Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Wheeze Signal Feature Engineered Deep Neural Network Model for Wheeze Segmentation and Wheeze Sound Detection from Lung Sound Data Signals

Gomathi R[1*], KarthikR[2]

[1,2]Department of Computer Science,Sri Krishna Adithya College of Arts and Science, KovaiPudur,Coimbatore - 641 042,India.

Corresponding Author: gomathiphd85@gmail.com

Abstract

A study of the lung sound plays an essential role in diagnosing respiratory diseases, with wheezing being one of the most important features indicating respiratory diseases such as asthma and Chronic Obstructive Pulmonary Disease (COPD). The classical methods of detecting wheezes used in traditional techniques of auscultation are dependent on the experience of the physician and thus caused inconsistency and disparity in diagnosing the symptoms. Computerized Respiratory Sound Analysis (CORSA) and sophisticated machine learning (ML) algorithms have been used to achieve higher accuracy in detecting wheezing. But, ML methods have limitations based on data size and might have difficulties with big data. A promising alternative to that is deep learning (DL) methods, but there are still issues with overfitting and class imbalance. Other optimized models such as AlexNet and VGG16 had improved performance, and harmonic and percussion features can complicate classification. To deal with the challenges, the present research suggests a Wheeze Signal Feature Engineered Deep Neural Network (WSFEDNN) model that can identify and categorize lung sounds, i.e., wheezing and so on, in a multi-stage approach. First, it uses Empirical Mode Decomposition (EMD) to break down respiratory sound signals into Intrinsic Mode Functions (IMFs), which contain non-linear and non-stationary oscillating patterns. These IMFs are then processed with a Power-Phase Harmonic-Percussive Source Separation (PP-HPSS) which breaks these IMFs down into harmonic (Power IMFs) and percussive (Phase IMFs) parts, maintaining amplitude and phase information that is important to use when features are to be processed. Subsequently, feature engineering identifies significant patterns in the processed data including lagged observations as a way of capturing temporal dependence and temporal features as a way of analysing timing and cyclic changes and statistical features as a way of quantifying sound intensity, variability and distribution. Lastly, the model utilises a CNN-LSTM to apply a hybrid combining spatial feature extraction together with a temporal pattern recognition with a softmax classification of the features as either wheezing or normal. The combined methodology makes sure that it adequately detects wheezing using the frequency domain and time domain information. The analysis of the experiment revealed that the WSFEDNN performed better as compared to other existing baseline models, HPSS and RNN-LSTM with a precision of 93.48% and 91.23, respectively. This highlights the prospects of WSFEDNN model on real world clinical applications, leading to timely identification and treatment of respiratory disabilities and ultimately patient recovery.

Keywords: Wheezing, HPSS, Deep Learning, Empirical Mode Decomposition, Feature Engineering, Classification

This is an open access article under CC BY 4.0, allowing unrestricted use with proper attribution, a license link, and indication of any changes made.

1. Introduction

Lung sounds serve as critical indicators of respiratory health. Abnormal lung sounds including wheezing may indicate underlying conditions. Wheeze refers to a high-pitched whistling sound that occurs when respiratory airways are partially obstructed during the process of breathing. The presence of wheezes serves as a diagnostic indicator utilized by medical professionals for the identification of lung and chronic pulmonary diseases, including asthma, COPD and bronchiolitis etc... [1]. Typically, wheezes persist for an extended length of time, as a result, it is important to detect and control the symptoms before the patient's respiratory state worsens. In order to find the right therapy, it is necessary to accurately detect even mild wheezes [2]. Currently, auscultation techniques are widely used for detecting respiratory problems due to its non-invasive, easy-to-implement nature [3]. However, these methods rely heavily on the physician experience as it introduces variability in the detection and interpretation of wheezing potentially leading to inconsistent diagnoses.

Therefore, there is a rising need to improve the precision and consistency of wheeze diagnosis by incorporating cutting-edge technology like Computerized Respiratory Sound Analysis (CORSA). CORSA identifies and interprets crackles and wheezes and similar sounds to identify respiratory problems. This non-invasive technique provides objective data, long-term documentation and graphical support and aids in early diagnosis and monitoring of respiratory diseases, which can often inform about severity earlier than alternative methods. Normal respiratory sound (NS) is over a broad spectrum of frequencies, most of the energy being focused in the 60Hz to 1000Hz range, but, Wheezing Sounds (WS) is more musical, sinuously shaped with frequencies of 100Hz-1000Hz and durations greater than 100 ms according to CORSA [4]. However, Variability in respiratory sounds among patients still remains a challenge.

The latest progress in automated respiratory sound analysis systems involves Machine Learning (ML) algorithms to improve the accuracy of wheezing detection. It encompasses a number of methods such as Support Vector Machine (SVM) [5], Random Forest [6], k-Nearest Neighbors (k-NN) [7], Logistic Regression (LR) [8], Decision Tree (DT) [9] and Extreme learning machine (ELM) [7]. These technologies make it possible to monitor and analyze the respiratory sounds in real-time. To identify wheeze sounds with the help of ML algorithms, the most prominent features identified are the frequency ones like Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCC) [10]. Other important characteristics are the statistical aspects such as entropy, kurtosis, spectral flatness (SF) and skewness [11]. All these features lead to the increased accuracy of wheeze sound classification in ML models.

In [12], the authors presented a ML-based multi-stage segmentation approach that precisely divides the respiratory sounds into inhale and exhale phases and used k-NN, SVM and ELM approaches as a classifier. Despite the advancements in ML, the effective training of ML models works well only on small datasets and may degrade the performance if the number of signals used are more.

To counteract these shortcomings, Deep Learning (DL) methods were developed. Since DL, and CNNs in particular, are so good at processing complicated and huge respiratory sound datasets, they have become increasingly popular for application in wheeze identification. In [13], the author developed a CNN model to enhance the accuracy of wheezing detection and classification which produced more accurate results when compared to conventional ML models. But CNN models exhibited overfitting, class imbalance and exploding gradient issues while training. The prediction of respiratory infections through transforming segmented sound signals (into Bump and Morse scalograms) into an optimized AlexNet model [14] was created. The other solution [15] applied Transfer Learning (TL) to VGG16 to discriminate between normal and abnormal respiratory sounds, and further sub-classify abnormal respiratory sounds into its type.

The above DL models make use of manually designed features such as Intrinsic Mode Functions (IMFs), MFCCs and deep features to classify diseased lung sounds. Nevertheless, the presence of both harmonic and percussioned features in the feature extraction process may adversely affect the performance of the classifier. Elements that deal with pitch, such as the tonal components, are the harmonic elements, and those that deal with transient sounds that resemble impacts and enhance rhythm are the percussive elements; a combination of the two can make it difficult to classify and analyse audio.

To mitigate all these challenges, the authors in [16] proposed an automatic wheeze segmentation method that integrates Harmonic-Percussive Source Separation (HPSS) with Empirical Mode Decomposition (EMD). This method isolates the harmonic component of wheezes using median filtering and a separation factor, effectively reducing interference from background noise. However, HPSS relies solely on power spectrograms, leaving the phase information intact for resynthesizing the signal. Furthermore, the effectiveness of rule generation from wheeze candidates decreases as signal or database size increases.

Having all the concerns noted, this study introduces a feature decomposition technique that integrates EMD and Power and Phase aware HPSS (PP-HPSS) methods to obtain IMFs and harmonic enhanced IMFs. To begin representing non-linear and non-stationary signals, the EMD approach iteratively extracts oscillatory modes from the signal, decomposing it into iteratively-maintained fields (IMFs). Because it learns from the data, EMD excels at extracting the signal's intrinsic properties. Following this, PP-HPSS is applied to produce harmonic enhanced IMFs by considering both amplitude and phase in complex-valued spectrograms, thereby improving smoothness through convex optimization. By combining these two advanced techniques, the proposed framework enhances the analysis and extraction of relevant features, leading to improved model accuracy in the study.

Afterward, the element of feature engineering is utilized to narrow in on the input data preparation towards modeling. This entails the establishment of a vector of input variables that encompasses lagged input variables, time features and statistical features. Subsequently, CNN-LSTM model is fed on these input features, where it reads and extracts deep features. The choice of the number of lagged observations is informed by the autocorrelation and partial autocorrelation functions which make certain that the most useful historical data is incorporated in the modeling. Lastly, in our CNN-LSTM model, the softmax classifier is used to classify the lung sounds. The suggested approach is dubbed as the Wheeze Signal Feature Engineered Deep Neural Network (WSFEDNN).

The rest of the research is shown below: When it comes to wheeze detection, Section II takes a look at the current DL models. Here is the methodology that has been suggested in Section III. Part IV assesses how well it worked, and Section V wraps up the research.

2. Literature Survey

To automatically categorize wheezing, a wearable stethoscope with a piezoelectric MEMS resonant microphone array was created [17].The system employed a 12-layer temporal convolutional network (TCN) for feature extraction and classification, achieving a 98% accuracy in wheezing detection. However, it requires more processing time and cost.

In order to convert the audio data input into a spectrogram representation, a CNN with a Mixture-of-Experts (MoE) model was presented, which employed a front-end feature extraction strategy [18].This model can classify abnormalities in respiratory cycles and identify diseases from recordings of respiratory sounds. Spectrogram characteristics were subsequently classified into different respiratory anomaly cycle or disease groups using the back-end DL model. Nevertheless, complexity is still a concern.

It was suggested to use a Deep CNN (DCNN) approach to categorize aberrant respiratory sounds with the help of Artificial Noise Addition (ANA) [19].This method utilized Fourier evaluation for visual inspection of abnormal respiratory sounds, enhancing the actual spectrogram of sounds. Then, ANA was used to accurately differentiate between subclasses of abnormal/adventitious sounds by extracting features through feature maps. Finally, DCNN was utilized to identify and categorize various abnormal respiratory sound classes.

The accuracy of a CNN-based ResNet model for wheeze classification was improved by adding a Convolutional Block Attention Module (CBAM) [20]. This enhancement allowed the model to focus on relevant wheezing patterns in mel spectrograms. However, the imbalanced data in the training set still posed challenges despite the use of the F1 score to mitigate its effects.

In order to analyze respiratory sounds in real-time while dealing with continuous data streams, a fusion model was created that combines 1D CNN and LSTM networks [21].In its turn, this model resulted in the emergence of a counting algorithm of wheezing when monitoring breathing functions, both in clinical and non-clinical practice. There is however a lack of accuracy.

2.1 Research Gap

Although there has been progress in using DL in realizing wheezing, a number of gaps are still present. There is sparse research on the use of EMD with DL models to improve intrinsic feature extraction of lung sounds and there is underutilisation of HPSS on respiratory sound analysis. Also, the work done in literature tends to ignore the effect of EMD and HPSS on pre-processing of sounds of lungs, which impedes effective training datasets. Sealing these disconnects can result in more robust DL models to detect wheezing, which can eventually enhance respiratory health results.

3. Proposed Methodology

This section will provide an in-depth analysis of the WSFEDNN structure.

3.1 Wheeze Signal Feature Engineered Deep Neural Network (WSFEDNN) model

The WSFEDNN model proposed describes the different steps involved in detection and classification of lung sounds. Figure 1 depicts the phases of WSFEDNN model.

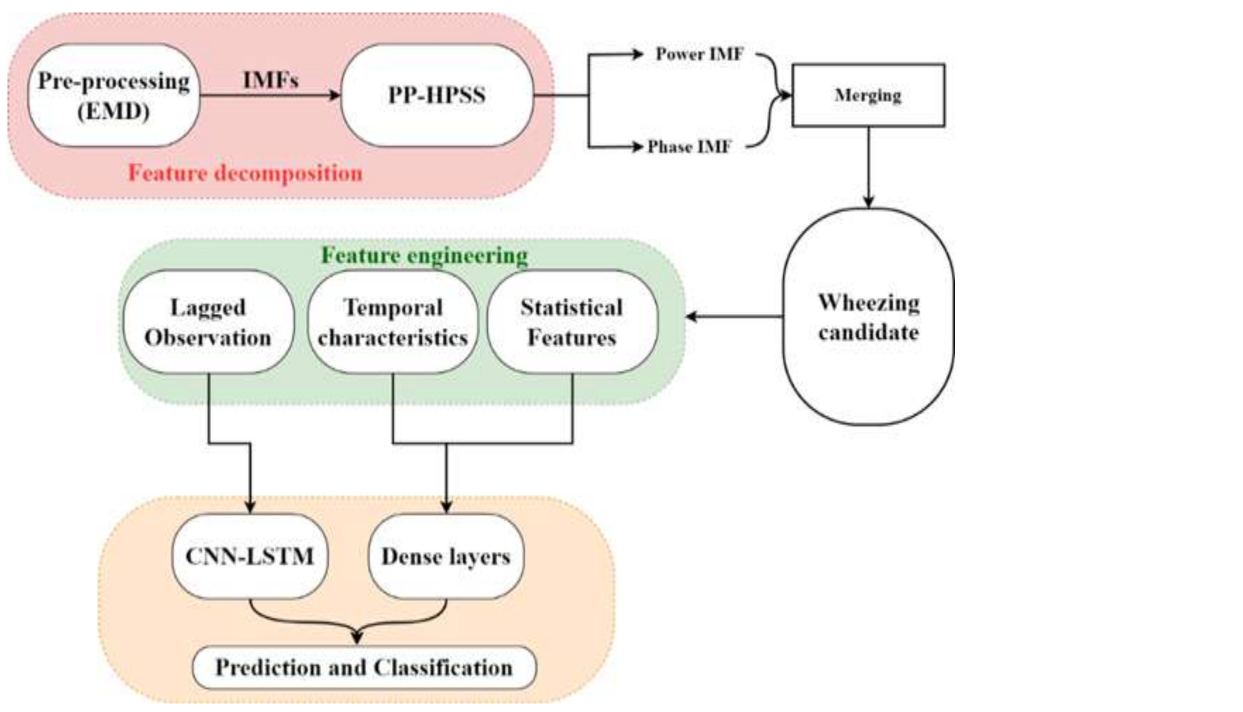


Fig. 1: WSFEDNN model

3.1.1 Feature Decomposition using EMD and PP-HPSS

The EMD, PP-HPSS are feature decomposition procedures in which EMD breaks down a signal into IMFs representing different oscillatory modes in non-linear and non-stationary signals, which are useful in extracting features. Cheaply separating mixed signals into harmonic (long-term, constant sounds) and percussive (little, brief bursts) signals at both amplitude and phase is then the goal of the PP-HPSS method. The resulting output has two kinds of IMFs, power IMF that depicts energy or intensity of the signal and phase IMF that carries phase details. Incorporating EMD with PP-HPSS, the method of decomposition improves the accuracy of decomposition and preserves the original properties of the signal in both parts.

Empirical Mode Decomposition

In Empirical Mode Decomposition (EMD) mechanism, the respiratory sound signals are broken down into a succession of n IMFs corresponding to various oscillatory modes inside the signal. Such a dynamic process enables EMD to retrieve non-linear and non-stationary characteristics of respiratory sounds that are essential to detect wheezing patterns. The high-frequency, periodic sounds caused by the constriction of airways normally occur in wheezing, and may be separated by studying the IMFs of frequency and amplitude parameters. EMD is able to detect the characteristic patterns of oscillation linked to wheezing by looking at the instantaneous frequency and amplitude of these IMFs.

Algorithm 1: Empirical Mode Decomposition

Input: Signal from the original audio source

Output: Group of n IMFs

1) Determine each and every local maximum and minimum in the signal $x(t)$.
2) Using interpolation from the local maxima, build the upper envelope $U(t)$.
3) Develop the lower bound $L(t)$ by extrapolating the vicinity of the minimums.
4) Find the average of the top and bottom envelopes:

$$M(t) = \frac{U(t) + L(t)}{2}$$

5) To get the first component, subtract the original signal from the mean envelope:

$$c_1(t) = x(t) - M(t)$$

6) Verify that $c_1(t)$ meets the requirements of the IMF, which state that the counts of extrema and zero crossovers need to be the same or differ by one.
7) If it doesn't, try again until the IMF requirement is satisfied.
8) After obtaining the initial IMF $c_1(t)$, deduct it from the initial signal:

$$r(t) = x(t) - c_1(t)$$

9) To extract other IMFs $c_2(t),\ c_3(t), \ldots, c_K(t)$, repeat the methods above on the residual $r(t)$.
10) The total of the acquired IMFs plus the final residual $r(t)$ is the original signal $x(t)$:

$$x(t) = \sum_{k=1}^{K} c_K(t) + r(t)$$

The produced IMFs are shown in Figure 2. This decompositionallows for the extraction of features relevant to both harmonic and percussive components, enabling further analysis through the PP-HPSS framework.

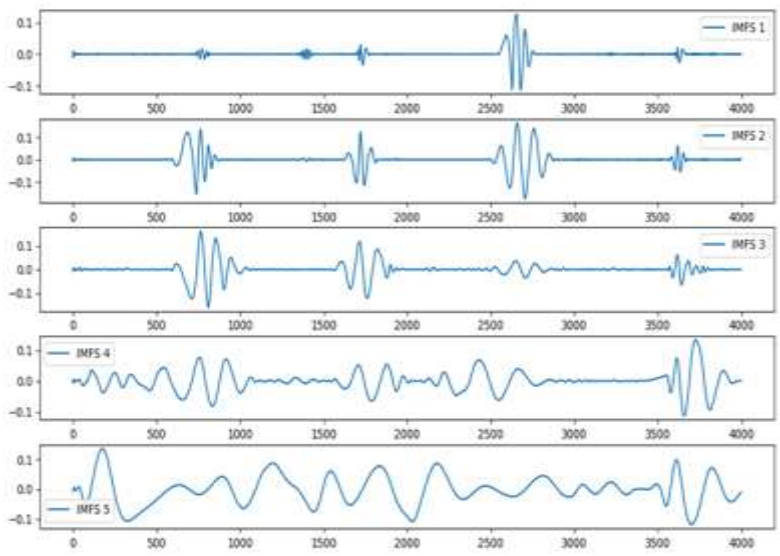


Fig. 2: Illustration of the Empirical Mode Decomposition Process

Power and Phase Aware Harmonic-Percussive Source Separation (PP-HPS)

IMFs obtained with the help of EMD can be refined with the help of PP-HPSS which is capable of isolating the signal to harmonic and percussive parts. When PP-HPSS is applied to IMFs based on respiratory sounds, it generates two different types of components: power-dominant IMFs, and phase-dominant IMFs. Power-dominant IMFs are mainly associated with capturing the harmonic parts of the IMF but phase dominant IMFs capture the percussion parts of the respiratory sound signal.The working principle of PP-HPSS is to separate a complex valued spectrogram $S(f, t)$of the respiratory signal into two parts, the Power IMF (harmonic component) $S_h(f,t)$ and Phase IMF (percussive component) $S_p(f,t)$. Mathematically this can be expressed to show its decomposition is:

$S\,(f,t) = H(f,t) + P(f,t)(1)$

- **Anisotropic Smoothness:** Smoothness is a crucial aspect in PP-HPSS which helps to preserve integrity of the harmonic and percussive elements. The Power IMF building blocks can be anticipated to be time- and frequency-continuous, with the form:

$H(f,t) \approx H\,(f,t \pm \Delta t)(2)$

Conversely, the Phase IMF elements are smooth mostly in the frequency direction as stated:

$P(f,t) \approx P(f \pm \Delta f,t)$ (3)

To ensure preferred smoothness properties, the PP-HPSS method poses the optimization problem as:

$\min_{H,P}(\frac{1}{2\epsilon_h^2}\|D_t(H)\|_{Fro}^2 + \frac{1}{2\epsilon_p^2}\left\|D_f(P)\right\|_{Fro}^2)$ (4)

where, $D_t$and D_f are the discrete directional derivatives in time and frequency, $\epsilon_h$and ϵ_p are regularization parameters that control the smoothness of the Power IMF and Phase IMF components, respectively. The given formulation guarantees that the natural properties of the original respiratory signal are maintained.

- **Sinusoidal Model:** Sinuoidal model is an additional signal that adds an enhancement on the Power IMF representation by including the amplitude and phase of signal. Short-Time Fourier Transform (STFT) of signal $x \in \mathbb{R}^L$ is defined as:

$F(x)_{f,t} = \sum_{l=0}^{L-1} x_l a_t g_l e^{-2\pi j\frac{b_t}{L}}$ (5)

where, g is the window function, a and b denote time and frequency shifting steps. At the same time, a sinusoidal component can be represented as:

$s_l = Ae^{2\pi j\frac{b_t}{L}+\varphi}$ (6)

A the amplitude and φ the initial phase. The phase spectrogram φ varies as:

$\varphi = \varphi_{f,t\pm\Delta t} + 2\pi v_{f,t\pm\Delta t}$ (7)

$v_{f,t}$ is the instantaneous frequency in each time-frequency bin. The optimization problem of the power and phase IMF separation is:

$\min_{H,P}\left(\sum_{t=1}^{T}\left\|F(H)_{f,t}A_{f,t}e^{j\varphi_{f,t}}\right\|_2^2 + \lambda_H\|D(t)_H\|_{Fro}^2 + \quad \lambda_P\|D(f)_P\|_{Fro}^2\right)$ (8)

where, λ_HH and λ_P are regularization factors of harmonic and percussional elements, D_t and D_f are directional difference to achieve smoothness.

In order to stabilize the components of Power IMF and remove the sudden change in phase, an instantaneous phase modification matrix E is applied to the complex-valued spectrogram:

$F_{iPM}(x) = E\ \odot F(x)(9)$

where,

$E_{f,t} = \prod_{l=0}^{t\pm\Delta t} e^{2\pi j v_{f,}\frac{\eta a}{L}},\ E_{f,0} = 1(10)$

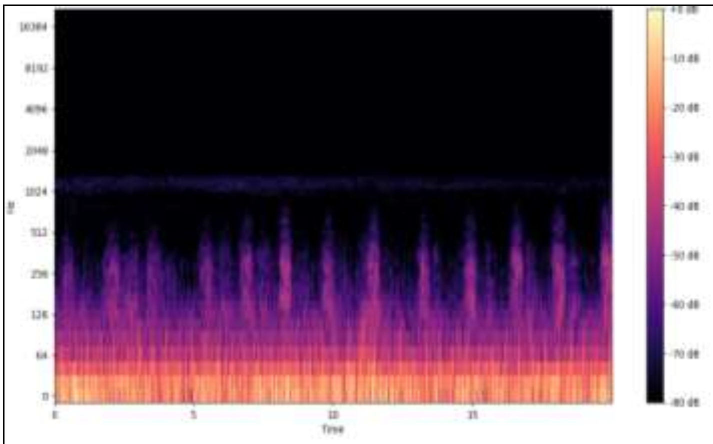
The change won't create hiccups in the harmonic Power IMF, enhancing the quality of the reconstruction.

Sparsity in Percussive Components

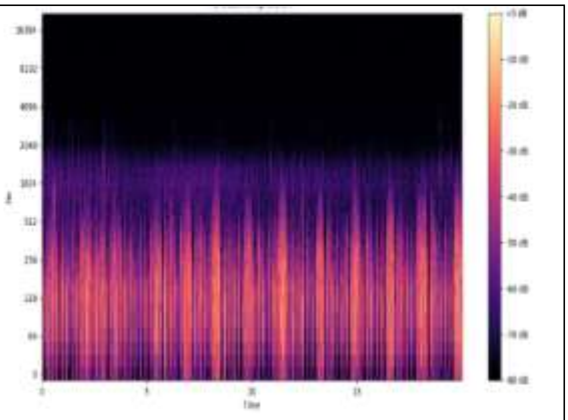
In case of the Phase IMF components, sparsity in time-frames should also be promoted, and this is achieved by using the $l_{2,1}$-norm as follows:

$\|X\|_{2,1} = \sum_{t=1}^{T}\left(\sum_{f=1}^{K}\left|X_{f,t}\right|^2\right)^{\frac{1}{2}}(11)$

This is a successful formula that isolates the percussion (noise in the background or abnormal breathing sounds) and the Power IMF wheezing elements, so that only the periodic, tonal wheezing calls can be audible.At last the power and phase IMFs are obtained which is illustrated in the Figure 3.



(a)



(b)

Fig. 3: The output of a) Power IMF (harmonic component) b) Phase IMF (percussive component)

The combination of EMD and PP-HPSS forms a powerful platform of the detection of wheezing in the lung sounds. The EMD step breaks respiratory signal into IMFs that represent important oscillatory patterns, which is vital in determining wheezing episodes. This process can be enhanced by the second use of PP-HPSS which correctly distinguishes the components of the Power IMF and Phase IMF along with their amplitude and phase. This segregation boosts the strength of recognition of wheezing particularly when background noise is present or other irregular sounds are present. Components of the Power IMF and Phase IMF, respectively tend to emphasize periodic patterns in wheezing and remove noise and other non-wheezing sounds of the respiratory process, respectively.

3.1.2 Feature Engineering

After using the PP-HPSS technique, the second important step associated with our methodology is feature engineering. It transforms processed audio data to meaningful features that enhance model performance. This is done by removing the input variables, such as lagged observations, temporal features and statistical features based on signal decomposition methods.

Lagged Observations:

Lagged observations encode the history of lung sounds by using previous audio observations, enabling the model to identify temporal features of wheezing. In particular, lagged observations contain lag of values, like lag_1, which is the current value at time t; lag_2, the value in time $t-1$; lag_3, the value in time $t-2$and so forth extending until lag_{10}which is the value in time $t-9$.. These values of history offer a more profound context to the CNNLSTM model, where CNN module picks out important features and the LSTM module translates sequential dependencies, to identify long-term trends. Autocorrelation analysis picks the most useful lagged observations, with this making sure that the model pays attention to lung sounds that are most likely to indicate wheezing. This CNN-LSTM approach is more effective since it uses both previous and present data to detect wheezing.

Temporal Features:

The timing of wheezing events with its cyclical characteristics can be determined by temporal features obtained on lung sound recordings which can be critical to the identification of respiratory disorders. These point out hour (actual time of recording), minute, day of the week and whether the episode is on a weekend or during a holiday. Other changes like the cosine of the minute and the sine of the second are used to represent time-based cycles. These attributes capture differences in the frequency and length of wheezing behavior at different times, and assist the model in identifying complicated behaviors in wheezing. After extraction, three dense layers are applied to these temporal features to reveal the complex relationships, improving predictions of the respiratory health outcomes.

Statistical Features:

Lung sound statistical analysis offers useful clues to the structure and behavior of lung sounds, which are essential in distinguishing between normal and wheezing lung sounds. Some of the most important features (f) are the average (f_{mean}), standard deviation (f_{std}), minimum (f_{min}) and maximum (f_{max}) values that provide the general image of the sound intensity and variability. Rolling is taken across intervals (i.e. $rolling_{mean_{10}}$, ... $rolling_{mean_{40}}$, capture the trend of statistics across different time periods, whilst expanding statistics like $expanding_{mean}$, $expanding_{min}$ and $expanding_{max}$ indicate cumulative change in properties of the sounds. Other characteristics like skew $(f_{skewness})$ and kurtosis $(f_{kurtosis})$ are useful to determine the distributions and severity of wheezing. Such features as $f_{peakcount}$reveal the frequency of wheeze peaks and $f_{diff}$is a measure of the range of the sound. All these features are then processed in three layers of high density, allowing the model to learn complex relationships of the data. This combination with signal decomposition methods provides the ability to capture complex temporal patterns, improving its accuracy and reliability in wheezing detection.

3.1.3 Wheezing detection

Following the phase of feature engineering, which results into the generation of deep features out of the input features, the softmax classifier is invoked to cope with the classification task. In training, CNN learns spatial characteristics and the LSTM learns time-related characteristics of lung sounds. The resulting features are then sent to the softmax layer which transforms them into a probability distribution over the potential classes (e.g., "wheezing" or "normal").

The softmax function is used to make sure that the total of all of the predicted class probabilities is equal to 1 and the highest probability class is chosen to be the predicted label. In the training stage, the model is trained by minimizing the cross-entropy loss, which is the comparison of the probabilities generated by the model with the true class labels. Backpropagation causes the parameters of the model to be changed, thus enhancing the separation of classes.

The trained CNN-LSTM model is used during classification allowing the determined class of new lung sound data to be predicted based on the softmax output with the highest probability. This composite feature derivation and categorization arrangement enhances the capability of the model to discern wheezing habits well.

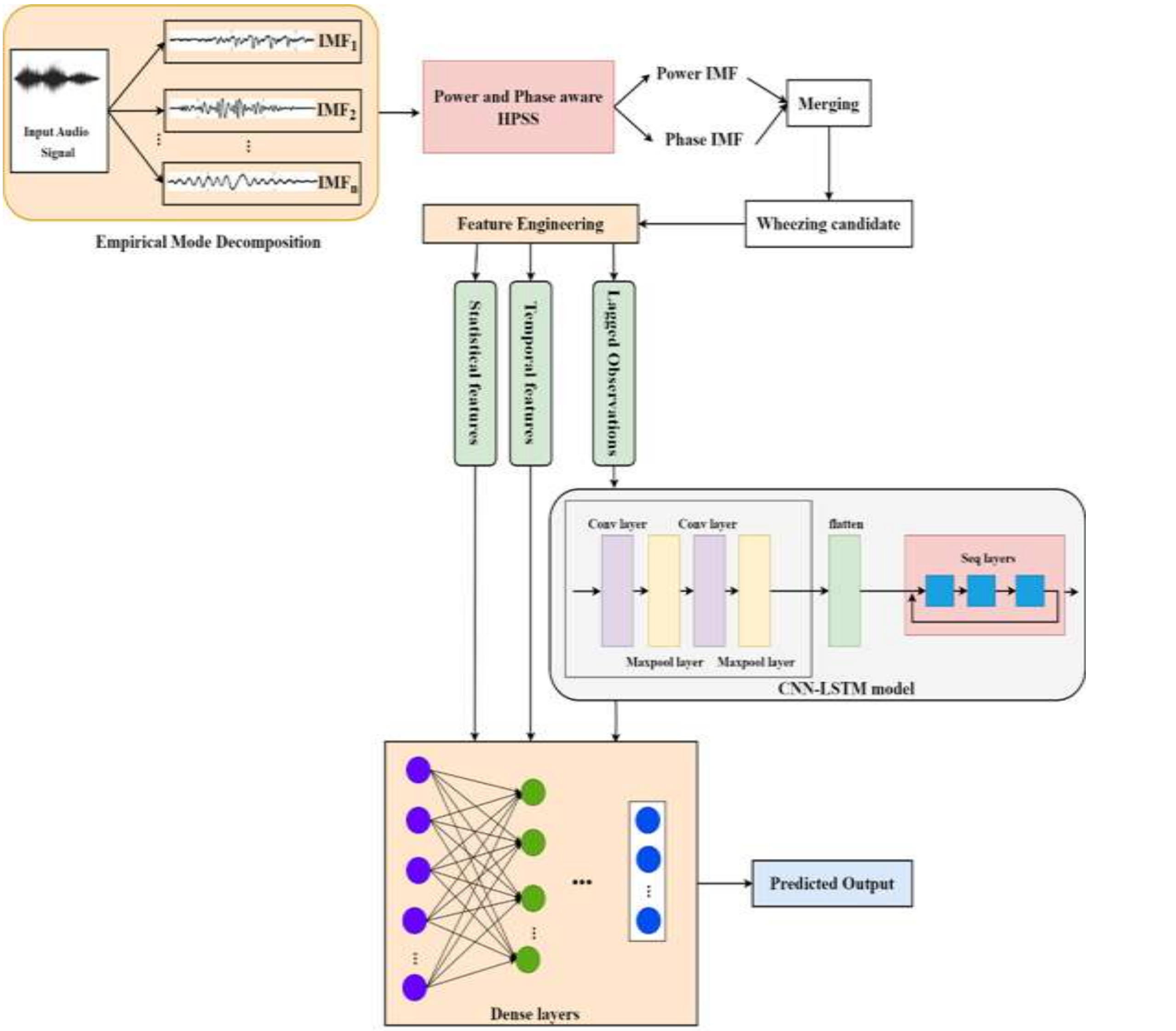


Fig. 4: Network structure of the proposed WSFEDNN model

Table 1: Optimal Hyperparameter Settings for the CNN–LSTM Model in the Proposed WSFEDN

Parameters	Range
Learning rate scheduling	0.001
Count of epochs	100
Size of batch	16
Dropout	0.5
Optimization	Adam
Kernel Size	3X3
Input size	40 × 862 × 1
Number of Dense Layers	3
Loss Function	Cross-Entropy
Output Activation	Softmax

Here are the hyperparameter values that were used for the CNN-LSTM architecture of the WSFEDNN model that was proposed: Table 1. These parameters were found by empirical means to improve the model's accuracy, generalizability, and efficient wheezing classification performance.

4. Performance Evaluation

4.1 Dataset description

Some publicly respiratory sound datasets include:

The ICBHI 2017 Respiratory Sound Database is a freely accessible resource that includes 920 audio files. These files contain 5.5 hours of recordings obtained from 126 participants, including 79 males, 46 females, and 1 unknown. The participants' ages range from 76 adults to 49 children, and 1 unknown. Two separate study groups in Greece and Portugal gathered audio samples over the course of several years and entered them into the database. Gathered in both professional and private situations, the recordings feature a variety of background noises and stethoscopes (WelchAllynMeditron, 3MLittmann 3200, 3MLittmann Classic II SE) or microphones (AKG C417 L) with varying sampling frequencies. The database includes 341 audio files with 1898 annotated wheezes. By age, sex, body mass index (BMI), diagnosis, and recording equipment, the distribution of total files, files with wheeze annotations, and annotated wheezes is revealed. All individuals under the age of 18 were deemed to be children for the purposes of age categorization. Based on the World Health Organization's criteria, body mass index (BMI) categories were established. Since there were only three underweight individuals, a category that combined normal and underweight was created. Patients with lower respiratory tract infections (LRTIs), upper respiratory tract infections (URTIs), bronchiolitis, or pneumonia were classified as non-chronic, while patients with chronic obstructive pulmonary disease (COPD), asthma, or bronchiectasis were classified as chronic. For this study, three classes were kept in the ICBHI 2017 dataset, including Normal (35 samples), Crackle (859 samples) and Wheeze (26 samples). COPD, pneumonia, bronchiectasis, and bronchiolitis cases are come under Crackle whereas URTI, LRTI, and asthma-related samples are come under Wheeze.

RALE dataset presents digital recordings of respiratory sounds in health and disease. The 26 participants included in the RALE dataset had lung sound recordings taken at various ages, from newborns to 76-year-olds. These recordings were made at a sampling rate of 10.240 KHz and serve as positive class examples for the wheezing model. The negative class examples utilize audio data collected from 21 people using a Samsung Gear S3 wearable at a sampling rate of 16 KHz. An internet repository is utilized to acquire throat clearing and supplementary cough sounds at a sampling rate of 44.1 KHz. Various audio samples of varying sampling rates are acquired.For this study, wheeze (26 samples) and non-wheeze (23 samples) classes are considered form this dataset.

The sample distribution of each class in both datasets is imbalancedthat can interfere with the ability of the model to acquire discriminative featuresand could result in prediction bias at the expense of overall performance in generalization. In order to handle this issue, the segmentation based on windowing has been considered on each signal. For ICBHI 2017, a 1-second window was used to segment each recording. Similarly, two window configurations of 0.5 seconds and 1 second were used to segment the recordings of RALE dataset. The segmentation of recordings significantly increased the number of samples by dividing lengthy recordings into fixed-length segments while keeping their original labels

After segmentation, preprocessing operations such as normalization, conversion of spectrogram and smart padding to ensure that the input representation is constant across signals of different lengths.In order to facilitate model building and assessment, the increased samples after segmentation were consistently divided into 60% for training and 40% for testing across both datasets. Tables 2 and 3, respectively, provide the ICBHI 2017 and RALE datasets' complete data distribution.

Table 2: ICBHI data distribution

Class	Original Data	After Segmentation (1s) and resampling	Training data (60%)	Testing data (40%)
Normal	35	520	312	208
Crackle	859	580	348	232
Wheeze	26	510	306	204
Total	**920**	**1610**	**966**	**644**

Table 3: RALE data distribution							
Class	**Original Data**	**After Segmentation (0.5s)**	**Training data**	**Testing data**	**After Segmentation (1s)**	**Training data (60%)**	**Testing data (40%)**
wheeze	26	478	287	191	239	143	96
Non-wheeze	23	422	253	169	211	127	84
Total	**49**	**900**	**540**	**360**	**450**	**270**	**180**

4.2 Performance metrics

On Windows 10 PCs running version 21H2, with a 64-bit operating system, an Intel(R) Xeon(R) 4.01 GHz CPU, and 64 GB of RAM, the efficacy of the WSFEDNN approach is assessed and evaluated using the existing DL algorithms. MATLAB 2019 B is used as the programming language. The PP-HPSS model is compared against HPSS [22], NMF-HPSS [23] and DNN-HPSS [24] highlighting the effectiveness of source separation. The WSFEDNN model is also compared to show its effectiveness in classification with the models such as CBAM [20] and RNN-LSTM [25].

The concepts to be known for predicting the results are as follows:

- Annotated Event (AE): The annotator decides on the time parameters that designate the beginning and finish of a session.
- Segmented Event (SE): Time intervals when a potential wheeze starts and stops, as determined by the segmentation method
- Detected Event (DE): AE that the algorithm identifies.
- Undetected Event (UE): AE for which the algorithm is inadequate.
- False Event (FE): SE that do not contain an AE at either its beginning or its end

The performance metrics used to evaluate the proposed and existing algorithms using the above concepts are described below:

Metrics for PP-HPSS

- **Signal-to-Distortion Ratio (SDR):** SDR is used to evaluate the quality of audio separation or enhancement tasks such as in lung sound analysis for wheezing detection. The formula for SDR is typically expressed as:

$$SDR = 10.log_{10}(\frac{\|s_{target}\|^2}{\|e_{noise}\|^2})(12)$$

where,

- s_{target} is the original signal.
- e_{noise} represents the distortion or noise present in the separated or predicted signal.
- $\|\cdot\|$ denotes the energy (squared magnitude) of the signal.

- A higher SDR value indicates a cleaner separation, meaning the predicted signal closely resembles the original lung sound with minimal distortion When applied to wheezing detection, a high SDR indicates how well the model focuses on and isolates the vital information about wheezing and maintains the originality of the wheezing features, but rejects the artifacts or noise in the audio signal.

- **Jaccard Index (JI):** The Jaccard Index adopted to measure the similarity between two sets at segmentation tasks to assess accuracy of a DE or SE as compared with the annotated ground truth.For an event detection task like wheeze segmentation, the Jaccard Index can be calculated as follows:

$$Jaccard\ Index\ (JI) = \frac{Intersection\ (overlapping\ region\ of\ SE\ and\ AE)}{Union\ (total\ combined\ region\ of\ SE\ and\ AE)} \qquad (13)$$

The result is a value between 0 and 1 with 1 being the perfect overlap (the segmented event and the annotated one coincide in their boundary) and 0 the total non-overlap (the events have no similarity at all). An increase in the Jaccard Index values is an indication that there is a stronger match between the identified event and the actual one indicating strong segmentation whereas a decrease is an indication that there is some wrong match which might translate into segmentation algorithm errors.

- **Overlap Coefficient (OC):**The SorensenDice coefficient (also called the Overlap Coefficient) is another method of assessing the similarity of two sets, giving emphasis to the degree of overlap between sets. The formula for the OC is:

$$Overlapping\ Coefficient\ (OC) = \frac{Size\ of\ the\ Intersection\ between\ SE\ and\ AE}{Size\ of\ the\ smaller\ set\ (either\ SE\ or\ AE)} \quad (14)$$

The coefficient will be between 0 and 1 with a value of 1 being a perfect overlap, i.e. the Segmented Event is completely covered by the Annotated Event and 0 means that they do not overlap at all between the two sets. The greater the value, (the closer it is to 1), the greater the overlap between the annotated events and the segmented events, implying the segmentation is very accurate. The smaller the values, the worse the overlap and this means that the segmentation or event detecting algorithm is not that good.

Metrics for WSFEDNN

- **Precision:** The accuracy of the model's wheeze detection is defined as the percentage of true positives relative to the total number of cases that the model identified as positive.

$$Precision = \frac{DE}{DE+UE} (15)$$

- **Recall:** The recall is a measure of how well the model detected positive wheezing occurrences relative to all the actual wheezing events. The model's ability to capture all wheeze instances is demonstrated by this.

$$Recall = \frac{DE}{DE+FE} (16)$$

- **F1-score:** When Precision and Recall are harmonically averaged, the result is the F1-Score. If it detects an imbalance between recall and precision, it offers an individual metric that balances both.

$$F1 - score = \frac{2 \times Precison \times Recall}{Precision+Recall} \quad (17)$$

4.3 Performance evaluations of PP-HPSS on ICBHI 2017 and RALE Datasets

The comparison of the performance of different methods of HPSS is presented in Figure 5. It is dedicated to the SDR values of the ICBHI 2017 and RALE datasets.The PP-HPSS method proposed shows a high improvement in SDR in comparison to the rest of the approaches.In the case of ICBHI 2017 dataset, the SDR of PP-HPSS is 15.0 dB, which is 2.5 dB better than that of HPSS, NMF-HPSS, and DNN-HPSS. On the same note, PP-HPSS has an SDR of 14.0 dB, which is larger than that of HPSS by 3.0 dB, NMF-HPSS by 1.7 dB, and DNN-HPSS by 1.3 dB.These findings point to how the PP-HPSS is always more effective in comparison with the current methods, and can be used to demonstrate its usefulness in enhancing the quality of wheezing detection.

Figure 6 compares the performance of different HPSS techniques, concentrating on values of Jaccard Index in the ICBHI 2017 and RALE datasets.The proposed PP-HPSS techniquedemonstrates a significant increase inJaccard Index as compared to other methods. The Jaccard Index of PP-HPSS is 0.61 in the ICBHI 2017 dataset, which is higher than that of HPSS at 0.14, NMF-HPSS at 0.13 and DNN-HPSS at 0.07.Likewise, PP-HPSS has a Jaccard Index that is higher by 0.58 compared with HPSS (by 0.13), NMF-HPSS (by 0.08), andDNN-HPSS (by 0.03) in the RALE dataset.

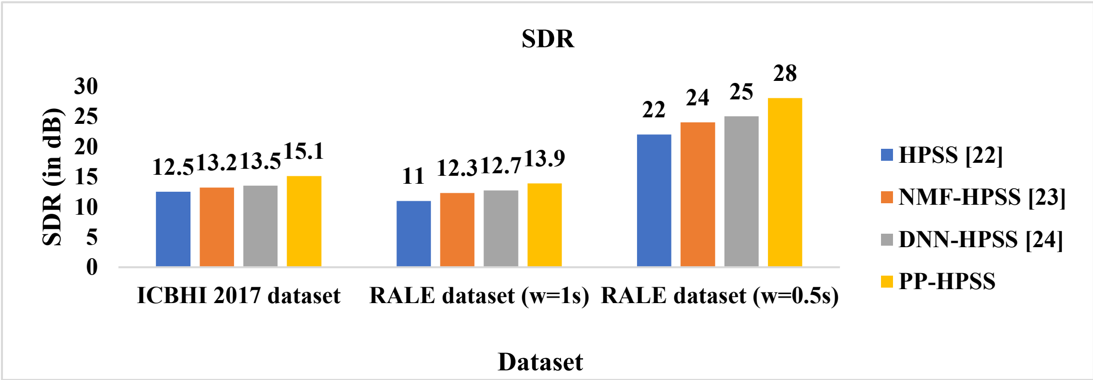


Fig. 5: Performance Comparison of various HPSS models in terms of SDR across ICBHI 2017 and RALE datasets

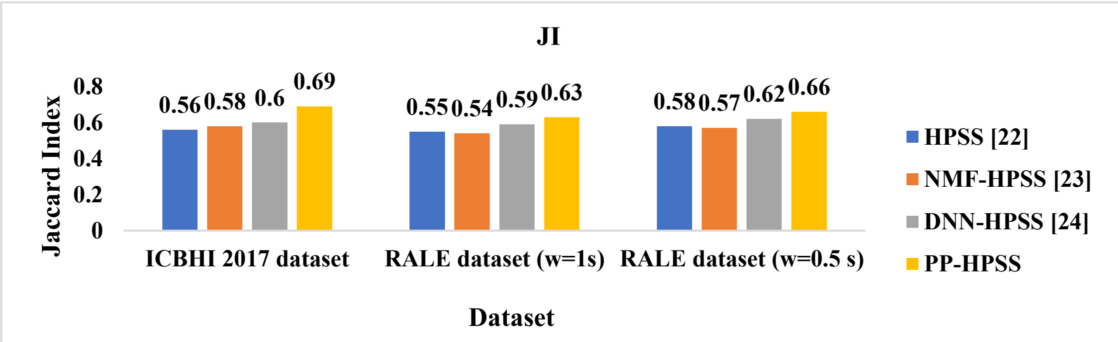


Fig. 6: Performance Comparison of various HPSS models in terms of JI across ICBHI 2017 and RALE dataset

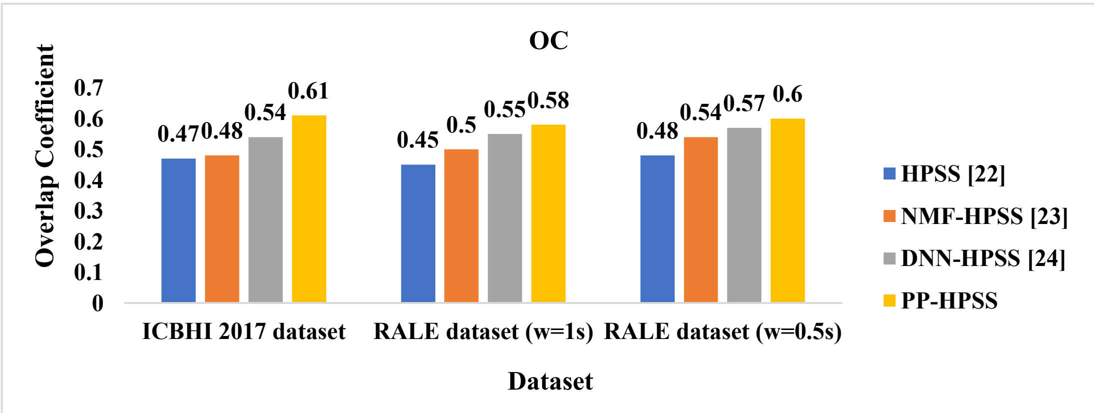


Fig. 7: Performance Comparison of various HPSS models in terms of OC across ICBHI 2017 and RALE dataset

These findings highlight the overall high quality of the PP-HPSS approach, emphasizing its efficiency in improving the quality of wheezing detection. Figure 7 compares the performance of the different methods that an HPSS can be implementedto by emphasizing the Overlay Coefficientvalues as applied to the ICBHI 2017and

RALE datasets. It can be noted that the Overlay Coefficient of the proposed PP-HPSS approach is significant in comparison with the others. With an Overlay Coefficient of 0.61 in the ICBHI 2017 dataset, PP-HPSS beats HPSS (0.14), NMF-HPSS (0.13) and DNN-HPSS (0.07). Likewise, in the RALE dataset, PP-HPSS has an Overlay Coefficient of 0.58, which is 0.13 higher than HPSS, NMF-HPSS, and DNN-HPSS. These findings highlight thesteady high-quality of PP-HPSS technique, and highlight the success of the technique in improving wheezing detection.

4.4 Performance evaluations of WSFEDNN on ICBHI 2017 and RALE Dataset

Figure 5 present the confusion matrices obtained from the testing sets, where the ICBHI 2017 dataset was segmented using a 1s window, while the RALE dataset wasevaluated under two window configurations of 0.5s and 1s in

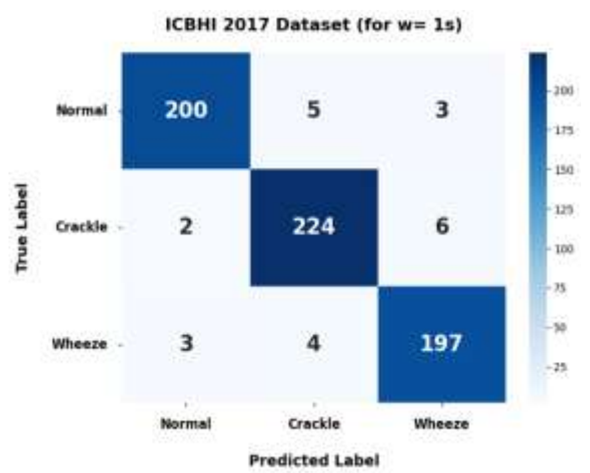


Fig. 8 Confusion matrix for ICBHI dataset for w=1s

Figure 6 and Figure 7.

4.4.1 Training and Testing Accuracy curve

Figure 11 and Figure 12 represent the training and testing accuracy curves of the ICBHI 2017 and RALE datasets.

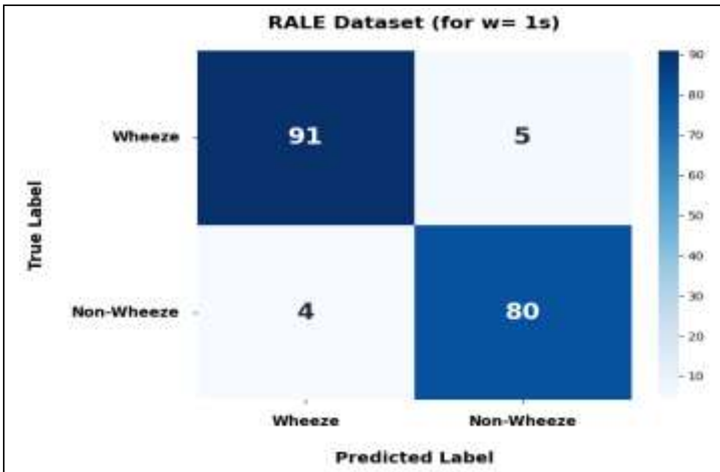


Fig. 9: Confusion matrix of RALE dataset for w= 1s

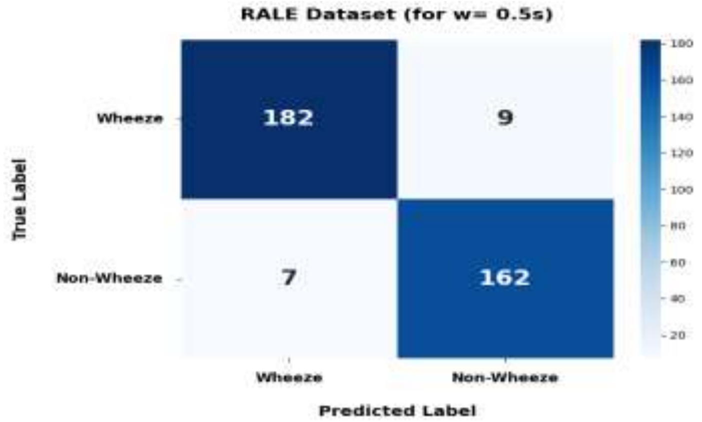


Fig. 10: Confusion matrix of RALE dataset for w= 0.5s

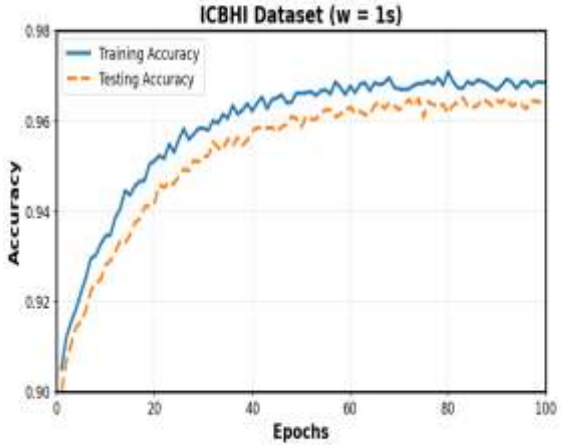


Fig. 11: Training and Testing Accuracy curve on ICBHI dataset (w=1)

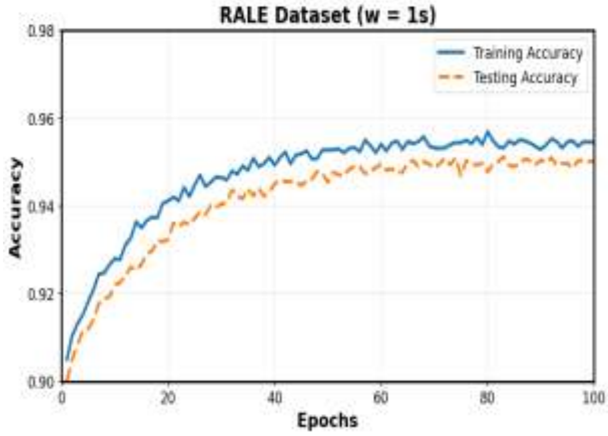


Fig. 12: Training and Testing Accuracy curve on RALE dataset (w=1s)

Table 4: Evaluation Measures based on Classes of RALE Dataset (w = 0.5s and 1s)

RALE Dataset (w=0.5s)				RALE Dataset (w=1s)				
Classes	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Wheeze	95.39	95.29	96.30	95.79	94.69	94.79	95.79	95.29
Non-Wheeze	95.76	95.86	94.74	95.29	95.34	95.24	94.12	94.67

Table 5. Evaluation Measures based on Classes of ICBHI Dataset (w = 1s)

ICBHI dataset (for w= 1s)				
Classes	Accuracy	Precision	Recall	F1-Score
Normal	96.18	96.15	97.56	96.85
Crackle	96.52	96.55	96.14	96.34
Wheeze	96.58	96.57	95.63	96.10

4.4.2 Class wise Performance analysis

The suggested model's accuracy, precision, recall, and F1-score were analyzed on the RALE and ICBHI data with window sizes of 0.5s and 1s, respectively, in Tables 4 and 5.

4.4.3 Performance Comparison of WSFEDNN with existing models

Figure 13 shows that the proposed model of the WSFEDNN compares its performance with two other algorithms, CBAM and RNN-LSTM with the ICBHI 2017 dataset. The WSFEDNN model has shown great improvements in the most important measures like Precision, Recall, and F1-Score. In particular, WSFEDNN is 10.28% more precise and 11.60% more recalls than CBAM and is 6.36% more precisely and 7.13% more recalls than RNN-LSTM. With regards to F1-Score, WSFEDNN is expected to have 10.84 percent higher acceptance rate than CBAM and 8.13 percent higher acceptance rate than RNN-LSTM.These findings demonstrate the use of the presented WSFEDNN model as the one successfully applied every time in terms of the quality of wheezing detection in comparison with the rest of the methods.Figure 14 draws a comparison between the performance of the proposed WSFEDNN model and two other models, HPSS [25] and RNN-LSTM [26], on RALE dataset. WSFEDNN model shows significant improvements in significant measures such as Precision, Recall, and F1-Score. In particular, at WSFEDNN a 10.05% improvement in Precision compared to CBAM, and a 6.77% improved compared to RNN-LSTM improvement. WSFEDNN is 9.29 and 6.02 percent better than CBAM and RNN-LSTM in Recall, respectively. Also, F1-Score increases by 9.23% with WSFEDNN as compared to CBAM and by 6.19% when compared to RNN-LSTM [26]. These findings underscore the performance of WSFEDNN by making it the best option in the detection of wheezing as compared to the other two.

Figure 15 presents the performance of the proposed WSFEDNN model against that of two other approaches: HPSS [25] and RNN-LSTM [26] on RALE dataset. The WSFEDNN model achieves significant advancements in crucial metrics, such as Precision, Recall, and F1-Score. In particular, WSFEDNN demonstrates an absolute increase in Precision of 10.57% with respect to CBAM and 7.32% with respect to RNN-LSTM. WSFEDNN beats CBAM and RNN-LSTM by 9.83% and 6.57 percent respectively in Recall. Also, WSFEDNN demonstrates the growth of F1-Score by 9.76% in comparison to CBAM and 6.70% in comparison to RNN-LSTM [26]. These findings reveal that WSFEDNN is quite effective, thus making it a better tool to detect wheezing as compared to the other two techniques.

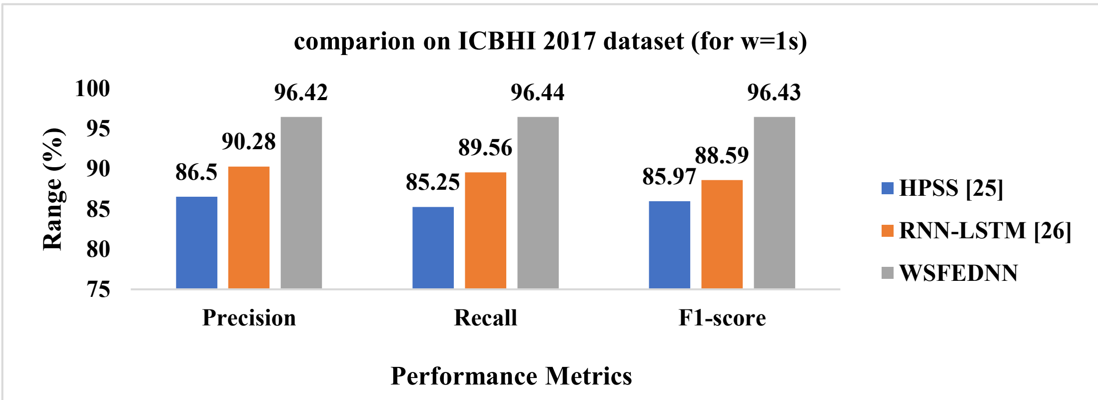


Fig. 13: Comparison of WSFEDNN with existing models on ICBHI 2017 dataset (for w=1s)

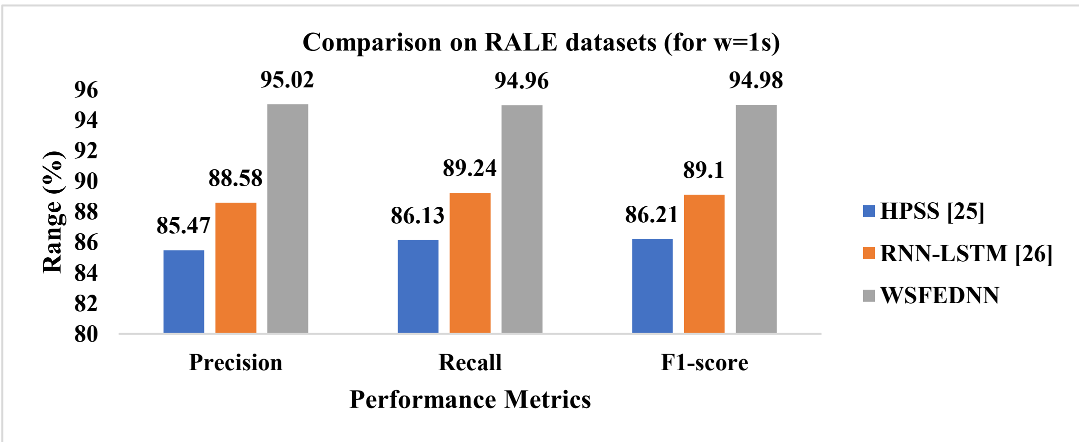


Fig. 14: Comparison of WSFEDNN with existing models on RALE datasets (for w=0.5s)

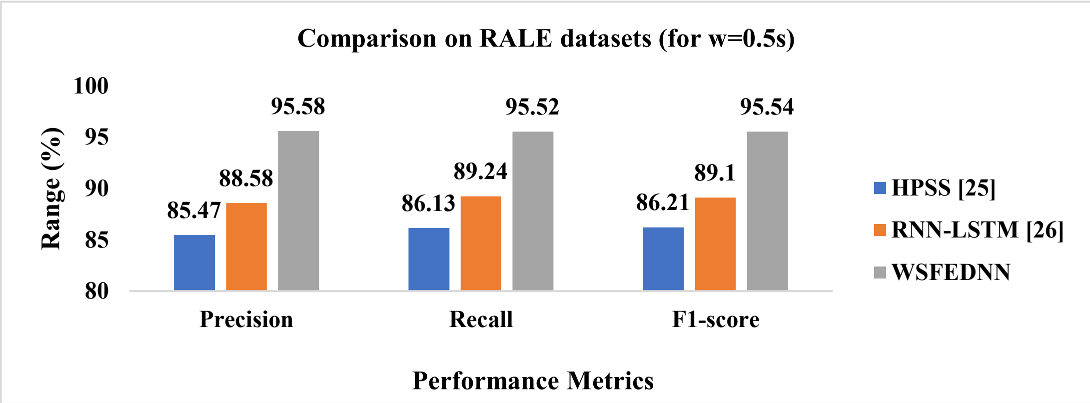


Fig. 15: Performance Comparison of various HPSS models in RALE datasets (for w=0.5s)

Conclusion

The suggested WSFEDNN type is much more efficient in detecting and classifying the wheezing sound instances in the lung sounds. The effective implementation of the WSFEDNN by combining PP-HPSS and EMD allows decomposing lung sounds without losing the necessary amplitude and phase information of sounds. This advancement underscores the model capacity to clearly differentiate wheezing sounds among being exposed to background noise. Moreover, feature engineering methods used such as lagged observations and temporal features, as well as the CNN-LSTM architecture, allowed the model to be able to detect both spatial and temporal patterns. The strong performance of the WSFEDNN highlights the fact that it can be used in the real-world clinical scenarios in order to detect respiratory conditions early and control them, which consequently leads to overall patient health improvements. Experimental work proved that the WSFEDNN was able to perform better than other existing baseline models such as HPSS and RNN-LSTM with a precision of 93.48 and 91.23 respectively. Future studies could aim at maximally refining and investigating its use in various patient groups.

References

1. J. Acharya, A. Basu and W. Ser, "Feature extraction techniques for low-power ambulatory wheeze detection wearables", *In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)IEEE*, pp. 4574-4577, (2017, July). [CrossRef] [Google Scholar] [Publisher Link]
2. C.Habukawa, N. Ohgami, N. Matsumoto, K. Hashino, K. Asai, T. Sato and K. Murakami, "A wheeze recognition algorithm for practical implementation in children", *PloS one*, vol. 15, no. 6, e0240048, 2020. [CrossRef] [Google Scholar] [Publisher Link]
3. N. Boucher, A. Prystupa, A. Witczak, E. Walczak, G. Dzida and L. Panasiuk, Lung auscultation–Identification of common lung sound abnormalities and associated pathologies, *Journal of Pre-Clinical and Clinical Research*, vol. 7, no.1, 2013.[Google Scholar]
4. L.D. Mang, F.J. Cañadas-Quesada, J.J. Carabias-Orti, E.F. Combarro and J. Ranilla, Cochleogram-based adventitious sounds classification using convolutional neural networks, *Biomedical Signal Processing and Control*, vol. 82, 104555, 2023. [CrossRef] [Google Scholar] [Publisher Link]
5. G.D. Sosa, A. Cruz-Roa, F.A. González, "Automatic detection of wheezes by evaluation of multiple acoustic feature extraction methods and C-weighted SVM", *In 10th International Symposium on Medical Information Processing and AnalysisSPIE*, Vol. 9287, pp. 32-37, (2015, January). [CrossRef] [Google Scholar] [Publisher Link]
6. Xu, K.G. Tantisira, A. Wu, A.A. Litonjua, J.H. Chu, B.E. Himes and S.T. Weiss, "Genome Wide Association Study to predict severe asthma exacerbations in children using random forests classifiers", *BMC medical genetics*, vol. 12, pp. 1-8, 2011.[CrossRef] [Google Scholar] [Publisher Link]
7. Z. Neili, M. Fezari and A. Redjati, "ELM and K-nn machine learning in classification of Breath sounds signals", *Int J ElectrComput Eng*, vol. 10, no. 4, pp. 3528-3536, 2020. [CrossRef] [Google Scholar] [Publisher Link]
8. D. Spathis and P. Vlamos, "Diagnosing asthma and chronic obstructive pulmonary disease with machine

learning", *Health informatics journal*, vol. 25, no. 3, pp. 811-827, 2019. [CrossRef] [Google Scholar] [Publisher Link]
9. A.B. Mahammad and R. Kumar, "Machine Learning Approach to Predict Asthma Prevalence with Decision Trees", *In 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) IEEE*, pp. 263-267, (2022, October).
10. [CrossRef] [Google Scholar] [Publisher Link]
11. G. Serbes, "Detection of Wheeze Sounds in Respiratory Disorders: A Deep Learning Approach", *International Advanced Researches and Engineering Journal*, vol. 8, no. 1, pp. 20-32, 2024.[CrossRef] [Google Scholar] [Publisher Link]
12. R.X.A. Pramono, S.A. Imtiaz and E. Rodriguez-Villegas, "Evaluation of features for classification of wheezes and normal respiratory sounds", *PloS one*, vol. 14, no. 3, e0213659, 2019. [CrossRef] [Google Scholar] [Publisher Link]
13. H. Chen, X. Yuan, J. Li, Z. Pei and X. Zheng, "Automatic multi-level in-exhale segmentation and enhanced generalized S-transform for wheezing detection", *Computer methods and programs in biomedicine*, vol. 178, pp. 163-173, 2019. [CrossRef] [Google Scholar] [Publisher Link]
14. P.S. Faustino, "Crackle and wheeze detection in lung sound signals using convolutional neural networks", 2019.[Google Scholar]
15. S. Jayalakshmy and G.F. Sudha, "Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks", *Artificial intelligence in medicine*, vol. 103, p. 101809, 2020. [CrossRef] [Google Scholar] [Publisher Link]
16. Y. Kim, Y. Hyon, S.S. Jung, S. Lee, G. Yoo, C. Chung and T. Ha, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning", Scientific reports, vol. 11, no. 1, pp. 1-11, 2021. [CrossRef] [Google Scholar] [Publisher Link]
17. B.M. Rocha, D. Pessoa, A. Marques, P. de Carvalho and R.P. Paiva, "Automatic wheeze segmentation using harmonic-percussive source separation and empirical mode decomposition", *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 1926-1934, 2023. [CrossRef] [Google Scholar] [Publisher Link]
18. H. Liu, S. Liu, A.A. Shkel, Y. Tang and E.S. Kim, "Multi-band MEMS resonant microphone array for continuous lung-sound monitoring and classification", *In 2020 IEEE 33rd International Conference on Micro Electro Mechanical Systems (MEMS) IEEE*, pp. 857-860, (2020, January). [CrossRef] [Google Scholar] [Publisher Link]
19. 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. [CrossRef] [Google Scholar] [Publisher Link]
20. R. Zulfiqar, F. Majeed, R. Irfan, H.T. Rauf, E. Benkhelifa and A.N. Belkacem, "Abnormal respiratory sounds classification using deep CNN through artificial noise addition", *Frontiers in Medicine*, vol. 8, p. 714811, 2021. [CrossRef] [Google Scholar] [Publisher Link]
21. B.J. Kim, B.S. Kim, J.H. Mun, C. Lim and K. Kim, "An accurate deep learning model for wheezing in children using real world data", *Scientific Reports*, vol. 12, no. 1, p. 22465, 2022. [CrossRef] [Google Scholar] [Publisher Link]
22. S. Im, T. Kim, C. Min, S. Kang, Y. Roh, C. Kim and S. Seo, "Real-time counting of wheezing events from lung sounds using deep learning algorithms: Implications for disease prediction and early intervention", *Plos one*, vol. 18, no. 11, p. e0294447, 2023. [CrossRef] [Google Scholar] [Publisher Link]
23. H. Tachibana, N. Ono, H. Kameoka, H and S. Sagayama, "Harmonic/percussive sound separation based on anisotropic smoothness of spectrograms", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, 2059-2073, 2014.
24. F. Canadas Quesada, P. Vera Candeas, N. Ruiz Reyes, P. Alonso Jordá and J. Ranilla Pastor,"Online harmonic/percussive separation using smoothness/sparseness constraints", *In Proceedings of the 12th International Conference in Sound and Music Computing, SMC 2015*, p. 337-342 ,2015.[Google Scholar]
25. K. Drossos, P. Magron, S.I. Mimilakis and T. Virtanen, "Harmonic-percussive source separation with deep neural networks and phase recovery", *In 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC) (pp. 421-425) IEEE*, (2018, September).
26. [CrossRef] [Google Scholar] [Publisher Link]
27. M. Tabata, H. Lu, T. Kamiya, S. Mabu and S. Kido, "Automatic Classification of Respiratory Sound Considering Hierarchical Structure", *In 2022 22nd International Conference on Control, Automation and Systems (ICCAS) (pp. 537-541)IEEE*,(2022, November).[CrossRef] [Google Scholar] [Publisher Link]
28. A. Dawadikar, A. Srivastava, N. Shelar, G. Gaikwad and A. Pawar, "Three Feature Based Ensemble Deep Learning Model for Pulmonary Disease Classification", vol. 10, no. 2, pp. 696-702, 2023. [Google Scholar]
29. Ashu Nayak. (2025). Nanotechnology-Enhanced Eco-Smart Building Materials for Carbon-Neutral and Climate-Resilient Urban Infrastructure. *Journal of Smart Infrastructure and Environmental*

Sustainability, *2*(3), 1-8.
30. M. Kavitha. (2025). Neural Mechanisms of Attention and Inhibition in Complex Decision-Making Tasks: An fMRI-Based Analysis. *Advances in Cognitive and Neural Studies*, *1*(3), 1-7.
31. Namrata Mishra. (2025). A Multidisciplinary Exploration of Intelligent Systems: Technical Advances and Societal Implications. *Bridge: Journal of Multidisciplinary Explorations* , *1*(3), 27-33.
32. Nisha Milind Shrirao, & Nidhi Mishra. (2025). Climate Change and Biodiversity Loss: An Ecological Network Analysis Perspective. *Frontiers in Life Sciences Research*, 22–28.