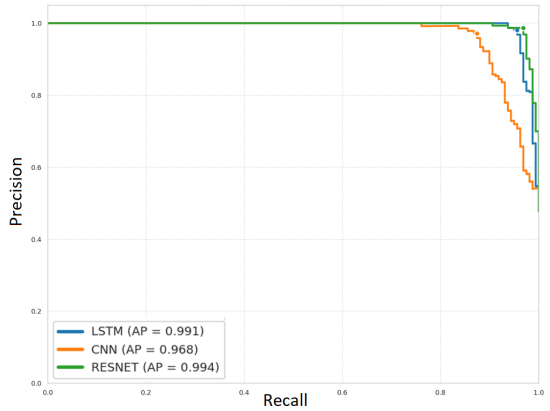

Figure 8. Precision-recall curve comparison

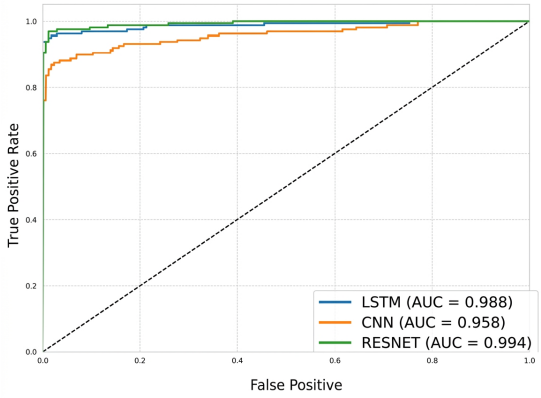

Figure 9. ROC curve comparison

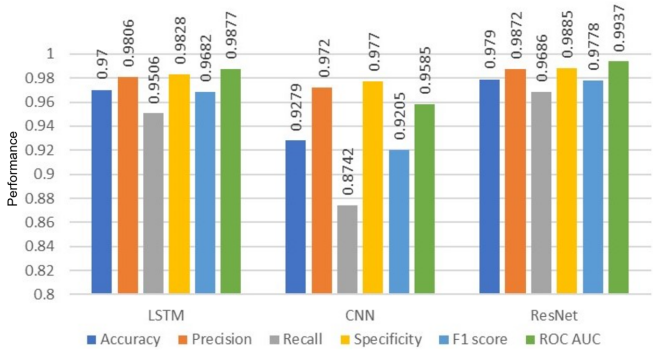

Figure 10. Performance metrics comparison

Table 1. Comprehensive comparison of performance for CLASP model architectures showing key metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	ROC-AUC
LSTM	97.00	98.06	95.60	98.28	96.82	0.9877
CNN	92.79	97.20	87.42	97.70	92.05	0.9584
ResNet	97.90	98.72	96.86	98.85	97.78	0.9937

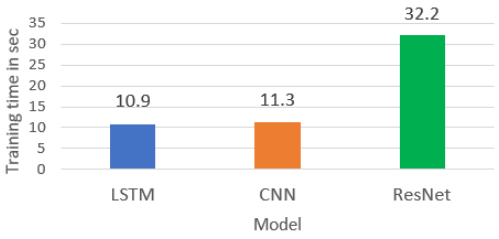

Figure 11. Training time comparison

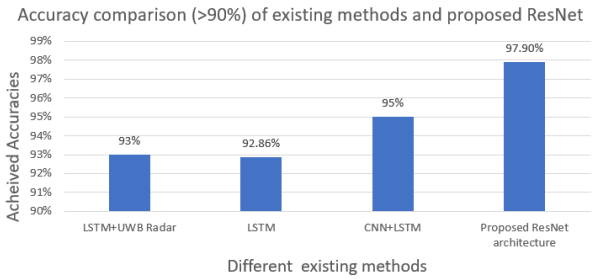

Figure 12. Accuracy of existing and proposed methods

6. CONCLUSION

The proposed model achieves a significant advancement over existing approaches. The combination of optimized MFCC feature extraction with ResNet architectures crosses the performance threshold required for practical clinical deployment in COPD screening programs. The limitation of the proposed approach is the dependence on high-quality audio recordings and the computational requirements of 32.2 seconds for ResNet architecture. Additionally, validation on larger and more diverse patient populations will be necessary to establish generalization across different demographic groups and recording conditions. The diagnostic principles established in this proposed work provide a pathway towards comprehensive respiratory health monitoring systems. By achieving clinical-grade accuracy in automated COPD detection, CLASP enables the development of accessible screening tools that could dramatically improve early intervention rates and patient outcomes in respiratory medicine.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Babitha Sudhakar Ullal	✓	✓		✓	✓	✓		✓	✓	✓		✓		
Veena Kalludi Narasimhaiah	✓	✓		✓	✓	✓				✓	✓	✓	✓	
Rithul Kamesh	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The supporting data of this study are openly available in Kaggle at https://www.kaggle.com/code/mariammagdy22/pulmonary-diseases-detection-system/input.

REFERENCES

[1] D. Singh *et al.*, "Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019," *European Respiratory Journal*, vol. 53, no. 5, 2019, doi: 10.1183/13993003.00164-2019.

[2] WHO, "Chronic obstructive pulmonary disease (COPD)," *World Health Organization*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(COPD)

[3] S. Banerjee, J. Khubchandani, C. Onukogu, C. Okpom, and M. Johnson, "Elevated C-reactive protein and mortality risk among COPD patients," *Egyptian Journal of Bronchology*, vol. 18, no. 1, 2024, doi: 10.1186/s43168-024-00291-0.

[4] J. Zhang, S. Wang, J. Courteau, L. Chen, G. Guo, and A. Vanasse, "Feature-weighted survival learning machine for COPD failure prediction," *Artificial Intelligence in Medicine*, vol. 96, pp. 68–79, 2019, doi: 10.1016/j.artmed.2019.01.003.

[5] A. Agusti *et al.*, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease," *Global Initiative for Chronic Obstructive Lung Disease*, 2021. [Online]. Available: https://goldcopd.org/wp-content/uploads/2020/11/GOLD-REPORT-2021-v1.1-25Nov20_WMV.pdf

[6] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, "CNN-MoE based framework for classification of respiratory anomalies and lung disease detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2938–2947, 2021, doi: 10.1109/JBHI.2021.3064237.

[7] H. Park *et al.*, "Diagnosis and evaluation of small airway disease and COPD using impulse oscillometry," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-79818-w.

[8] J. M. G. Blake, C. A. Senger, E. M. Webber, R. A. Mularski, and E. P. Whitlock, "Screening for chronic obstructive pulmonary disease evidence report and systematic review for the US preventive services task force," *JAMA - Journal of the American Medical Association*, vol. 315, no. 13, pp. 1378–1393, 2016, doi: 10.1001/jama.2016.2654.

[9] D. M. Huang, J. Huang, K. Qiao, N. S. Zhong, H. Z. Lu, and W. J. Wang, "Deep learning-based lung sound analysis for intelligent stethoscope," *Military Medical Research*, vol. 10, no. 1, 2023, doi: 10.1186/s40779-023-00479-3.

[10] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 3, pp. 535–544, 2020, doi: 10.1109/TBCAS.2020.2981172.

[11] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 58–69, 2018, doi: 10.1016/j.artmed.2018.04.008.

[12] T. M. Al-Hasan, M. Noorizadeh, F. Bensaali, N. Meskin, and A. A. Hssain, "Current trends and future orientation in diagnosing lung pathologies: a systematic survey," *Intelligent Medicine*, vol. 5, no. 1, pp. 23–36, 2025, doi: 10.1016/j.imed.2024.09.004.

[13] F. Lin *et al.*, "AutoCOPD–a novel and practical machine learning model for COPD detection using whole-lung inspiratory quantitative CT measurements: a retrospective, multicenter study," *eClinicalMedicine*, vol. 82, 2025, doi: 10.1016/j.eclinm.2025.103166.

[14] A. Idrisoglu, A. L. Dallora, A. Cheddad, P. Anderberg, A. Jakobsson, and J. S. Berglund, "COPDVD: automated classification of chronic obstructive pulmonary disease on a new collected and evaluated voice dataset," *Artificial Intelligence in Medicine*, vol. 156, 2024, doi: 10.1016/j.artmed.2024.102953.
[15] B. M. Rocha *et al.*, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, 2019, doi: 10.1088/1361-6579/ab03ea.
[16] D. K. Lee, J. S. Choi, S. J. Choi, M. H. Choi, and M. Hong, "Classification of chronic obstructive pulmonary disease (COPD) through respiratory pattern analysis," *Diagnostics*, vol. 15, no. 3, 2025, doi: 10.3390/diagnostics15030313.
[17] H. U. R. Siddiqui *et al.*, "Respiration-based COPD detection using UWB radar incorporation with machine learning," *Electronics*, vol. 11, no. 18, 2022, doi: 10.3390/electronics11182875.
[18] C. Abineza, V. E. Balas, and P. Nsengiyumva, "Deep learning model for early subsequent COPD exacerbation prediction," *Studies in Informatics and Control*, vol. 32, no. 3, pp. 99–107, 2023, doi: 10.24846/v32i3y202309.
[19] S. Mei *et al.*, "Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series," *npj Systems Biology and Applications*, vol. 11, no. 1, 2025, doi: 10.1038/s41540-025-00489-y.
[20] C. Yin *et al.*, "Fractional dynamics foster deep learning of COPD stage prediction," *Advanced Science*, vol. 10, no. 12, 2023, doi: 10.1002/advs.202203485.
[21] V. K. Bairagi and A. B. Kanwade, "Classification of chronic obstructive pulmonary disease (COPD) using electromyography," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 45, no. 1, 2020, doi: 10.1007/s12046-020-01336-y.
[22] S. Kumar *et al.*, "A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases," *Computer Methods and Programs in Biomedicine*, vol. 243, 2024, doi: 10.1016/j.cmpb.2023.107911.
[23] A. Ullah, M. S. Khan, M. U. Khan, and F. Mujahid, "Automatic classification of lung sounds using machine learning algorithms," *Proceedings - 2021 International Conference on Frontiers of Information Technology (FIT)*, 2021, pp. 131–136, doi: 10.1109/FIT53504.2021.00033.
[24] V. Nunavath, M. Goodwin, J. T. Fidje, and C. E. Moe, "Deep neural networks for prediction of exacerbations of patients with chronic obstructive pulmonary disease," *Communications in Computer and Information Science*, vol. 893, pp. 217–228, 2018, doi: 10.1007/978-3-319-98204-5_18.
[25] A. Jenefa, V. E. Naveen, V. Ebenezer, K. Rajkumar, M. Vargheese, and K. A. Jeyaraj, "COPD assessment through multimodal analysis: exploiting the synergy of CNNs and LSTM networks," *International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2023 - Proceedings*, 2023, pp. 13–18, doi: 10.1109/ICSSAS57918.2023.10331673.

BIOGRAPHIES OF AUTHORS

Babitha Sudhakar Ullal holds Bachelor of Engineering (B.E.) in Electronics and Communication from Golden Valley Institute of Technology, Visvesvaraya Technological University in 2002; M.Tech. in VLSI and Embedded Systems (gold medalist) from REVA University, Bengaluru in 2022; and is currently pursuing Ph.D. in artificial intelligence, REVA University. She is currently working as an assistant professor in Department of Electronics and Communication Engineering, PES University, Bengaluru. She has 12 years of teaching experience and 1 years of industry experience. She has 8 research publications in national and international conferences and journals. She has 4 patents filed and 1 patent grant. Her research areas are artificial intelligence and its application in the medical field. She can be contacted at email: babithas@pes.edu.

Dr. Veena Kalludi Narasimhaiah holds Bachelor of Engineering (B.E.) in Electronics from from Bengaluru University; M.Tech. in Digital Electronics and Advanced Communication from NITK Surathkal in 2001, Mangalore University; and Ph.D. in Computational Intelligence and Sensor Networks from Kuvempu University in 2015. She is currently working as a professor in School of Electronics and Communication Engineering, REVA University, Bengaluru. She has 22 years of teaching experience and 2 years of industry experience. She has more than 50 peer-reviewed research publications in national and international conferences and journals. She has 15 patents filed and 2 patent grants. Her research areas are artificial intelligence, networking, and robotics. She is currently guiding 5 Ph.D. students. She can be contacted at email: veenakn@reva.edu.in.

Rithul Kamesh pursuing Electronics and Communication Engineering (2027) with research interests in photonics, embedded systems, artificial intelligence, and machine learning. He has experience as a full-stack software developer. His interdisciplinary approach combines technical expertise with creative pursuits, exploring the intersection of interactive technologies and creative expression. His research areas focus on photonics applications, embedded systems design, and AI implementation in practical systems. He can be contacted at email: hi@rithul.dev.