Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

ATwC^{2}N^2: Automatic Chronic Obstructive Pulmonary Disease Classification Using Adaptive Two-Way Cascade Convolutional Neural Network

Nagendra Kumar V V.[1]**, Kavisankar L.**[2]

Department of Computing Technologies, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India.

E-mail: [1]vn5082@srmist.edu.in, [2]kavisanl@srmist.edu.in

Orcid ID: [1]0009-0009-3149-4326, [2]0000-0003-2984-5763

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a widespread lung condition that impacts the quality of life and lung functioning, leading to significant morbidity and mortality on a global scale. Non-invasive lung sound analysis is a method that has garnered growing interest in detecting COPD at an early stage and seeking immediate medical assistance. However, manually assessing respiratory sounds to determine the severity of COPD is subjective and prone to human error, often resulting in inaccurate assessments. This highlights the need for automated and accurate monitoring systems. This paper proposes an Adaptive Two-Way Cascade Convolutional Neural Network (ATwC^{2}N^2) optimized by the Adaptive Snake Swarm Optimization (AS^{2}O) algorithm for classifying COPD stages using respiratory sounds. The process involves cleaning and normalizing the respiratory sound, transforming it into a spectrogram for feature extraction and disease classification. The model utilizes a dual-pathway feature extraction mechanism that processes raw lung sound signals using a 1D CNN and spectrogram images using a 2D CNN simultaneously, integrating temporal and frequency domain features for superior classification performance. The AS^{2}O algorithm is used to optimize network hyperparameters, such as the learning rate, batch size, convolutional filters, dropout rate, and fully connected layer size, thereby enhancing feature learning, convergence stability, and overall classification performance. The proposed system categorizes subjects into four groups: healthy, mild COPD, moderate COPD, and severe COPD. The ATwC^{2}N^2 model achieves an accuracy of 97.14%, specificity of 96.01%, precision of 95.91%, recall of 98.32%, and an F1-score of 97.10%. The proposed model improves overall accuracy by 0.88%, 67.82%, 2.06%, and 3.07% compared to EMD-IMF-ellipse area from 2D-PSR, STFT + Wavelet, Expert Diagnostic System, and ALSD-Net, respectively.

Keywords: Pulmonary Disease, Convolutional Neural Network, Snake Swarm Optimization, Deep Learning.

DOI: https://doi.org/10.36548/jiip.2026.3.021
Received: 18.05.2026, received in revised form: 21.06.2026, accepted: 13.07.2026, published: 23.07.2026

1. Introduction

Pulmonary abnormalities constitute a major global health concern encompassing several fatal respiratory diseases such as COPD and pneumonia. Early diagnosis and effective medical intervention can greatly improve patient outcomes and lower mortality rates, although some diseases are life-threatening. COPD patients are especially susceptible to pneumonia, which is a deadly infection that can cause complications such as death [1]. UNICEF reports that infectious diseases and disorders, with pneumonia ranking as the most common in 2015, are to blame for half of the morbidity from the 5.9 million under-five fatalities that have been reported. The symptoms of pneumonia and an acute exacerbation of COPD are similar, making it challenging to distinguish between the two components [2]. By listening to the internal sounds of air passing through the lungs, a clinical procedure named auscultation is used to diagnose Respiratory Disease (RD). A stethoscope is frequently used to auscultate pulmonary sounds across the trachea, anterior, and posterior chest walls [3]. Auscultation is a medical examination in which a physician listens for abnormal lung sounds layered on top of normal breathing patterns. Wheezes, stridors, coarse or fine crackles, and pleural rubs are some of the commonly identified Adventitious Lung Sounds (ALS) [4].

Acoustic signals having frequencies between 110 Hz and 3 kHz are commonly represented as Lung Sounds (LS). The human ear can only detect waves with frequencies between 20 Hz and 20 kHz [5]. A manual stethoscope has limitations in terms of the lung sounds it can detect, which can lead to incorrect diagnoses of respiratory diseases. Consequently, during the auscultation process, essential information about the respiratory system conveyed by shorter-range waves is lost. To overcome the limitations of manual diagnosis, researchers have developed automatic systems that use deep learning to analyse lung sounds and detect respiratory diseases. Various machine learning algorithms, such as Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) and Artificial Neural Networks (ANN), have been employed for disease classification. Additionally, feature extraction techniques such as Higher-Order Statistics (HOS), time-frequency analysis, wavelet transform, Mel-Frequency Cepstral Coefficients (MFCCs) and the Hilbert–Huang Transform (HHT) are employed to extract relevant information from lung sounds prior to classification.

Classifying breathing sounds using traditional ML approaches such as Random Forest (RF) has demonstrated potential. However, these approaches rely heavily on manual extraction of features and construction of models that may lead to bias, making computation more complex and the applicability of results to a large population of patients less general. The time- and frequency-domain characteristics of the sounds of breathing are also difficult to choose, making it more challenging to discern minor issues. Moreover, they do not perform well on large or real-life clinical data, so they are not trusted in real healthcare practice. These issues demonstrate the significance of having automatic DL techniques that are capable of learning powerful models on raw data. The study added the most significant things, which are:

- The proposed $ATwC^2N^2$ considers both the 1D CNN-based lung sound data and spectrogram representations (2D CNN) and provides high-quality classification of the COPD stages due to its ability to simultaneously consider both the time-varying and frequency-based characteristics.
- A pre-processing approach based on wavelet smoothing, displacement artifact elimination, and Z-score normalization is employed to eliminate noise and artifacts and enhance the resulting features.

- The hyperparameter model has been optimized using the AS^{2}O method, which has shown improved learning, convergence stability, and object classification.
- Finally, the fully connected layers categorize respiratory sounds into four stages: Healthy, Mild COPD, Moderate COPD, and Severe COPD.

The primary novelty of ATwC^{2}N^2 is not only the integration of 1D and 2D CNN branches but also the use of AS^{2}O-driven optimization and adaptive cascade feature learning for stage-wise COPD severity classification. In this study, the proposed ATwC^{2}N^2 framework uses adaptive two-way cascade learning to progressively fuse temporal and spectral features, thereby improving feature complementarity. The AS^{2}O algorithm also optimises important hyperparameters to enhance convergence and feature learning. The unified framework proposed provides more discriminative feature representations than conventional dual-branch CNN architectures, achieving robust stage-wise COPD severity classification.

The remainder of this research is organized as follows: Section 2 provides an overview of the relevant work, the proposed ATwC^{2}N^2 model is described in Section 3 the analysis and results of the research are presented in Section 4, and Section 5 provides the conclusion.

2. Related Work

In recent years, extensive research has been carried out on the automatic analysis of respiratory sounds for detecting pulmonary diseases such as asthma, pneumonia, and COPD. Traditional Machine Learning (ML) approaches have relied heavily on handcrafted features like MFCCs, wavelet coefficients, and higher-order statistics, combined with classifiers such as SVM, KNN, and RF.

Roy, A., et al., [6] suggested the TriSpectraKAN model improves early diagnosis by utilizing audio-based lung sound characteristics. MFCCs, chromograms, and Mel spectrograms are used in the analysis of lung sounds using TriSpectraKAN, a hybrid model that combines spectral characteristics with the KAN. TriSpectraKAN shows good performance with an F1 score of 0.98, precision of 0.97, recall of 0.98, and accuracy of 93%. Luay Fraiwan et al., [7] investigate the applicability of various uniform ensemble learning approaches to categorize respiratory diseases under a variety of groups. Using differentiated classifiers and decision trees as basic learners, aggregation via bootstrap and adaptive boosting groups was built. The optimal structure of the ensemble models was determined using Bayesian hyperparameter optimization. All of the aforementioned algorithms were subjected to Bayesian optimization to determine the ideal training settings.

Naqvi et al., [8] provided a method to diagnose pneumonia and COPD by using a machine learning and signal processing technique. The lung sound is first cleaned and segmented. The most important combined features are then selected using the backward elimination method, which consists of EMD and DWT. By using the adaptive synthetic (ADASYN) sampling technique, class imbalance is eliminated. In 2021 Mukherjee, H., et al., [9] created a technique to identify breathing noises produced by patients who are suffering from respiratory infections. To describe respiratory sounds, the authors used features based on the LPCC. For classification purposes, the MLP classifier, a feed-forward artificial neural network is used. Their system was trained using a respiratory sound database that was publicly available and linked to the ICBHI.

Khan, S.I. and Pachori, R.B., [10] proposed a system for automatically dividing lung sounds into chronic and non-chronic groups. To produce intrinsic mode functions (IMFs), the LC signals are processed via empirical mode decomposition (EMD). Using a hybrid technique, IMFs were found throughout the dataset; four IMFs were selected for feature extraction. The analysis of the LS data used the entire signal without any segmentation. The ensemble classifier was then given the chosen features to distinguish between non-chronic and chronic lung sound data.

Tena, A., et al., [11] proposed for COPD detection that leverages the time–frequency characteristics of user-recorded breathing sounds. The model integrating sociodemographic data and respiratory features achieved an accuracy of 0.761, sensitivity of 0.836, specificity of 0.871, and an AUC of 0.901. This represents a significant improvement over the model relying solely on sociodemographic data, which had an AUC of 0.818. Sahu, P., et al., [12] proposed early diagnosis and classification of COPD according to severity. They proposed to use the statistical analysis of the samples of lung sounds. The proposed framework extracts features from the Open Respiratory Sound Database and the Respiratory@TR dataset. These are CSTFT, the MFCCs, the Mel-Spectrogram (mSpec), the Mel spectrogram-snippet and a format that integrates the three attributes. With two-class classification, this model obtained 95.76% accuracy and with multiclass classification has 93% accuracy.

Baghel N et al. [13] presented a ML technique for automatic classification technique for diagnosing different pulmonary illnesses using lung sounds. Doctors compiled and classified a large database of LS, which was then employed in a DL network. To diagnose asthma patients, this method made use of a neural network model due to its high efficacy in identifying lung noises like wheezing. Utilizing data augmentation approaches, the method trained and multi-classified lung sounds to increase robustness and enhance diagnosis accuracy in noisy conditions. Tripathy, R.K., et al., [14] proposed using LS signals to automatically identify pulmonary illnesses (PDs). Based on the properties of LS signals, classifiers like PDs, support vector machines, random forests, extreme gradient boosting, and LGBM have been selected to find automatically. The proposed technique outperformed the currently used techniques with the LS signals, with the scheme to classify regular, pneumonia, and asthma having an accuracy value of 84.76%.

Variational Inference Pre-Trained Audio Neural Networks (VI-PANNs) [15] present a Bayesian deep learning framework that combines transfer learning and uncertainty-aware audio classification. The model, based on a variational inference version of the ResNet-54 architecture and pre-trained on AudioSet, passes on both learned knowledge and calibrated epistemic uncertainty to downstream tasks. Results demonstrate the potential of Bayesian transfer learning for building more reliable and interpretable audio-based diagnostic systems. Efficient Pre-trained Audio Neural Networks (E-PANNs) [16], are a lightweight and resource-efficient framework for audio tagging and sound event recognition. Redundant parameters are removed by using model pruning techniques on the original PANNs architecture, which dramatically decreases both the computational complexity and the memory demand. The E-PANNs model reduces computational cost by 36% and memory usage by 70% with a slight improvement in sound recognition performance. The improvements make E-PANNs suitable for deployment on edge devices and large-scale acoustic monitoring systems with limited resources.

In this work [17], an innovative framework based on the Vision Transformer (ViT) for automatic diagnosis of respiratory diseases from breath sounds' spectrograms is presented. The model uses attention mechanisms and advanced data augmentation techniques to classify five respiratory diseases namely Normal, Asthma, Pneumonia, COPD and Lung Fibrosis. EBU normalization considerably improves the quality of features and classification accuracy. Authors [18] proposed a multi-scale transformer deep learning model for lung cancer surgery risk assessment by integrating imaging, clinical and genomic data. The framework employs self-attention and multi-scale feature extraction to accurately predict surgical outcomes and classify patients into risk categories. Experimental results show it to be robust, reliable, and to have the potential to advance precision medicine in healthcare.

This study [19] proposes a deep learning-based framework for COPD diagnosis using auscultation lung sounds for non-invasive and early diagnosis of respiratory abnormalities. The proposed approach uses advanced preprocessing, Mel spectrogram feature extraction and Vision Transformer (ViT-S1) model for the classification of COPD, healthy, and non-COPD cases. Data augmentation based on VQVAE-2 was performed to address the dataset imbalance and significantly improved diagnostic performance.

He, W. et al. [20] propose a Multi-View Spectrogram Transformer (MVST) which leverages multi-scale spectrogram patches to learn the complementary time-frequency characteristics of respiratory sounds. The extracted features are combined by a transformer-based self-attention mechanism and an adaptive gated fusion strategy to enhance the feature representation. The experimental results on ICBHI respiratory sound dataset demonstrate that the proposed method is superior to existing state-of-the-art methods in terms of classification performance. Kim, B. J. et al. [21] introduce an Audio Spectrogram Transformer (AST) using the transformer for the automatic classification of pediatric respiratory sounds from electronic stethoscope recordings. The model uses pre-trained transformer weights to capture long-range temporal dependencies, resulting in better performance than traditional CNN-based methods for wheeze detection and disease diagnosis.

Saky, S. M. et al. [22] proposed an explainable multimodal deep learning framework by integrating CNN–BiLSTM-attention spectro-temporal encoder with handcrafted acoustic features for multiclass respiratory sound classification. The model combines data-driven and domain-specific representations through late fusion and demonstrates promising internal performance, while providing interpretable predictions using Grad-CAM, Integrated Gradients and SHAP. The results show substantial dataset-shift effects, indicating the need for further generalization before clinical deployment. Sahu, P. et al. [23] proposed a multimodal framework for respiratory disease classification by combining FBSE, spectrogram and MFCC features with weighted feature fusion and BiGRU classifier. The model can differentiate between COPD, asthma and pneumonia at an early stage with an overall precision of 94.1%.

The literature review reveals that lung sound analysis has been applied previously to identify and name respiratory diseases. Many of them rely on hand-crafted features and statistical models that are not effective with intricate time and frequency structures. DL techniques are also applied in other methods that are not always effective with noise and errors. In addition, existing models are also unable to distinguish between severity levels of COPD due to the inability to examine local and global acoustic characteristics at the same time. To avoid such issues, a different ATwC^{2}N^2 approach to COPD diagnosis is proposed. It employs superior pre-processing, dual-path CNN feature extraction, and optimize flexibly to acquire superior stage-wise classification.

3. Proposed ATwC²N² Model

The ATwC²N² is proposed in this research to utilize breathing sounds to categorize the COPD stages. The processed signal is then converted to a spectrogram picture that can record the signal in the time domain and frequency domain. These spectrograms are processed by an adaptive two-way cascaded CNN. The lung sounds are first pre-processed with wavelet smoothing, displacement artifact removal and Z-score normalization to remove the noise as well as make the data more homogenous. The data are then converted to spectrogram pictures. Adaptive Snake Swarm Optimization (AS²O) techniques is used to achieve speedy and more precise learning. Lastly, a fully CNN classifies the extracted features into one of four categories Healthy, Mild COPD, Moderate COPD, and Severe COPD. It becomes possible to determine the severity of COPD using LS recordings with adequate precision and automaticity. Figure 1 describes the proposed ATwC²N² architecture.

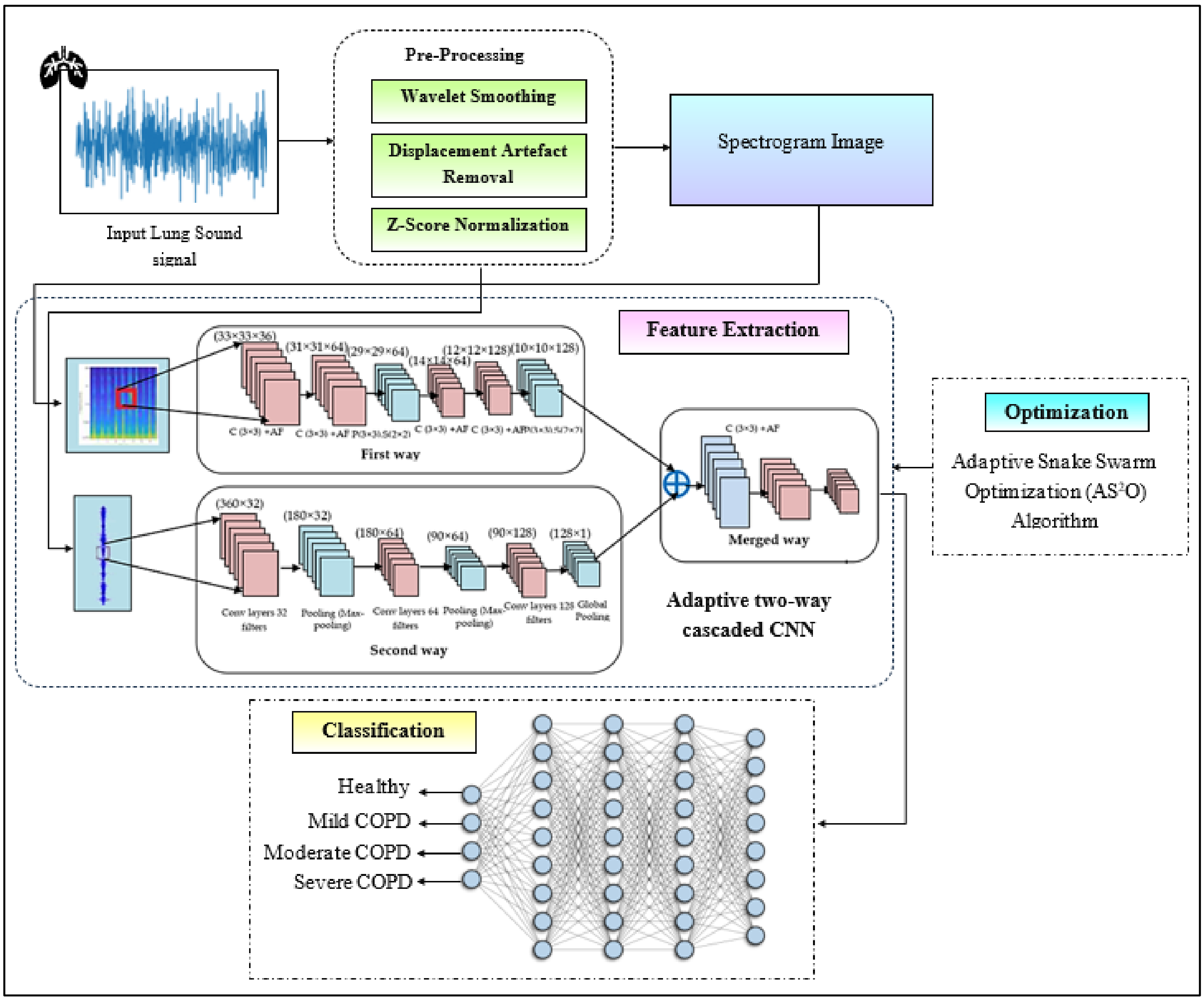

Figure 1. The overall workflow of the proposed ATwC²N²

3.1 Dataset Description

The Respiratory Sound Database was developed through a collaboration between experts dispatched from Spain and Greece. It contains 920 recordings of 126 individuals which have been labelled and range in length from 10 to90 seconds. The dataset on the breathing cycles across 5.5 hours shows that 6,898 breathing cycles are recorded, out of which 1,864 include crackles, 886 include wheezes, and 506 include both crackles and wheezes as shown in Table1. Both recordings have distinct sounds of breathing and noisy sounds that are more similar to those in real life.

The ICBHI Respiratory Sound Database does not provide clinically validated COPD severity labels. To address this issue, an acoustic severity-label generation process was employed. COPD recordings were identified using available metadata and classified based on respiratory acoustic characteristics such as wheeze intensity, crackle occurrence, breath sound intensity, dominant spectral patterns, and breathing irregularity. Heart rate, dominant frequency, and signal amplitude were additional acoustic descriptors, rather than clinical diagnostic criteria. Therefore, the proposed severity groups are not based on the Global Obstructive Lung Disease (GOLD) classification or spirometry, but are acoustically derived labels for machine-learning evaluation.

Table 1. Calculation of COPD disease classification based on respiratory sound signal

Severity Group	Respiratory Acoustic Characteristics	Supplementary Acoustic Indicators	Number of Patients
Healthy	Normal breath sounds with regular respiratory pattern	Heart rate: 70–90 bpm; dominant frequency: 100–250 Hz; relatively higher signal amplitude	40
Mild COPD	Mild wheezes with occasional crackles and slight breathing irregularity	Heart rate: 90–110 bpm; dominant frequency: 150–300 Hz; moderate signal amplitude	28
Moderate COPD	Frequent wheezes and crackles with noticeable respiratory irregularity	Heart rate: 100–120 bpm; dominant frequency: 150–400 Hz; reduced signal amplitude	32
Severe COPD	Persistent wheezes, prominent crackles, reduced breath sound intensity, and pronounced respiratory irregularity	Heart rate: 110–130 bpm; dominant frequency: 200–500 Hz; low signal amplitude	26

3.1.1 COPD Severity Label Generation and Clinical Validation

The ICBHI Respiratory Sound Database does not include standardised annotations of COPD severity, such as 'mild', 'moderate' or 'severe'. In this study, we employed an acoustic approach to classify severity. We examined respiratory sounds for their dominant frequency range, signal amplitude, wheeze strength, the presence of crackles, and breathing irregularities. First, COPD recordings were selected based on available metadata. The machine learning-based classification severity levels in Table 1 were established using auditory indicators. It should be noted that these categories are not intended to replace the standard GOLD criteria based on spirometry data (e.g. FEV1 and FEV1/FVC ratios), nor are they clinically confirmed grades of COPD severity. The Healthy, Mild, Moderate and Severe classes proposed can thus be considered as severity groups for calculation purposes.

3.1.2 Dataset Partitioning Strategy

After generating the severity labels, we split the dataset into training, validation and testing subsets to develop and evaluate the proposed ATwC^{2}N^2 model. A multilayer partitioning strategy was employed to preserve the class distribution across all subsets. Specifically, 70% of the recordings were used for training, 15% for validation, and the remaining 15% for testing. To avoid data leakage, we evaluated the model using the patient-wise partitioning method. All the breathing data and cycles for each participant were included in the training and validation subgroups. Therefore, no patient was included in more than one subset. This domain-neutral evaluation approach provides a more robust performance measure and better reflects the model's ability to generalise to new contexts.

3.2 Pre-processing

Patterns of respiratory sounds are preprocessed to improve them and prepare them to be classified and feature extracted. An initial application of wavelet smoothing is to remove high-frequency noise but retain the crackles and wheezes in breathing sounds. Then, a displacement artifact removal is performed to eliminate distortions and ensure that the signals accurately represent actual breathing activity. Finally, z-score normalization is applied to bring the data to a better level of consistency by placing all the values on the same scale, centered at the mean and divided by the standard deviation. These measures are combined to eliminate noise, correct artifacts, and standardize for case-to-case differences, increasing reliability and consistency in the analysis.

3.2.1 Wavelet Smoothing

Wavelet smoothing is a signal processing technique used to reduce high-frequency noise in non-stationary signals. Unlike Fourier transform, which decomposes a signal into complex sinusoids, wavelet transform breaks a signal into a set of wavelets. Wavelet smoothing is based on the core concept of translating and dilating a signal into various functions using an initial wavelet. The mother wavelet is a mathematical function employed to analyze the signal at different scales and positions.

$$\chi_{pq}(n) = |p|^{-1/2}\chi\left(\frac{n-q}{p}\right) \tag{1}$$

Equation (1) denotes n as the time instance, p as the dilation function, and q as the translation function. In this work, the Discrete Wavelet Transform (DWT) is performed using a level 4 soft Maximum Overlap Discrete Wavelet Transform (MODWT) based on the db5 parent wavelet. Level 4 of the MODWT with the db5 wavelet was chosen because it effectively reduced noise while preserving important respiratory sounds, such as wheezes and crackles. Decomposition reduced noise at lower levels to some extent but lost sensitive breathing details at higher levels. The db5 wavelet is ideal for feature extraction in respiratory sound analysis due to its superior time-frequency localization. This value was selected to limit local volatility while preserving clinically significant respiratory events (Loess span = 0.1). Level-4 MODWT decomposition using the db5 wavelet provided an effective compromise between noise reduction and preservation of minor respiratory characteristics. Lower decomposition levels had little effect on denoising, but at a higher level, some significant auditory information was lost.

The respiratory sound data is processed using the wavelet denoising tool wden() in MATLAB to reduce noise and improve signal quality. The one-dimensional respiratory audio signals are directly fed into the denoising procedure before feature extraction and spectrogram generation. This pre-processing enhances the detection of clinically relevant acoustic events, such as wheezes and crackles, while retaining the necessary respiratory features for categorization.

$$\overline{\psi}_{c,n} = \sum_{d=0}^{D_c-1} \overline{j}_{c,d}\, X_{n-d} \tag{2}$$

$$\overline{\gamma}_{c,n} = \sum_{d=0}^{D_c-1} \overline{g}_{c,d}\, X_{n-d} \tag{3}$$

In equations (2) & (3) where $\overline{\psi}_{c,n}$ is the wavelet coefficient, $\overline{\gamma}_{c,n}$ is the scaling coefficient, $\overline{j}_{c,d}$ and $\overline{g}_{c,d}$ are represents the high-pass and low-pass filters, correspondingly, X_n represents the level of decomposition and the endless sequence.

In non-stationary signals such as lung sounds, wavelet smoothing effectively reduces high-frequency noise. By reducing the noise in the signal, wavelet smoothing enhances the accuracy of the deep learning model in identifying lung diseases from lung sounds. The Two-Way feature involves extracting temporal and spectral properties simultaneously from raw respiratory signals using a 1D CNN and from spectrogram representations using a 2D CNN. The cascade method improves feature complementarity by gradually combining features from both branches, including intermediate representations before final classification. In the AS^{2}O-based optimization method, adaptive refers to the dynamic adjustment of network parameters and feature learning weights during training. Compared with the dual-branch CNN with fixed feature fusion, our adaptive cascade learning enhances feature discrimination, convergence stability, and classification performance.

3.2.2 Displacement Artifact Removal

Displacement artifact removal is a pre-processing step that is utilized to remove unwanted noise caused by the displacement of the electronic stethoscope during the recording of lung sounds. Any stethoscope movement has the potential to overlay a low-frequency, wave-shaped signal on top of the relevant signal, contaminating the lung sound signal's structure and lowering the DL model.

A Loess function is employed to eliminate the displacement artifact. Loess is a non-parametric regression method that aligns the data points locally to a polynomial function to create a smooth curve. The function is weighted so that points closer to the fitted line have higher weights and points farther away have lower weights. The function is denoted as,

$$\omega(y) = (1 - |l|^3)^3 \tag{4}$$

In equation (4) which l is the scaled distance between each point and the fitted curve, ranging between 0 and 1.

3.2.3 Z- score Normalization

Z-score normalization is a further pre-processing step that creates a normalized signal with a standard deviation of 1 and an average value of 0. This stage is critical since it aids in the elimination of any patterns in the signal that overpower the lesser ones, which might impair the performance of DL algorithms. As a result, the signal has a mean value (τ) of 0 and a standard deviation (ρ) of 1. The Z-score normalization is given in equation (5),

$$z = \frac{z-\tau}{\rho} \tag{5}$$

DL algorithms work more effectively if there is a distinct signal with no trend fluctuations over time.

Table 2. Signal quality evaluation before and after preprocessing

Metric	Before Preprocessing	After Preprocessing
SNR (dB)	18.42	27.83
MSE	0.0214	0.0086
RMSE	0.1463	0.0927

The impact of the pre-processing steps was evaluated by comparing the signal quality metrics before and after wavelet smoothing and artifact removal. Noise reduction and signal retention were measured using the signal-to-noise ratio (SNR), mean square error (MSE) and root mean square error (RMSE). Table 2 demonstrates that the proposed preprocessing framework effectively improves the quality of respiratory sounds by suppressing noise and motion artifacts while preserving diagnostically relevant acoustic features.

3.2 Spectrogram Generation

In the temporal domain, examine the digital audio signal. $f(t)$ represents the continuous time domain audio signal, where t is the time index. An analysis of the amplitude variations over time provides insight into events such as crackles and wheezes. To capture respiratory sounds' complex frequency components, spectral or time-frequency analysis must be combined with temporal analysis. The Discrete Fourier Transform (DFT) is a mathematical operation used to achieve this conversion using equation (6),

$$F(m) = \sum_{n=0}^{N-1} f(n) e^{-\frac{i2\pi}{N}mn} \tag{6}$$

A fast way to determine the DFT of a sequence is to utilize the FFT. For magnitude scaling, the audio signal $f(n)$ is transformed to decibels after the Fourier coefficients have been calculated using,

$$f_{dB} = 10\,log(|f|) \tag{7}$$

Next, the 50x34 pixel spectrogram image is created to display the frequency-time dependence of the audio signal's amplitude or intensity in dB. The spectrograms were resized to 50 x 34 pixels to allow an efficient trade-off between computational efficiency and preservation of discriminative respiratory acoustic information. This resolution preserves the important time-frequency features required for the classification of COPD severity, while reducing the memory and computational complexity during the training of the CNN. Lower resolutions may result in the loss of subtle spectral patterns associated with wheezes and crackles, and much higher resolutions increase computational cost with little increase in performance. Data from a single audio recording is contained in a single spectrogram. By analyzing the spectrogram from left to right, the variation of frequencies and their corresponding intensities in the audio signal was observed over time.

3.3 Adaptive Two-Way Cascade CNN

The Adaptive Two-Way Cascade Convolutional Neural Network (ATwC^{2}N^2) is employed as a feature extraction framework to effectively capture both local and global characteristics of lung sound signals. The cascaded two-way architecture allows for the simultaneous extraction of useful information, in contrast to conventional single-path CNNs that limit the focus to either global structures or local patterns. By processing raw lung sound signal data and spectrogram-based features together, the model is able to develop more discriminative and richer representations. By adaptively determining the cascade weight values, the AS^{2}O algorithm optimizes this process so that the most pertinent aspects are focused. The ATwC^{2}N^2 architecture processes the pre-processed lung sound signals. The resultant significant feature embeddings serve as the foundation for the following classification of pulmonary disease.

The proposed ATwC^{2}N^2 integrates three complementary design principles. The Two-Way architecture consists of two parallel feature extraction branches: a 1D CNN that learns temporal features directly from respiratory sound signals, and a 2D CNN that extracts spectral representations from spectrogram images. The Cascade strategy allows for richer respiratory representations than independent feature processing, by progressively integrating the complementary features extracted from both branches using a feature-level fusion before the final classification. The adaptive behavior is realized by the AS^{2}O algorithm, which dynamically optimises the network hyperparameters during training to improve convergence stability and feature learning. Different from the conventional dual-branch CNN architecture with fixed architecture and simple feature concatenation, the proposed framework accomplishes adaptive optimization and cascade feature integration simultaneously, producing more discriminative feature representations for COPD severity classification.

The architecture consists of two paths: a local path and a semi-global path. The local path extracts feature from a small region around the target object, while the semi-global path extracts feature from a larger region around the target object. To classify the target item, the result of both paths is concatenated and sent into a classification layer. This method can increase classification accuracy by taking into account each local and semi-global feature surrounding each target object. A two-pathway CNN framework is suggested, with the first path processing 3×3 patches and extracting the local features (local pathway) from the spectrogram images and the second path processing 1-D CNN patches and obtaining the features (global pathway) from the lung signals. In a two-way cascaded CNN architecture for classifying lung sounds into healthy, mild COPD, moderate COPD, and severe COPD, two distinct input branches are utilized. The first input branch feeds spectrogram images created on the signal of lung sounds into a 2D CNN.

The second input branch is a 1D CNN designed to deal directly with raw lung sounds. It removes characteristics of the audio data, such as breathing patterns, variations in sound strength and frequency, as well as harmonic and nonharmonic components. By combining these characteristics, the strengths of the two approaches are integrated and applied collectively to ensure that the network can obtain information from both time and frequency domains. The classifier provides a probability distribution encompassing the various classes. The predicted class is the one with the highest probability score. This technique classifies lung sounds by utilizing information from the time and frequency domains to discern between regular sounds and those indicative of COPD. Figure 2 illustrates the structure of the two-way cascaded CNN.

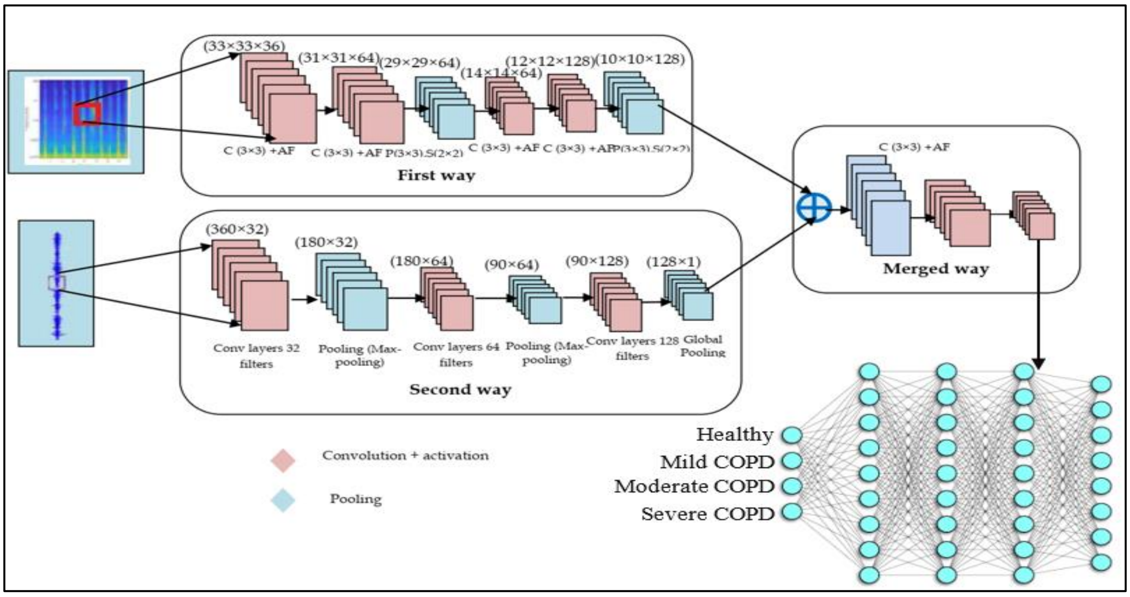

Figure 2. Architecture of two-way cascaded CNN

The proposed ATwC^{2}N^2 architecture uses a 2D convolutional neural network branch to extract spectral features and a 1D CNN branch to extract temporal information. The feature vector extracted by the 1D CNN is named (F_{1D}). The symbol for the feature vector extracted by the 2D CNN is (F_{2D}). Concatenation of features by,

$$F_{fusion} = [F_{1D}, F_{2D}] \tag{8}$$

The fused feature vector is denoted by F_{fusion}. A fully connected layer then receives the concatenated features,

$$F_c = ReLU(WF_{fusion} + b) \tag{9}$$

Where W and b are trainable weights and biases, respectively. To predict the severity class of COPD, the Softmax classifier receives the processed feature vector. The model may use the spectral and temporal information of respiration due to the fusion technique, which improves classification accuracy.

The loss of classification is measured using the softmax log loss function. The softmax log loss forward transform is expressed as follows,

$$L = -\,log\frac{e^{j_1}}{\sum_{x=1}^{m} e^{j_x}} = log\left(\sum_{x=1}^{m} e^{j_x}\right) - j_y = log(Z) - j_y \tag{10}$$

$$Z = \sum_{y=1}^{m} e^{j_y} \tag{11}$$

In equation (10) & (11) where y denotes the class index and m denotes the number of classes. During the back propagation, $\frac{\partial z}{\partial x}$ is calculated by equation (12),

$$\frac{\partial z}{\partial x} = -\frac{1}{b_y}\left(\left[\varphi_{yx} b_y\right] - b_y b^K\right) = -\frac{1}{b_y}\left(\left[0, \ldots, b_y, \ldots, 0\right] - b_y[b_1, \ldots, b_c]\right) - \left[0, \ldots, 1_y, \ldots, 0\right] \tag{12}$$

3.3.1 AS²O Algorithm for Optimization

In this section, AS²O optimization is applied to improve the classifier's accuracy, efficiency, and overall reliability in detecting COPD diseases. The hyperparameters are then utilized to optimize the model's accuracy and performance. The results are compared to manually selected hyperparameters. A new intelligent optimization technique called Snake Optimization (SO) algorithm, which mathematically models the feeding, fighting, and mating behaviours of snakes. The two phases of feeding are called exploitation and exploration. The intricate and interesting tactics for the survival of snakes are amazing. The population is divided into males and females using snake optimization, in contrast to previous metaheuristic techniques. Randomly generated populations are used to build up SO. Furthermore, the ambient temperature influences how snakes eat and copulate because they are cold-blooded creatures. The step-by-step process of hyperparameter optimization is presented below;

Step 1: Solution Initialization

Each search agent in the algorithm represents a candidate solution in the search space. The parameters of two-way cascaded CNN including weight (ψ) and bias (ℓ) are taken as solutions. The population's initialization is defined by equation (13),

$$S_n = \{x_1, x_2 ..., x_n\} \tag{13}$$

Where, x_n depicts the nth solution and it is given by equation (14),

$$x_n = \{\psi_1, \psi_2, ... \psi_i; \ell_1, \ell_2, ... \ell_i\} \tag{14}$$

Step 2: Fitness Calculation

This algorithm evaluates fitness according to accuracy. The fitness computation is outlined as follows in equation (15),

$$Fitness_N = Max(Accuracy) \tag{15}$$

Step 3: Update the Solution

The population convergence speed is significantly increased in future stages; however, the population individuals are more prone to the phenomenon of accumulation, which decreases the algorithm's performance during the ideal search. The snake optimization algorithm's population position update equations throughout the development stage are as follows.

In this case, the sample size of men is believed to be equal with that of women, which is half each. The population is divided into two categories; the females and the males. The division of the group is done using the following equations (16) & (17),

$$N_m \approx 0.5 * N \tag{16}$$

$$N_f \approx N - N_m \tag{17}$$

Where the numbers of candidates are represented by and represent the numbers of women and men in the population, respectively. The food quantity (G)can be offered by equation (18),

$$G = v_1 \cdot exp\left(\frac{n-N}{N}\right) \tag{18}$$

The temperature ($temp$) is estimated via the following equation (19),

$$temp = exp\left(-\frac{n}{N}\right) \tag{19}$$

- n →The number of current iterations
- N →The maximum number of iterations

$G < 0.25$ denotes a lack of food in the surroundings, implying that snakes are in the stage of exploration and are searching for food at random. The following is how the exploring phase is expressed in equation (20) & (21),

$$Z_{q,m}(n+1) = Z_{R,m}(n) \pm v_2 \cdot D_m \cdot ((Z_{max} - Z_{min}).R + Z_{min}) \tag{20}$$

$$Z_{q,f}(n+1) = Z_{R,f}(n) \pm v_2 \cdot D_f \cdot ((Z_{max} - Z_{min}).R + Z_{min}) \tag{21}$$

The following equation (22) is used to determine the extent to which males can locate food:

$$D_m = exp\left(-\frac{f_{q,f}}{f_{q,m}}\right) \tag{22}$$

Equation 23 can be used to test the ability of women to find food:

$$D_f = exp\left(-\frac{f_{q,m}}{f_{q,f}}\right) \tag{23}$$

- $Z_{q,m}(n+1)$and $Z_{q,f}(n+1)$ → the male and female snakes' separate placements have been passed down to the next generation.
- D_m and D_f →The mating ability of the male and female snakes, respectively
- $f_{q,m}$and $f_{q,f}$ →The male and female snakes' respective fitness values for their places.

There is enough food accessible in the environment throughout the exploitation period. If the, $temp > 0.6$ snakes remain to look for food. Equation (24) was used to update the male and female locations.

$$Z_{q,r}(n+1) = Z_{food}(n) \pm v_2 \cdot temp \cdot R \cdot (Z_{food} - Z_{q,r}(n)) \tag{24}$$

$$G \in (0.25,1] and\ temp \in (0.6,1]$$

- $Z_{q,r}(n+1)$ →The optimal position of the snake in $(n+1)$iterations.
- Z_{food} → The ideal location for the target food.
- v_1 andv_2 →Constants
- R →A random number into [0; 1]

- $Z_{q,r}$ →The current individual position of the snake

The equations that follow are used to modify the population position during the development stage of the snake optimization algorithm's mating mode, in equation (25) & (26),

$$Z_{q,m}(n+1) = Z_{q,m}(n) + v_2 \cdot D_m \cdot R \cdot \left(G \cdot Z_{q,f}(n) - Z_{q,m}(n)\right) \tag{25}$$

$$G \in (0.25,1] and temp \in (0,0.6)$$

$$Z_{q,f}(n+1) = Z_{q,f}(n) + v_2 \cdot D_f \cdot R \cdot \left(G \cdot Z_{q,m}(n) - Z_{q,f}(n)\right) \tag{26}$$

$$G \in (0.25,1] and temp \in (0,0.6)$$

$Z_{q,m}(n+1)$ and $Z_{q,f}(n+1)$ →The new generation of individual positions of the qth male and female snakes respectively.

Algorithm 1: Optimization of two-way cascaded CNN parameters using the AS²O algorithm

Input: Weight ($\boldsymbol{\psi}$), bias ($\boldsymbol{\ell}$), and coefficient factors
Output: Optimal weight ($\boldsymbol{\psi}$) and bias ($\boldsymbol{\ell}$)
Initialize the population randomly
Evaluate the fitness for each solution using equation (15)
Divide the population $\boldsymbol{N}$into two equal sections $\boldsymbol{N_m}$ and $\boldsymbol{N_f}$ using equations (16) and (17)
Define $(\boldsymbol{G})$ utilizing equation (18)
Denote $\boldsymbol{temp}$ utilizing equation (19)
if $(\boldsymbol{G} < \mathbf{0.25})$ then
Execution Exploration Phase by the equations (20) and (21)
else if $(\boldsymbol{temp} > \mathbf{0.6})$ then
Exploration Phase of execution by the equation (24)
else
Perform mating mode using equations (25) and (26)
Update equations in mating mode by merging the snake dynamic adaptive weight using equations (27) and (28)
end if
end if
Output the optimal solution

The SO algorithm is used for optimization; however, its mating population converges rapidly, making it prone to stagnation in local optima. To overcome this limitation and ensure more effective optimization in the classification process, the snake population's optimal positions are diversified, and a dynamic adaptive weight is incorporated. This adaptive weight improves the reproductive ability of individuals in the subsequent generation, thereby enhancing exploration and exploitation balance. Consequently, the snake position update equations in the mating mode are modified as shown in Equations (27) and (28),

$$Z_{q,m}(n+1) = Z_{q,m}(n) + \omega \cdot D_m \cdot R \cdot \left(G \cdot Z_{q,f}(n) - Z_{q,m}(n)\right) \tag{27}$$

$$Z_{q,f}(n+1) = Z_{q,f}(n) + \omega \cdot D_f \cdot R \cdot \left(G \cdot Z_{q,m}(n) - Z_{q,f}(n)\right) \tag{28}$$

The calculation of the snake dynamic adaptive weight is as follows in equation 29:

$$\omega = 1 + \frac{e^{n\beta}}{10N} \cdot cos\left(\frac{\pi n}{\gamma N} + \frac{\pi}{2}\right) \tag{29}$$

β →The adaptive range coefficient, defines the snake population's location update range.γ → The adaptive density coefficient, which establishes the snake population's update density.

Step 4: Termination

The algorithm continuously updates the weights and biases of the Two-Way Cascaded CNN until the best solution is found. Once the optimal solution is obtained, the optimization process stops.

The proposed ATwC^{2}N^2 model was fine-tuned for key hyperparameters, such as the learning rate, batch size, number of convolutional filters, dropout rate, and size of the fully connected layer, using the AS^{2}O method. The search space is determined by the following parameters: learning rate [0.0001, 0.01], batch size [16, 128], number of convolutional filters [16, 128], dropout rate [0.1, 0.5] and number of fully connected neurons [64, 512]. The optimization process was allowed to run for a maximum of 100 iterations, with a population size of 30 viable solutions. Optimization ended when either the maximum number of iterations was reached or classification accuracy did not improve significantly in subsequent rounds.

Instead of using the traditional dual-branch CNNs that extract temporal and spectral features separately and then concatenate them, the proposed ATwC^{2}N^2 framework is able to gradually fuse complementary representations through adaptive cascade feature learning. The AS^{2}O algorithm also optimizes the network hyperparameters in the training to make the convergence stability and feature discriminability better. Our proposed framework differs from other existing respiratory sound classification models with adaptive cascade learning and optimization, thus achieving improved performance in COPD severity classification.

4. Results and Discussion

The proposed model is implemented in MATLAB (2020b) and can be successfully run on an Intel Xeon with an NVIDIA graphics card, with the condition that the processing times are sufficiently low to be used in the medical field. The model requires approximately 12GB of memory and performs optimally using a powerful graphics card, such as the NVIDIA RTX 3090. This setup was tested on the suggested ATwC^{2}N^2 to determine its ability to classify COPD classes using this framework.

Figure 3 presents an experimental outcome of the COPD classification process. The first column shows the input lung sound signals. The second column depicts the signals after artifact removal, where displacement and unwanted disturbances are eliminated to produce a cleaner waveform. The spectrograms are displayed in the third column and indicate the sound of lungs as per their time and frequency and how they sound in different circumstances. The fourth column displays extraction of features, which is the process of extracting the significant spatial and temporal features of the spectrograms and signals to display characteristics that are unique to the disease. Finally, the fifth column displays the classification results, where the processed inputs are categorized into four classes: Healthy, Mild COPD, Moderate COPD, and Severe COPD, clearly showing the model’s ability to differentiate between disease stages.

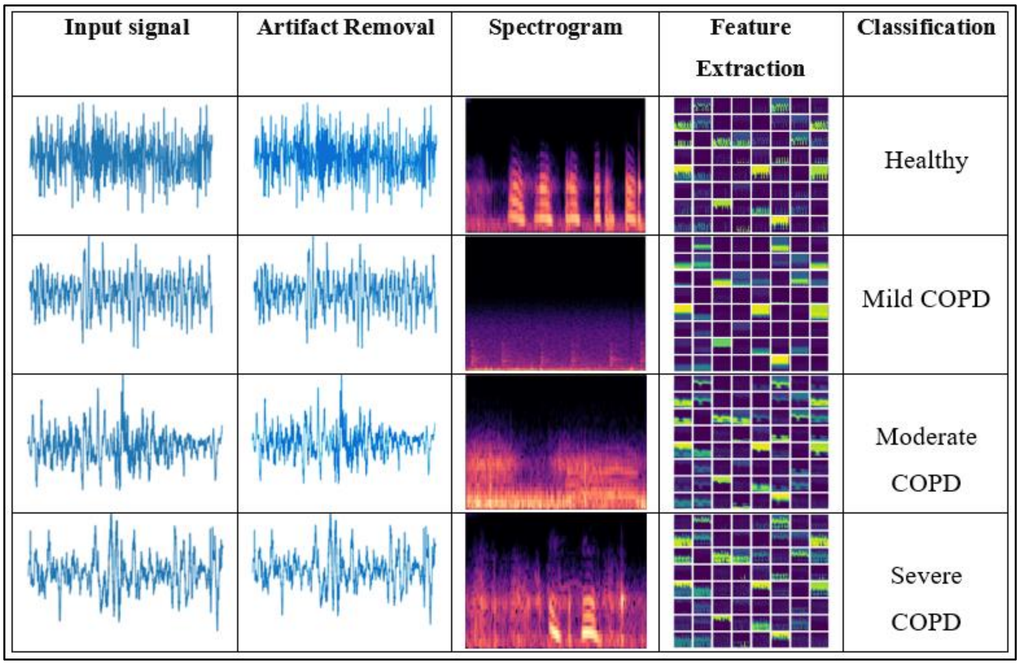

Figure 3. Experimental results of the ATwC^{2}N^2 model

4.1 Performance Analysis

We evaluated the proposed system using the accuracy (AY), specificity (SY), recall (RL), precision (PN), and F1 score (FS) of the ATwC^{2}N^2 model.

$$SY = \frac{TN}{TN+FP} \tag{30}$$

$$PN = \frac{TP}{TP+FP} \tag{31}$$

$$RL = \frac{TP}{TP+FN} \tag{32}$$

$$AY = \frac{TP+TN}{TP+FP+FN+TN} \tag{33}$$

$$FS = \frac{2XPNXRL}{PN+RL} \tag{34}$$

Each thing is denoted by true positive (TP), true negative (TN), false positive (FP) as well as false negative (FN).

Table 3. Efficiency analysis of the proposed model

Classes	AY	SY	PN	FS	RL
Healthy	97.74	96.61	96.51	97.70	98.92
Mild COPD	97.34	96.21	96.11	97.30	98.52
Moderate COPD	96.94	95.81	95.71	96.90	98.12
Severe COPD	96.54	95.41	95.31	96.50	97.72

Table 3 demonstrates that the ATwC^{2}N^2 model can be used to explain the various COPD stages well. In every class of COPD, the factors are AY, SY, PN, RL, and FS. The fact that the algorithm is obtaining an FS of 97.10% overall COPD classes shows that the algorithm provides a good balance between PN and RL. Correctness and recall scores are always able to ensure low levels of FP and FN. The proposed model is also highly accurate on all levels with the average AY of 97.14. According to this test, the proposed model is more useful in the search of COPD classes.

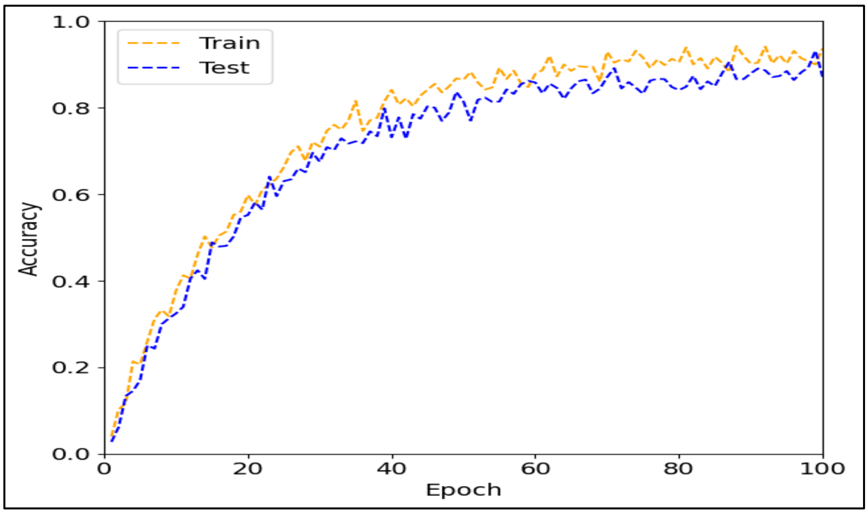

Figure 4. Accuracy curve of the proposed ATwC^{2}N^2 model

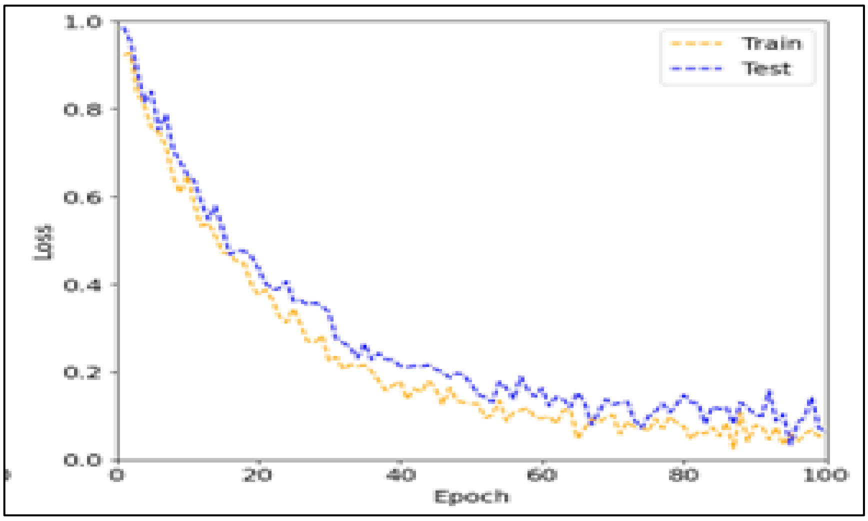

Figure 5. Loss curve of the proposed ATwC^{2}N^2 model

The proposed model was applied to sort cases of COPD 100 times, and the accuracy curve (Figure 4) and the loss curve (Figure 5) were presented.In case the training curve is slightly above the test curve, it indicates that the person is learning effectively and improving with time. The proposed model is always able to reduce classification errors, demonstrating its effectiveness. Scores of testing and training are virtually parallel to one another, indicating that the model is generalizing and not overfitting. The proposed ATwC2N2 model appears to be trained properly, with minimal loss and high accuracy.

Figure 6 presents CM to sort COPD cases into groups. Large diagonal values imply that the identifications were correct, whereas off-diagonal values imply that classes were wrong. The proposed model is highly effective as it can distinguish healthy people (97.74%), mild COPD (97.34%), moderate COPD (96.94%), and severely COPD (96.54%). However, moderate is confused with mild as the two most frequently confused severity levels that are near each other. The findings indicate that the ATwC2N2 model does a decent job of grouping various COPD cases.

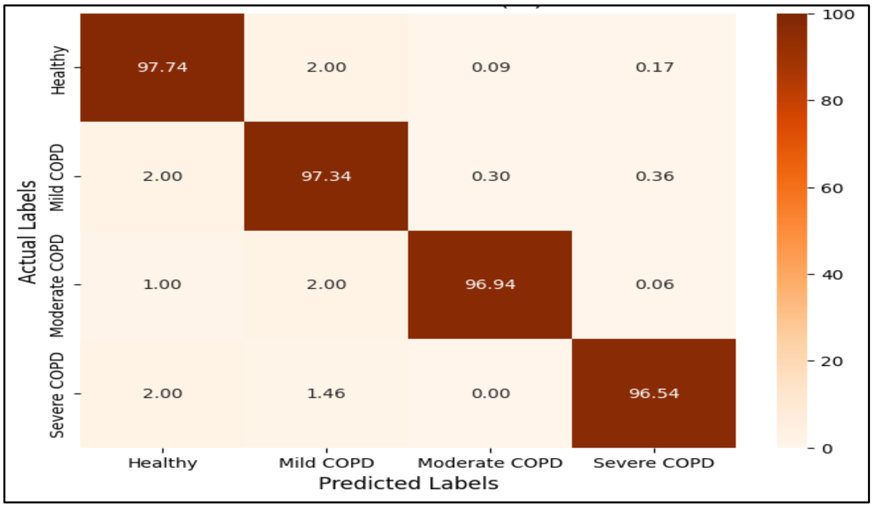

Figure 6. Confusion matrix (CM) of the ATwC^{2}N^2 model

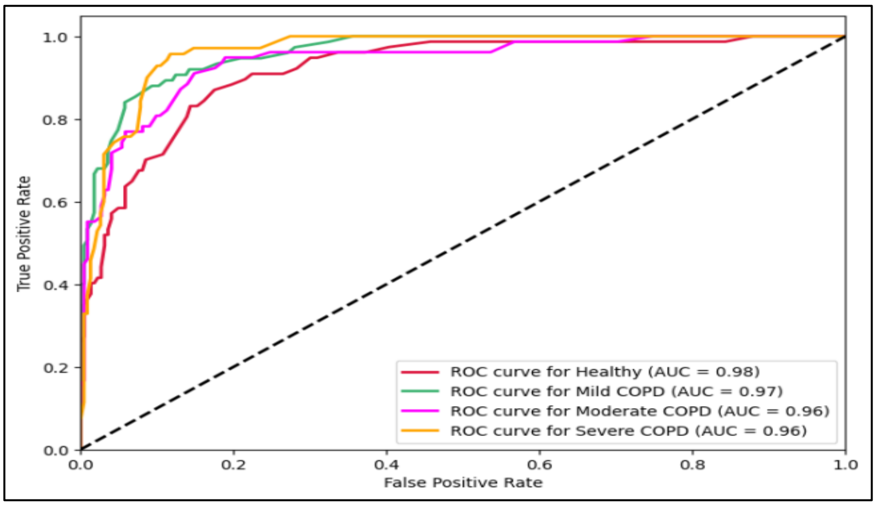

Figure 7. ROC curve of the ATwC^{2}N^2 model

Misclassification Analysis: In the confusion matrix, most errors in classification are made between close severity levels of COPD, especially between the Mild and Moderate classes. This misclassification pattern can be explained by the considerable overlap of respiratory acoustic features, including wheeze characteristics, the occurrence of crackles, and spectral signature, which tend to become more similar as the disease stages get closer. On the other hand, Healthy and Severe COPD classes have more distinguishable respiratory patterns, resulting in fewer classification errors. The results showed that the proposed model could capture important acoustic variations related to severity. However, it is difficult to separate closely related intermediate stages because COPD progression is continuous and gradual.

Class Distribution and Error Analysis: The dataset has a moderate imbalance in the distribution of the classes Healthy, Mild, Moderate, and Severe COPD, which can harm the learning of the model by biasing it toward the majority classes and decreasing sensitivity to the underrepresented classes. To mitigate this, we used a stratified data split to preserve class proportions during training and evaluation. Overall error analysis shows that the classification performance is strongly related to the class representation and the degree of acoustic similarity of the disease stages. Healthy and Severe COPD samples were classified with high accuracy due to their distinctive respiratory sound characteristics, but increased error levels were observed in the intermediate severity categories, indicating the challenges of learning subtle disease progression patterns from respiratory acoustic signals.

Figure 7 indicates the ROC curve of the proposed ATwC2N2 model for all classes of COPD. According to the AUC, a graph will indicate the difference between the rates of TP and FP for a particular class. The ROC analysis was then interpreted quantitatively by deriving the AUC values for each COPD severity category. The ATwC2N2 model proposed AUC values of 0.981, 0.974, 0.969, and 0.962 for the healthy, mild, moderate, and severe COPD groups, respectively. The proposed framework demonstrates high discriminative ability, distinguishing between COPD severity levels with a mean AUC of 0.972.

4.2 Comparative Analysis

In this section, various performance measures are applied to compare the proposed ATwC^{2}N^2 model with other COPD classification models. The obtained dataset is also utilized to test and compare the performance of other current deep learning models with that of the proposed ATwC^{2}N^2 model. Comparing the collected data with this, it is possible to conclude that the proposed ATwC^{2}N^2 model is more effective when compared to VI-PANN, ViT Transformer, Transformer Model, and VQVAE-2 model.

Figure 8 compares the performance of the proposed ATwC^{2}N^2 model with that of other well-known deep learning approaches. The accuracy of the Vision-Informed Pretrained Audio Neural Network (VI-PANN) was 95.1%; the Vision Transformer achieved 91.04%; and the Transformer achieved 94.2%. VQVAE 2 improved accuracy to 95.42%. The ATwC^{2}N^2 model achieved a higher accuracy (97.14%), recall (98.3%) and F1 score (97.1%) than the other techniques. This indicates its suitability for extracting additional respiratory information in the temporal and spectral domains for COPD severity classification.

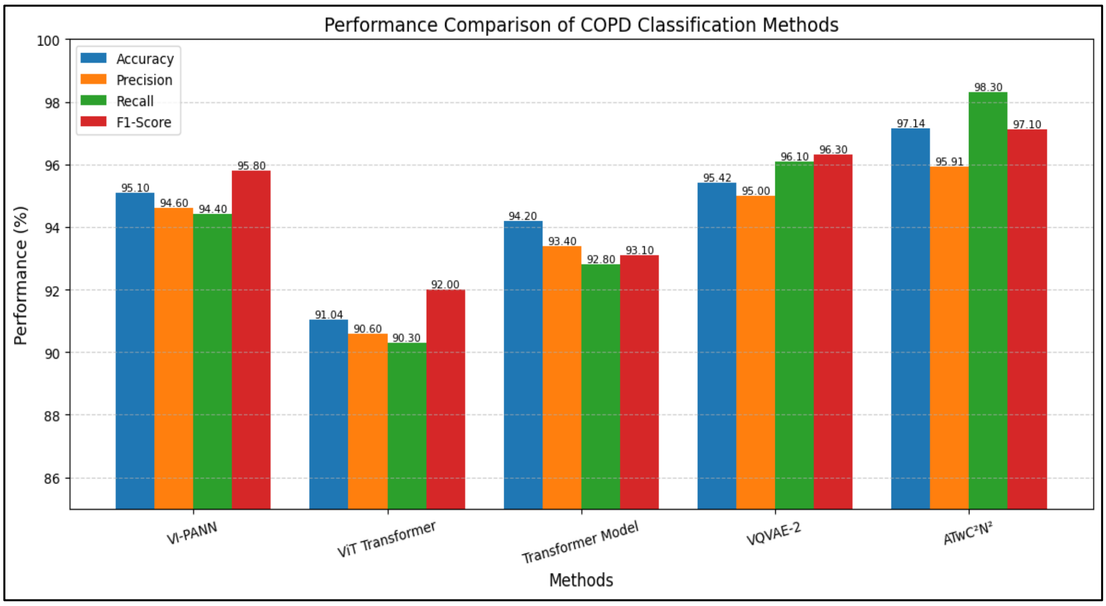

Figure 8. Performance comparison of COPD classification methods

4.2.1 Cross-Validation Analysis

A five-fold cross-validation study was conducted to evaluate the reliability of the proposed ATwC^{2}N^2 model. The model performed consistently well across all folds, achieving an average accuracy of 97.14% ± 0.24%, a precision of 95.91% ± 0.28%, a recall of 98.32% ± 0.23%, a specificity of 96.01% ± 0.28%, and an F1 score of 97.10% ± 0.24%. The low values of standard deviation demonstrate that the proposed approach is robust and reliable across different data segments, as shown in Table 4.

Table 4. Five-fold cross-validation results of the proposed ATwC^{2}N^2 model

Fold	AY (%)	PN (%)	RL (%)	SY (%)	FS (%)
Fold 1	96.82	95.53	98.05	95.72	96.77
Fold 2	97.41	96.12	98.61	96.35	97.35
Fold 3	97.08	95.84	98.24	95.93	97.02
Fold 4	97.36	96.21	98.48	96.27	97.29
Fold 5	97.03	95.85	98.22	95.78	97.07
Mean± SD	97.14±0.24	95.91 ± 0.28	98.32 ± 0.23	96.01 ± 0.28	97.10 ± 0.24

4.2.2 Computational Complexity and Deployment Analysis

The classification performance and computational components of the proposed ATwC^{2}N^2 model were investigated. The model has a processing speed of 1.8 GFLOPs per inference and around 2.3 million trainable parameters. The experimental platform had an average inference delay of 18 ms per respiratory recording, and an average training time of 42 minutes. The results demonstrate that the proposed framework can accurately predict COPD severity in real time. Additionally, due to its small size and low computational requirements, the framework can be used in edge-assisted healthcare applications and computer-aided respiratory monitoring systems.

4.2.3 Explainability and Robustness Analysis

To improve the interpretability of the model, we used Grad-CAM visualisation to identify the spectrogram segments that influenced the algorithm's categorisation decisions. Robustness analysis under different noise levels shows stable performance, with accuracy rates of 97.14%, 96.82%, 96.21% and 94.73% for clean signals and SNRs of 30, 20, and 10 dB, respectively. This steady degradation in performance indicates the robustness of the proposed ATwC^{2}N^2 architecture with respect to realistic recording settings and background noise, as shown in Table 5.

Table 5. Robustness analysis under different noise conditions

SNR Level (dB)	AY (%)	PN (%)	RL (%)	FS (%)
Clean Signal	97.14	95.91	98.32	97.10
30 dB	96.82	95.54	98.05	96.76
20 dB	96.21	94.88	97.42	96.10
10 dB	94.73	93.35	95.84	94.58

4.3 Ablation and Optimization Analysis

The impact of AS^{2}O optimization on the proposed ATwC^{2}N^2 framework is depicted in Table 6. The optimized model performed better than the non-optimized Two-Way Cascaded CNN, with an increased accuracy from 91.22% to 97.14%, and respective improvements in precision, recall, specificity, and F1-score. The experimental results show that AS^{2}O can effectively tune the model hyperparameters to improve the feature learning and convergence performance.

Table 6. Impact of AS²O hyperparameter optimization

Model	AY (%)	PN (%)	RL (%)	SL (%)	FS (%)
Two-Way Cascaded CNN (Without AS²O)	91.22	90.14	92.08	89.73	91.05
ATwC²N² (With AS²O)	97.14	95.91	98.32	96.01	97.10
Improvement (%)	+5.92	+5.77	+6.24	+6.28	+6.05

Table 7. Ablation study of the proposed ATwC²N² framework

Configuration	AY (%)
1D CNN	87.70
1D CNN + Wavelet Smoothing	89.15
1D CNN + Wavelet Smoothing + Artifact Removal	90.03
1D CNN + 2D CNN (Without Feature Fusion)	91.22
1D CNN + 2D CNN + Feature Fusion	94.68
1D CNN + 2D CNN + Feature Fusion + AS²O	97.14

The ablation study results of the proposed ATwC²N² framework are shown in Table 7. We started from the baseline 1D CNN with 87.70% accuracy and added wavelet smoothing, artifact removal, and 2D CNN branch one by one to gradually improve the performance. Feature fusion improves accuracy up to 94.68% fusing temporal and spectral information. The AS²O optimization achieved the highest accuracy of 97.14%, which shows that each component has a positive contribution to the overall framework.

Table 8. Impact of spectrogram resolution on classification performance

Spectrogram Resolution	AY (%)	PN (%)	RL (%)	FS (%)
32 × 32	94.82	93.76	95.41	94.57
50 × 34	97.14	95.91	98.32	97.10
64 × 64	97.32	96.08	98.45	97.26
128 × 128	97.41	96.14	98.53	97.34

Table 8 shows the influence of various spectrogram resolutions on the classification performance. The accuracy, precision, recall, and F1-score achieved by 32 × 32 resolution were 94.82%, 93.76%, 95.41% and 94.57%, respectively, which showed that low image resolution results in a loss of information. The selected 50 × 34 resolution improved the performance to 97.14% accuracy and 97.10% F1-score. For the 64 × 64 and 128 × 128 resolutions, the accuracies obtained were 97.32% and 97.41%, respectively, but the improvement in terms of computational complexity was not significant. The resolution of 50 x 34 was chosen as a good compromise between classification performance and computational efficiency.

The 128 × 128 spectrogram resolution achieved the highest classification accuracy (97.41%) but its advantage over the 50 × 34 resolution (97.14%) was only 0.27 percentage points, indicating that increasing the spectrogram resolution past 50 × 34 yields marginal performance improvement. However, the larger input size significantly increases the computational complexity, memory consumption and training and inference time because of the larger number of input features that the network has to process. On the other side, the 50 × 34 resolution can efficiently extract the discriminative respiratory sound patterns with significantly lower computational cost. The extra computation cost brought by the practical deployment and model efficiency perspectives exceeds the marginal improvement of the 128 X 128 resolution. Thus, the resolution of 50 x 34 spectrogram was selected to be the best input size for the proposed framework.

5. Conclusion

The study proposed an Adaptive Two-Way Cascade Convolutional Neural Network with Adaptive Snake Swarm Optimization (ATwC^{2}N^2) framework for COPD severity classification from respiratory sound recordings. The proposed method consists of wavelet-based pre-processing, dual-branch feature extraction from temporal and spectral representations, feature fusion and AS^{2}O based hyperparameter optimization to enhance the classification performance. The experimental evaluation on ICBHI Respiratory Sound Database provided promising results with an accuracy of 97.14%, precision of 95.91%, recall of 98.32%, specificity of 96.01%, and F1-score of 97.10%. Our results show that respiratory acoustic features contain relevant information about the severity levels of COPD and that the proposed framework is able to efficiently learn the complementary temporal and spectral features. However, the results should be considered in the light of the dataset and the experimental settings used in the study. The COPD severity categories used in this study were based on respiratory acoustic features rather than clinically confirmed GOLD-stage annotations. No external validation on independent datasets was performed. Future clinical validation is also required to confirm the association between acoustic patterns and disease progression. Thus, the proposed framework should be regarded as a computational decision support approach for respiratory sound analysis rather than a clinically validated diagnostic system. In future work, the proposed ATwC^{2}N^2 model can be improved for real-time clinical applications by integrating with portable digital stethoscopes and validating on larger datasets for effective early COPD diagnosis.

References

[1] Ariefiansyah, Muhammad Nurman, Yana Aurora Prathita, and Heidy Agustin. "Pneumonia in Lung Cancer Patients: Clinical Challenges, Diagnostic Pitfalls, and Management Strategies." International Journal of Medicine and Public Health 2, no. 2 (2025): 34-44.

[2] Zhu, Xiaodan, and Changxing Shen. "Development of a Novel Model for Accurate Prediction of Secondary Candida Pneumonia in Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease." Frontiers in Medicine 12 (2025): 1594934.

[3] Pessoa, Diogo, Bruno Machado Rocha, Maria Gomes, Guilherme Rodrigues, Georgios Petmezas, Grigorios-Aris Cheimariotis, Nicos Maglaveras et al. "Ensemble Deep Learning Model for Dimensionless Respiratory Airflow Estimation Using Respiratory Sound." Biomedical Signal Processing and Control 87 (2024): 105451.

[4] Saif, Engr Ahmed Ali, Syeda Zainab Mousavi, Haris Anwar, Syed Arslan Mustafa, Muhammad Faraz, and Syed Zohaib Hassan Naqvi. "Smart Diagnostic System for Classification of COPD and Healthy Subjects from Lung Sound Analysis." In 2024 26th International Multi-Topic Conference (INMIC), IEEE, 2024, 1-6.

[5] Ravikiran, Anaghaa. "Development of Enhanced Feature Set for Improved Performance on Classification of Stridor V/S Other Respiratory Auscultation." Master's thesis, Georgia Institute of Technology, 2025.

[6] Roy, Abhinav, Bhavesh Gyanchandani, Aditya Oza, and Anurag Singh. "TriSpectraKAN: A Novel Approach for COPD Detection via Lung Sound Analysis." Scientific Reports 15, no. 1 (2025): 6296.

[7] Fraiwan, Luay, Omnia Hassanin, Mohammad Fraiwan, Basheer Khassawneh, Ali M. Ibnian, and Mohanad Alkhodari. "Automatic Identification of Respiratory Diseases from Stethoscopic Lung Sound Signals Using Ensemble Classifiers." Biocybernetics and Biomedical Engineering 41, no. 1 (2021): 1-14.

[8] Naqvi, Syed Zohaib Hassan, and Mohammad Ahmad Choudhry. "An Automated System for Classification of Chronic Obstructive Pulmonary Disease and Pneumonia Patients Using Lung Sound Analysis." Sensors 20, no. 22 (2020): 6512.

[9] Mukherjee, Himadri, Priyanka Sreerama, Ankita Dhar, Sk Md Obaidullah, Kaushik Roy, Mufti Mahmud, and K. C. Santosh. "Automatic Lung Health Screening Using Respiratory Sounds." Journal of Medical Systems 45, no. 2 (2021): 19.

[10] Khan, Sibghatullah I., and Ram Bilas Pachori. "Automated Classification of Lung Sound Signals Based on Empirical Mode Decomposition." Expert Systems with Applications 184 (2021): 115456.

[11] Tena, Alberto, Ivan Juez-Garcia, Iván D. Benítez, Francesc Clariá, Jessica González, Jordi de Batlle, and Francesc Solsona. "Chronic Obstructive Pulmonary Disease Screening Using Time–Frequency Features of Self-Recorded Respiratory Sounds." JAMIA open 8, no. 4 (2025): ooaf083.

[12] Sahu, Prakash, Santosh Kumar, and Ajoy Kumar Behera. "Soundnet: Leveraging Deep Learning for the Severity Classification of Chronic Obstructive Pulmonary Disease Based on Lung Sound Analysis." In 2024 IEEE international conference on electronics, computing and communication technologies (CONECCT), IEEE, 2024, 1-6.

[13] Baghel, Neeraj, Vivek Nangia, and Malay Kishore Dutta. "ALSD-Net: Automatic Lung Sounds Diagnosis Network from Pulmonary Signals." Neural Computing and Applications 33, no. 24 (2021): 17103-17118.

[14] Tripathy, Rajesh Kumar, Shaswati Dash, Adyasha Rath, Ganapati Panda, and Ram Bilas Pachori. "Automated Detection of Pulmonary Diseases from Lung Sound Signals Using Fixed-Boundary-Based Empirical Wavelet Transform." IEEE Sensors Letters 6, no. 5 (2022): 1-4.

[15] Fischer, John, Marko Orescanin, and Eric Eckstrand. "VI-PANN: Harnessing Transfer Learning and Uncertainty-Aware Variational Inference for Improved Generalization in Audio Pattern Recognition." IEEE Access 12 (2024): 33347-33360.

[16] Singh, Arshdeep, Haohe Liu, and Mark D. Plumbley. "E-PANNs: Sound Recognition Using Efficient Pre-Trained Audio Neural Networks." In Inter-Noise and Noise-Con Congress and Conference Proceedings, vol. 268, no. 1, Institute of Noise Control Engineering, 2023, 7220-7228.

[17] Zhu, Pengfei, Tingmin Wang, Fan Yang, Meng Wang, and Yunjie Zhang. "A Transformer-Based Multi-Scale Deep Learning Model for Lung Cancer Surgery Optimization." IEEE Access (2025).

[18] Zhang, Xiangqing, Junyi Fu, Wei Wang, and Lu Yu. "Diagnosis of Chronic obstructive Pulmonary Disease Based on Deep Learning and Auscultation Lung Sound." Biomedical Signal Processing and Control 112 (2026): 108438.

[19] Aljaddouh, Batoul, D. Malathi, and Feisal Alaswad. "Multimodal Disease Detection and Classification Using Breath Sounds and Vision Transformer for Improved Diagnosis." Procedia Computer Science 235 (2024): 1436-1444.

[20] He, Wentao, Yuchen Yan, Jianfeng Ren, Ruibin Bai, and Xudong Jiang. "Multi-View Spectrogram Transformer for Respiratory Sound Classification." In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2024, 8626-8630.

[21] Kim, Beom Joon, Jeong Hyeon Mun, Dae Hwan Hwang, Dong In Suh, Changwon Lim, and Kyunghoon Kim. "An Explainable and Accurate Transformer-Based Deep Learning Model for Wheeze Classification Utilizing Real-World Pediatric Data." Scientific reports 15, no. 1 (2025): 5656.

[22] Saky, SM Asiful Islam, Md Saiful Arefin, Md Rashidul Islam, Mohammad Saiful Islam, Rashadul Islam Sumon, Md Mostafizur Rahman Masud, Maria Lapina, Mikhail Babenko, and Mohammed Muthanna. "Explainable Multi-Modal Deep Learning for Recording-Level Classification of Respiratory Audio Signals Under Internal and Domain-Shift Evaluation." Life 16, no. 7 (2026): 1108.

[23] Sahu, Prakash, Santosh Kumar, and Ajoy Kumar Behera. "DeepFusion: Early Diagnosis Of COPD, Asthma, and Pneumonia Using Lung Sound Analysis with a Multimodal BiGRU Network." Computer Methods in Biomechanics and Biomedical Engineering (2025): 1-14.