Machine Learning-Based Classification of Pulmonary Diseases through Real-Time Lung Sounds

Sangeetha Balasubramanian*, Periyasamy Rajadurai

Department of Instrumentation and Control Engineering, National Institute of Technology, Tamil Nādu, India

Received 24 May 2023; received in revised form 14 August 2023; accepted 15 August 2023

DOI: https://doi.org/10.46604/ijeti.2023.12294

Abstract

The study presents a computer-based automated system that employs machine learning to classify pulmonary diseases using lung sound data collected from hospitals. Denoising techniques, such as discrete wavelet transform and variational mode decomposition, are applied to enhance classifier performance. The system combines cepstral features, such as Mel-frequency cepstrum coefficients and gammatone frequency cepstral coefficients, for classification. Four machine learning classifiers, namely the decision tree, k-nearest neighbor, linear discriminant analysis, and random forest, are compared. Evaluation metrics such as accuracy, recall, specificity, and f1 score are employed. This study includes patients affected by chronic obstructive pulmonary disease, asthma, bronchiectasis, and healthy individuals. The results demonstrate that the random forest classifier outperforms the others, achieving an accuracy of 99.72% along with 100% recall, specificity, and f1 scores. The study suggests that the computer-based system serves as a decision-making tool for classifying pulmonary diseases, especially in resource-limited settings.

Keywords: cepstral coefficients, discrete wavelet transform, machine learning classifiers, pulmonary diseases, variational mode decomposition

1. Introduction

Pulmonary diseases have emerged as a significant cause of the highest mortality in society. World Health Organization (WHO) categorizes the “big five” respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), acute lower respiratory tract infections, lung cancer, and tuberculosis, are responsible for causing the deaths of over 3 million individuals globally each year. These respiratory diseases share identical symptoms, such as adventitious breathing, which can complicate the diagnostic procedure. Due to their severe consequences, an early and precise diagnosis of these types of diseases has become crucial [1].

Diagnosis of pulmonary illnesses can be done clinically in a variety of ways. Imaging techniques like chest X-rays, computer tomography scans, and magnetic resonance imaging are used to diagnose pulmonary diseases. Contrarily, adopting these imaging modalities presents several difficulties, including the risk of repeated exposure to harmful radiation, the expense of equipment, and the challenge of deploying these methods in remote areas. A spirometer is a common technique used to diagnose lung function. It measures the air inhaled and exhaled by the lungs and identifies irregular breathing patterns. However, this technique has challenges such as the cooperation of patients in forced breathing, and requires a professional operator. Due to its high cost, this device can be used only in clinical settings, needs regular calibration, and is inefficient in detecting obstructive-restrictive abnormalities [2].

* Corresponding author. E-mail address: sangeetha27may@gmail.com

English language proofreader: Yen-Chun Hsieh

In recent years, detecting pulmonary disease through lung sounds (LSs) has been an area of interest in bioinformatics. LSs provide valuable information about pulmonary diseases and can be heard throughout the posterior and anterior regions of the chest. Auscultation is a cost-effective, non-invasive technique for hearing sounds of the lungs, heart, and internal organs of the body. The present technique utilizes a stethoscope, a common clinical tool used by healthcare professionals to listen for LSs and detect various pulmonary diseases. Conventional auscultation is widely used, but it has limitations in clinical applications and research suitability. These limitations are due to: (i) expertise needed to annotate LSs, (ii) inability to monitor continuously, and (iii) inherent inter-listener unpredictability. Therefore, an automated respiratory sound analysis could circumvent these restrictions with the help of an electronic stethoscope [3].

LSs are categorized into two major classes: normal and abnormal (adventitious). Normal LSs are non-musical sounds produced by breathing over the chest wall and trachea in the frequency range of 150 to 1000 Hz. However, adventitious LSs are musical and can be continuous or discontinuous, occurring between 200 and 2000 Hz. Vesicular and bronchial LSs are typically observed in the absence of respiratory disorders. Abnormal sounds, often indicate complications in the lungs or airways. Examples of abnormal breath sounds include rhonchi, crackles, wheezing, and harsh stridor [4]. Identifying pulmonary diseases by computer-based analysis of these abnormal LSs leads to an automated computer-based pulmonary disease classification system. Combining the clinical method of auscultation with machine learning (ML) algorithms enhances the accuracy and efficiency of diagnosing lung illnesses. Significant efforts have been dedicated to the classification of lung disorders; most studies rely on publicly available online repositories. Only a limited number of studies involve real-time data collection from hospitals. The denoising is a preliminary step before proceeding to disease classification.

In the proposed work, two denoising method VMD-DWT-based decomposition is also incorporated to further enhance the signal quality and denoise the respiratory signal. By employing two stages of denoising, the LS signal is free from noise, thus improving its overall quality and reliability. Mel frequency cepstral coefficients (MFCC) and gammatone frequency cepstral coefficients (GFCC) are employed for feature extraction. The MFCC and GFCC have gained popularity in processing various human sound signals, including voice, cardiac sounds, and certain LSs. These two feature extraction methods show promising outcomes in application to audio signals from the human body.

Four ML models are compared in this work: LDA, k-NN, DT, and random forest (RF). These models are selected based on their distinct features, which include the LDA can effectively handle scenarios where the number of features is significantly larger than the number of training samples. k-NN is a versatile algorithm that can accommodate different proximity calculations. Its intuitive nature and memory-based approach make it easy to understand and implement. DT simplicity, as the decision-making process is visually and conceptually straightforward. The tree structure allows for easy interpretation and understanding of the classification process. RF is a powerful algorithm capable of performing regression and classification tasks. It generates reliable predictions that are easily interpretable. Additionally, RF can efficiently handle large datasets, making it suitable for scalability and computational efficiency.

Therefore, this study aims to emphasize these aspects and identify an effective and reliable method for categorizing three major classes of diseases: asthma, COPD, bronchiectasis, and the normal category. The research gap from the literature review reveals that many studies lack in investigating the crucial steps of LS denoising before feature extraction and classification. Furthermore, the combination of cepstral features has yet to be explored in the literature. The proposed work addresses these gaps by implementing a two-stage approach for LS denoising, followed by combined cepstral feature extraction and classification using ML techniques. The contribution of the proposed study is as follows:

(1) To denoise the collected LS signal by applying the variational mode decomposition technique and DWT to the raw data.

(2) To extract combinations of MFCC and GFCC features to classify asthma, COPD, bronchiectasis, and normal LSs.

(3) To examine various 4-class classifiers like LDA, DT, k-NN, and RF classifiers.

2. Literature Survey

The automatic classification of pulmonary diseases based on LSs using different ML and deep learning approaches has been published in various studies [5]. Numerous publications have notable results in denoising LSs followed by feature extraction and pulmonary disease classification. However, many studies lack a solid experimental design, which causes exaggerated outcomes. The literature has employed several methods for lung sound denoising, feature extraction, and lung disease classification. Some of the methods used for the classification of LSs can be found in a recent paper as follows:

Yan Shi et al. [6] utilized a wavelet-based denoising method to effectively reduce noise in LSs. By extracting wavelet-based features and combining them with linear discriminant analysis (LDA) and backpropagation neural network algorithms, an impressive accuracy of 92.5% was achieved. Georgios et al. [7] proposed a methodology of convolutional neural network (CNN) and long short-term memory (LSTM) networks that implement focal loss for categorizing four classes of sounds. This hybrid model validated the result on the ICBHI 2017 dataset and achieved an accuracy of 76.39% using the hybrid CNN-LSTM model. Acharya and Basu [8] applied the mel spectrogram in a hybrid convolutional neural network-recurrent neural network (CNN-RNN) architecture to classify four sound classes to achieve a score of 71.81%. Shuvo et al. [9] introduced a lightweight CNN architecture that utilizes a hybrid scalogram derived from continuous wavelet transform and empirical mode decomposition (EMD). This architecture was designed for respiratory disease classification and evaluated using the ICBHI 2017 scientific challenge dataset, achieving an impressive accuracy of 99.20%.

Acar Demirci et al. [10] employ Mel-frequency cepstral coefficients (MFCC), EMD, wavelet transform for feature extraction, and k-nearest neighbor (k-NN) to reach 98.8% accuracy. Gökçen [11] participated in a study on detecting COPD using empirical wavelet transform in conjunction with the Adaboost classifier. The study aimed to achieve a high level of accuracy and successfully reached an accuracy of 95.28%. Haider and Behera [12] conducted a study focused on denoising LSs using a combination of EMD, hurst analysis, spectral subtraction denoising, and wavelet packet decomposition to extract wavelet-based features. A decision tree (DT) classifier with wavelet features is employed to classify normal, COPD, and asthma conditions. This approach achieved an accuracy of 99.3% in classifying these respiratory diseases. Asatani et al. [13] introduced a respiratory sound classification that utilizes spectrograms as the primary feature representation. The study employed a convolutional RNN to perform the classification task. Rizal and Puspitasari [14] conducted a study to explore the combination of wavelet packet decomposition and discrete wavelet transform (DWT) for the classification of respiratory sounds. This study utilized a multilayer perceptron as a classification model to achieve an accuracy of 98.99%.

3. Materials and Methods

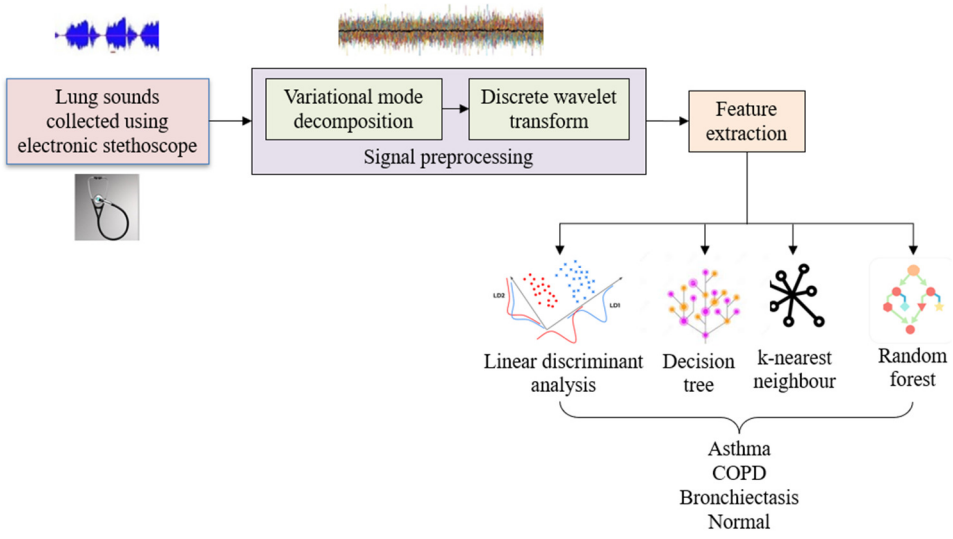


Fig. 1 Workflow of the proposed method

An overview of the theory and knowledge of the study basis is explained in this section. Fig. 1 illustrates the detailed block diagram of the steps involved in LS analysis for pulmonary disease classification. Since applying ML techniques directly to audio data is challenging, the signal is initially denoised by two different denoising techniques, followed by relevant feature extraction. Significant cepstral features like MFCC and GFCC are given to ML classifiers for accurate classification. The DT, k-NN, RF, and LDA classifiers are examined to identify the best outcome of the classifier. The following sections briefly explain the data collection procedure, denoising technique, feature extraction, and ML classifiers.

3.1. Experimental study

The LS data for the proposed study was recorded in hospitals from patients with various diseases. Table 1 shows detailed demographic information about the patients. The pathological sounds were collected from the outpatient section, thoracic medicine department, Thanjavur Medical College and Hospital, Thanjavur, Tamil Nadu, India.

Table 1 Demographic information of the patients

Pulmonary condition	Number of patients	Number of recordings	Age group	Gender	Mean±standard deviation
Asthma	58	222	38-78	Male (14), Female (44)	48±12.5
COPD	11	57	45-72	Male (8), Female (3)	52±13.6
Bronchiectasis	25	54	36-58	Male (12), Female (13)	41±7.8
Healthy	26	35	35-62	Male (15), Female (11)	55±5.2

The data collection process was carried out according to existing clinical standards and agreed upon by the institutional ethical committee for human studies (Reg No. EC/NEW/INST/2020/1058). The LSs included 58 asthma patients, 11 COPD patients, 25 bronchiectasis patients, and 26 normal subjects, with 368 recordings. The patients were asked to relax in a sitting position before the recording procedure. Subjects were instructed to breathe normally and remain motionless to minimize motion artifacts during digital signal recording.

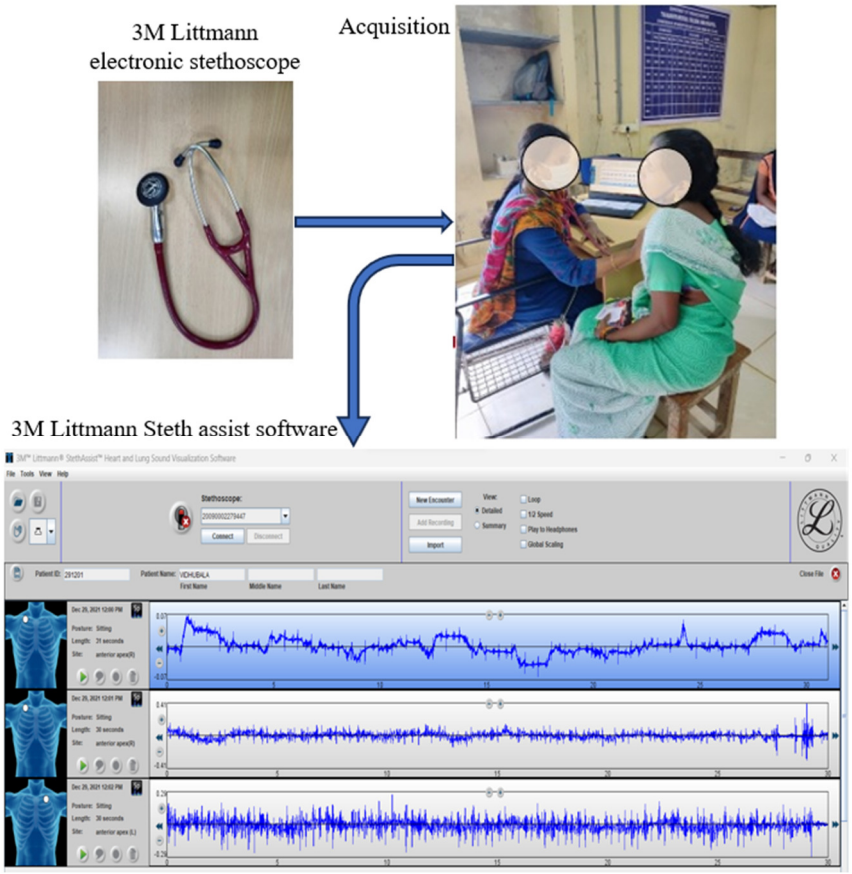


Fig. 2 Recording procedure in hospital

The LSs are collected using a single-channel electronic stethoscope (3M Littmann model 3200) placed on the anterior and posterior sides of the chest wall. Pulmonologists recommend pathological patients and auscultation sites based on the disease in the outpatient department. All signals were bandlimited to 20 Hz to 2 kHz and recorded at a 4 kHz sampling frequency. The sounds are collected for 20 seconds and stored in .wav file format. The healthy subjects recruited for this study are non-smokers with no history of serious lung diseases. The collected LSs are stored in Littmann Steth Assist software installed on a laptop connected via Bluetooth. The validation of real-time acquired LS signal is compared with the publicly available online ICBHI 2017 dataset [4]. The LS analysis is performed on MATLAB 2021a software, installed on a Lenovo laptop with 8 GB RAM and an AMD Ryzen processor. Fig. 2 shows the recording procedure.

3.2. Signal preprocessing

Pre-processing the collected data is the initial step before applying it to the next classification stage. Background noise and heart sound interference have been recorded with LSs from the hospital. These noises distort LS characteristics and lead to misdiagnosis. Therefore, it is essential to eliminate noise before feature extraction. Filtering between the specified frequency bands employs two filtering techniques: variational mode decomposition and discrete wavelet transformation techniques. The next section discusses the implementation of two denoising algorithms.

3.2.1. Variational mode decomposition

Variational mode decomposition is a technique to decompose a signal into a sum of intrinsic mode functions (IMFs) called modes. These modes exhibit localization in both time and frequency [15]. VMD is founded on a variational principle, enabling the estimation of IMFs and their corresponding time-frequency properties in a mathematically well-posed and computationally efficient manner. For LS denoising, VMD can be employed by decomposing the LS signal into its constituent modes and eliminating noise-associated ones. Noise detection involves identifying modes with low energy in the LS frequency band and high energy in the noise frequency band. The identification of noisy modes is based on the computation of their instantaneous frequency values, which can be derived from the modes. Subsequently, these noise-related modes can be eliminated from the original signal to obtain a denoised LS signal.

The general steps of the VMD algorithm include:

Step 1: Set the signal as first IMF and residue as zero.

Step 2: Iterate till the stopping criterion is met and the residue is zero.

Step 3: Calculate Hilbert transform.

Step 4: The signal is divided into an envelope and instantaneous frequency.

Step 5: Define a Gaussian function using an envelope.

Step 6: Define the cost function using a weighting function.

Step7: Obtain IMF by minimizing a cost function.

Step 8: Obtain residue by subtracting the IMF from the signal.

Step 9: Signal is set as residue and the process is repeated.

Step 10: Repeat the process until the stopping criterion is met and the residue is zero.

3.2.2. Discrete wavelet transforms

The wavelet transform involves transforming a signal into a set of basic functions called wavelets. These wavelet functions form orthogonal groups and are variations of a single mother wavelet that are dilated, translated, and scaled. Wavelet transform is a technique used for signal analysis in the time-frequency domain. The signal undergoes decomposition into

different stages by passing through two half-bands of frequencies: the low-pass and high-pass band frequencies. This process yields approximation coefficients and detailed coefficients [16]. The first discrete mother wavelet, represented by h(n), is a high-pass filter, while its mirror image, g(n), is a low-pass filter. The high-pass filter produces detailed coefficients, while the low-pass filter generates approximate coefficients. Following signal filtering and adhering to Nyquist sampling criteria, the signal is down-sampled by a factor of two.

In signal analysis utilizing wavelet transform, the selection of a suitable wavelet and determining the number of decomposition layers play a crucial role. Multiple wavelet types are available, and the most efficient one is chosen based on the specific application requirements. In this study, the Daubechies 4 (db4) wavelet is considered the mother wavelet and a hard threshold is applied to the detailed coefficients. The denoised signal is reconstructed using both the approximation and detailed coefficients. The dominant frequency of the signal influences the determination of the number of decomposition levels. The levels are carefully chosen to ensure that the wavelet coefficients retain the signal components that correlate strongly with the frequencies essential for signal classification.

3.3. Feature extraction

The next stage is extracting features from the denoised signal to train a machine-learning model. The denoised data is converted into useful characteristics that aid the ML classifier in categorizing various diseases. Extracting useful information from signals is necessary to reduce the classifier computation time [17]. Respiratory disease detection reliability depends on properly identifying prominent features in the filtered signal indicating diseases. Physiological changes in the lungs mean an obstruction in the lungs. Patients with pulmonary diseases exhibit structural breathing cycle variations compared to healthy subjects. Since the LS signal is non-stationary and non-linear, one feature extracted from the data is insufficient to discriminate between diseases. A total of 26 features, including 13 MFCC features and 13 GFCC features, were derived from VMD-DWT-filtered data. These extracted features are used to measure the asymmetric variations of the respiration signal based on the morphological changes observed in the abnormal condition. Noticeable changes can be observed between normal subjects and pathological patients. MFCC and GFCC feature extractions are shown in the next section.

3.3.1. Mel-frequency cepstral coefficients

MFCC is a widely utilized technique for feature extraction in speech recognition and audio signal processing [17]. These coefficients are derived from the power spectrum of a speech signal. This spectrum transforms a Mel-frequency spectrogram through the application of a filter bank. Subsequently, the Mel-frequency spectrogram is converted into the cepstral domain using the discrete cosine transform (DCT). This transformation yields a set of coefficients representing the audio signal's spectral envelope. It has been observed that the human perception of voice signals' frequency components is not linear. To address this, the Mel scale was developed as a more suitable scale. The Mel scale converts frequencies on the hertz scale into a hierarchy where listeners perceive equal distances between pitches. The mapping is based on the observation that the perceived gap between lower frequencies is much larger than the perceived distance between higher frequencies. The Mel scale was designed to establish this more appropriate scale. The Mel frequency equation is expressed as follows.

$$M(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

3.3.2. Gammatone frequency cepstral coefficients

GFCCs share similarities with MFCCs but are derived using a different type of filter bank called the gamma tone filter bank [17]. This filter bank offers a closer approximation of the human auditory system's frequency response. Unlike spectrograms, which have certain limitations, gamma tone grams obtained from gamma tone filters are more suitable as they

closely resemble the human ear's characteristics. The cochlea, a component of the human ear, possesses a membrane called the basilar membrane. When subjected to external sound, this membrane vibrates and generates energy corresponding to the incoming sound's frequency. The design of the gamma tone filters aims to replicate the vibrations occurring in the human ear, producing the gamma tone gram. This auditory map resembles a spectrogram and provides valuable information about the acoustic response to different frequencies. The impulse response of gamma tone filters is represented by the equation below. This is denoted as the product of a gamma distribution and a sinusoidal tone centered at the frequency f_c.

$$g_m(t) = pt^{n-1}e^{-2\pi qt}\cos\left(2\pi f_c t + \varphi\right);\ t > 0 \tag{2}$$

3.4. Classifier

Lung disease classification is based on distinct features extracted from the cepstral domain features in the previous section. Four ML classifiers were trained and compared for pulmonary disease classification. The combined features were fed to four classifiers (LDA, DT, k-NN, and RF) to find the highest classification accuracy.

3.4.1. Linear discriminant analysis

A supervised learning approach called LDA is used in ML for classification applications [12]. It is a method for determining the optimum linear combination of features for classifying a dataset into different classes. For LDA to function, the data are projected onto a lower-dimensional space with maximum class separation. This algorithm determines the directions in the feature space that effectively distinguish the various data classes. LDA has two assumptions: that the covariance matrices of the different types are equal and that the data is Gaussian. It also presumes that the data can be separated linearly, meaning that a linear decision boundary is enough to classify different classes.

3.4.2. Decision tree

A predictive model known as a DT maps decisions and their outcomes in a tree-like structure. The nodes and branches of the DT reflect a decision or an attribute; each branch is a potential result or value. The tree's root node serves as the initial choice or forecast, and the leaf nodes serve as the conclusion. They are especially helpful when there are a lot of variables and a complex decision-making process. The most crucial features are chosen, and the data is then divided into smaller groups depending on these attributes to create a DT. The objective is to build a tree that can correctly categorize or forecast results based on the input data [18].

3.4.3. k-nearest neighbor

The number of nearest neighbors used to produce a forecast for a new data point is indicated by the "k" in a k-NN. The method measures how far each new data point is from each training data point in the dataset. Following the selection of the k closest neighbors based on the calculated distances, the prediction is made using either the average value (for regression) or the most frequent class label (for classification) of the k closest neighbors. k-NN is a straightforward and efficient technique that can be used to solve specific types of problems, including those with few features or plenty of data [18].

3.4.4. Random forest

The RF method operates by building DT on each subset of the training data and characteristics. A greedy approach creates DT, which recursively separates the data according to the most important feature. This procedure is repeated until the tree's maximum depth is reached or all samples at a certain node belong to the same class. The method aggregates the outcomes of each DT in the forest at prediction time to provide the final prediction. The final prediction is often the mean prediction for regression or the mode of the classes for classification [19].

3.5. Evaluation metrics

The performance of two denoising techniques is evaluated using two metrics: signal-to-noise ratio (SNR) and mean square error (MSE) The SNR represents the ratio of signal power to noise power in dB. MSE is the quantity that measures the average of the square of the mean differences between predicted and actual values. To assess the performance of the classifier, accuracy, recall, precision, and f1 score are used, and they are given below. The dataset is split into 80% of the data for training and the remaining 20% for testing the proposed ML classifier. 10-fold cross-validation and a 20% holdout are used to evaluate the classifier's performance [19].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$Recall = \frac{TP}{TP + FP} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{6}$$

where true positives (*TP*) are events that are accurately classified into the relevant class. Events from other classes that are accurately classified are considered true negatives (*TN*). False negatives (*FN*) are events of the specified class that are mistakenly classified, and false positives (*FP*) are events that are incorrectly classified as the specified class. An accuracy under curve-receiver operating characteristics (AUC-ROC) curve is also obtained to visualize the performance of the classifier along with the four metrics mentioned above [20]. The training time and testing time of all ML algorithms are also considered to evaluate the performance of a classifier.

4. Results

This section reveals the outcomes achieved in each stage of the proposed method. The decomposition outcome of two denoising techniques and their performance measures show the effective removal of artifacts. Followed by extraction of combined cepstral features. Finally, the performance of four machine learning algorithms in this section is illustrated using the respective confusion matrices and ROC curves.

4.1. Signal preprocessing

The signal processing of the collected raw data is carried out in two steps. The raw LS is initially applied to the VMD technique, and the filtered signal is again given to DWT for further denoising to increase the SNR. Fig. 3 shows two stages of the denoising technique.

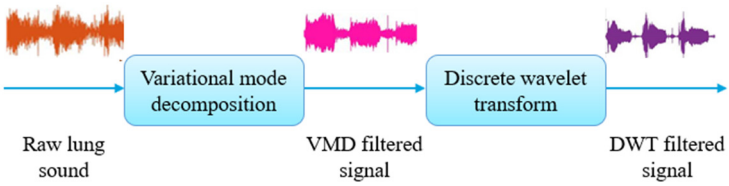


Fig. 3 Two stages of denoising

4.1.1. Variational mode decomposition

The collected signal contains noise, requiring the initial step to involve the removal of these noises from the raw LS signal. It is worth noting that the VMD algorithm is an iterative algorithm that decomposes the signal into a sum of IMFs at each

iteration. The number of IMFs obtained by the VMD algorithm is not predetermined and varies depending on the characteristics of the original signal and the specific noise present. The specific parameters and stopping criteria have also been adjusted based on the application and signal quality. Fig. 4 displays the time-domain representation of the original LS (COPD) signal.

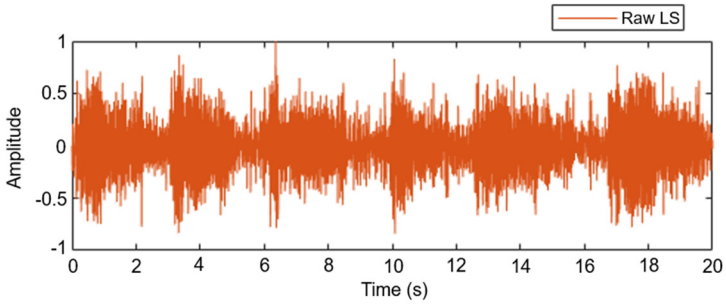
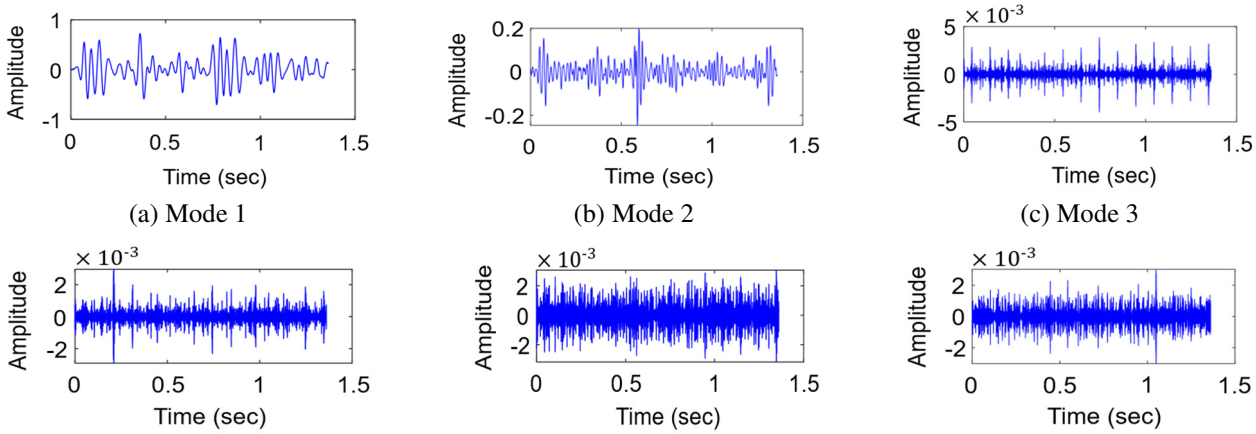


Fig. 4 Time domain representation of raw LS for COPD

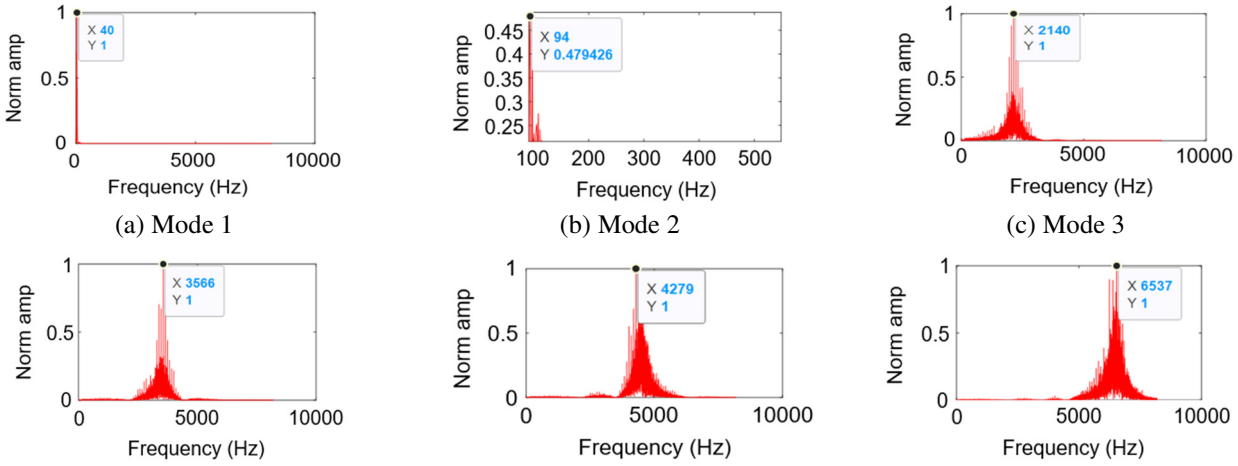


(a) Mode 1 (b) Mode 2 (c) Mode 3

(d) Mode 4 (e) Mode 5 (f) Mode 6

Fig. 5 Graphical representation of Mode1 to Mode 6 obtained by VMD technique

The VMD algorithm decomposes raw LS data into different modes and residues. Fig. 5 displays the six modes generated by applying the VMD algorithm to raw LS data. The raw LS is sliced into 1.5 seconds of data for analysis and reconstructed to 20 seconds of data after filtering. The fast Fourier transform (FFT) is applied to each decomposed mode to determine the dominant frequency of all modes.



(a) Mode 1 (b) Mode 2 (c) Mode 3

(d) Mode 4 (e) Mode 5 (f) Mode 6

Fig. 6 Graphical representation of frequency spectrum of Mode 1 to Mode 6 obtained using VMD technique

Fig. 6 shows the frequency spectrum of all modes obtained using FFT. Each mode's dominant frequency helps to identify the region of interest in LS. Mode 1, with a lower frequency of 40 Hz, falls below the region of interest and it is eliminated for signal reconstruction. Mode 2 and Mode 3 have 94 and 2140 Hz within the LS frequency range, respectively. As these two

modes fall within the region of interest, these two modes are considered for signal reconstruction. Mode 4 has a frequency of 3566 Hz, Mode 5 has 4279 Hz, and Mode 6 has a very high frequency of 6537 Hz as their dominant frequency. Since these three modes have frequencies exceeding 3000 Hz, they are discarded in the subsequent signal-filtering stage. For the next level of decomposition, IMFs with a dominant frequency of 100 to 2000 Hz are considered. The reconstructed signal using Modes 2 and 3 is presented in Fig. 7. The SNR of the reconstructed signal is 12.6 dB, which is relatively low. Further, DWT-based filtering is required to reduce noise on the reconstructed waveform in the subsequent filtering stage.

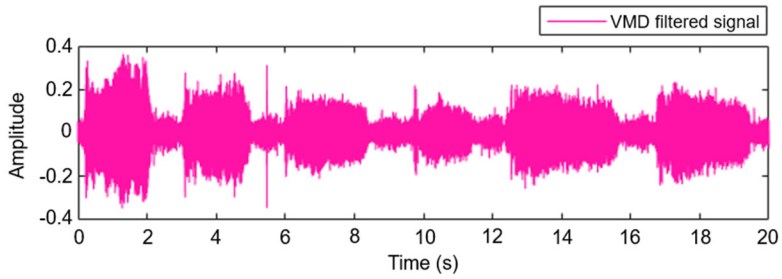


Fig. 7 Time domain representation of denoised signal using VMD

4.1.2. Discrete wavelet transforms

The DWT denoising method encompasses signal decomposition, coefficient thresholding, and reconstruction. The VMD-reconstructed signal is given as input for further denoising through a DWT-based method. The VMD-reconstructed signal is decomposed into four levels within the 94 to 2140 Hz frequency band. The waveforms obtained using DWT decomposition are illustrated in Fig. 8. Different wavelets with various thresholds are tested on the signal using a trial-and-error approach to achieve the optimal denoising output. Among these, the db4 wavelet applied at the fourth level of decomposition with a soft threshold produces superior output. Subsequently, the signal is reconstructed using approximation and detailed coefficients with a soft threshold. Fig. 9 displays the time-domain representation of the denoised signal, demonstrating a significant reduction in high-frequency noise. The SNR increases from the initial filtering stage, ensuring a near-perfect reconstruction of the denoised signal.

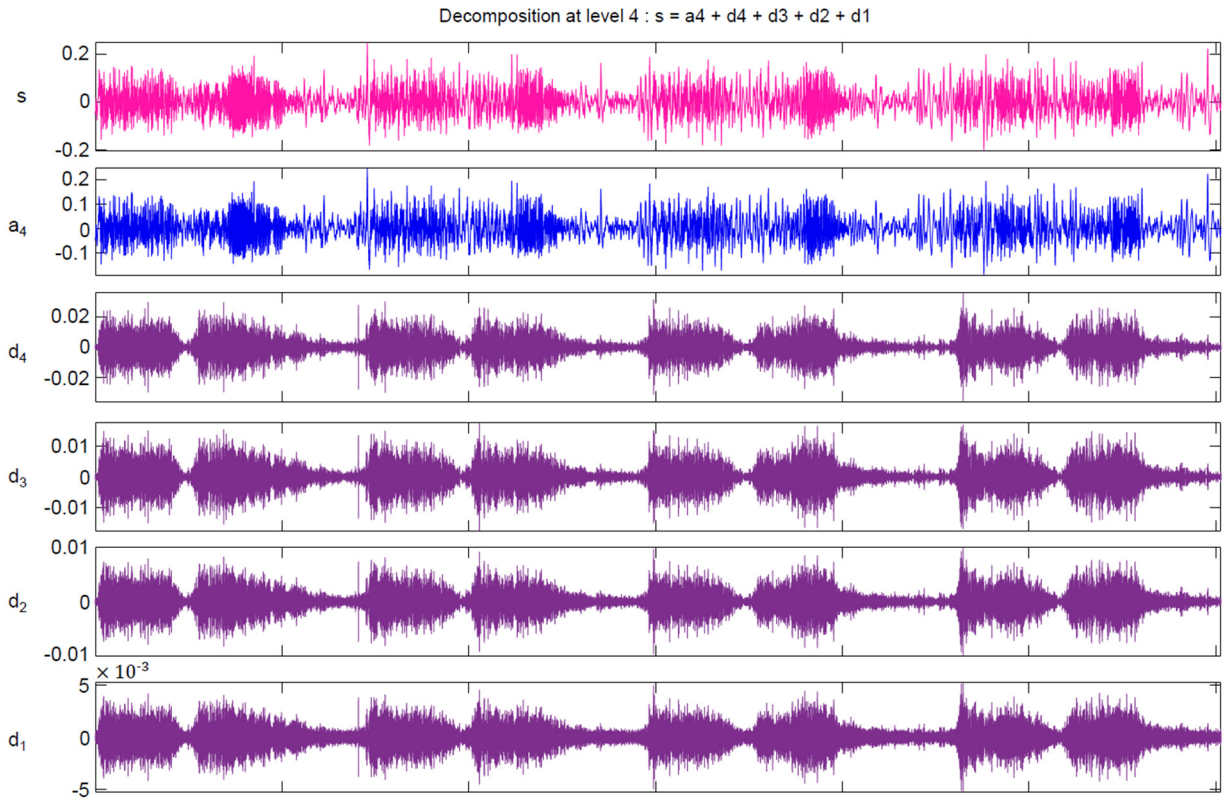


Fig. 8 Decomposition using DWT output at level 4

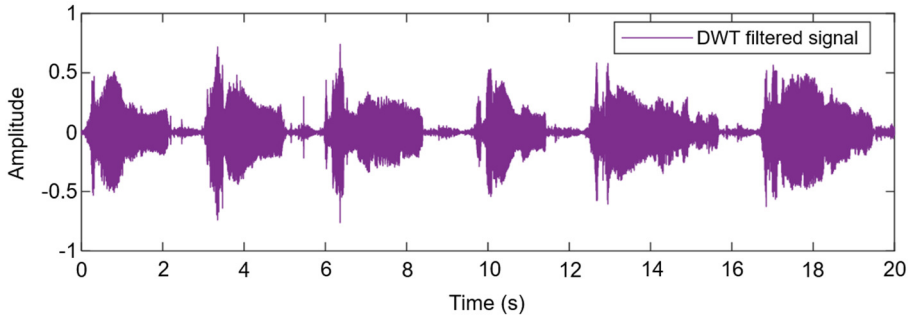


Fig. 9 Graphical representation of denoised signal using DWT

Table 2 illustrates the metrics obtained after denoising using VMD and DWT. The table shows metrics obtained for one LS from each category of disease. The MSE values indicate the difference between the denoised signals and the original noise-free signals and the MSE error should be as small as possible. The MSE obtained after VMD for healthy individuals is 0.0765. Further denoising using DWT results in an MSE of 0.0075. This indicates that the proposed method achieves better denoising performance with minimal signal loss, as evidenced by the smaller MSE values. Similarly, the pattern repeats for other diseases. To assure the quality of the signal, another metric SNR is measured which should be as high as possible. The SNR obtained after VMD is 15.68 dB whereas after applying DWT the SNR has increased to 19.47 dB. Hence two stages of the denoising method effectively reduce noise while preserving the strength and quality of the LS signal.

Table 2 Evaluation metrics of the VMD-DWT denoising method

Metrics/Diseases	MSE after VMD	MSE after DWT	SNR after VMD (dB)	SNR after DWT (dB)
Asthma	0.0852	0.0202	12.81	16.78
COPD	0.0185	0.0154	14.59	18.96
Bronchiectasis	0.0845	0.0185	14.99	19.57
Healthy	0.0765	0.0075	15.68	19.47

4.2. *Feature extraction*

In the classification of lung diseases, features play a vital role as they serve as distinguishing characteristics used by the classifier. Extracting significant features from the LS signals is essential to ensure accurate differentiation between the various classes. However, due to the non-stationary nature of LSs, a single feature alone is inadequate for precise categorization. Hence, multiple features were extracted for this study. At this stage, 13 MFCC and 13 GFCC features were extracted from denoised LS signals. It is possible to comprehensively represent signal characteristics by capturing relevant information about the spectral envelope and auditory perception of LSs. By incorporating these 26 features, the classifier can effectively analyze and classify the different classes of lung diseases.

4.3. *Classification*

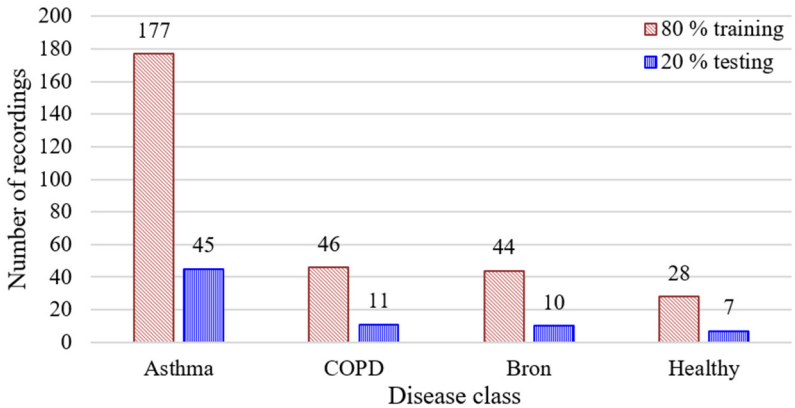
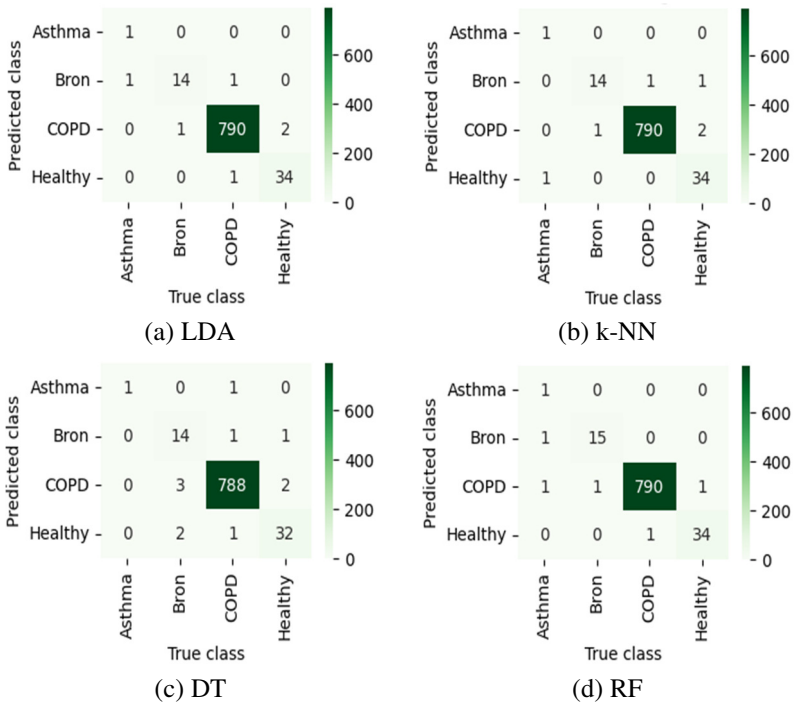


Fig. 10 Class split of audio files into training and testing datasets

This study employs four ML algorithms, namely LDA, DT, k-NN, and RF, to classify pulmonary diseases. The performance of these algorithms is evaluated using 10-fold cross-validation and a 20% holdout method, where 80% of the data is used for training and 20% for testing. Fig. 10 shows the class-wise split of audio files for training and testing. Fig. 11 shows the confusion matrix of all classifiers for the ICBHI dataset. Table 3 displays the evaluation metrics obtained for both datasets: ICBHI and hospital collected dataset. The metrics resulting for both datasets are closer to each other except for a few models. The difference between the metrics of both datasets is that a large data imbalance is observed in the ICBHI dataset, as it contains only one asthma sound and a maximum number of COPD sounds. RF achieved superior results with a remarkable accuracy of 99.72%, 100% recall, 100% precision, and a 100% f1 score.



(a) LDA (b) k-NN

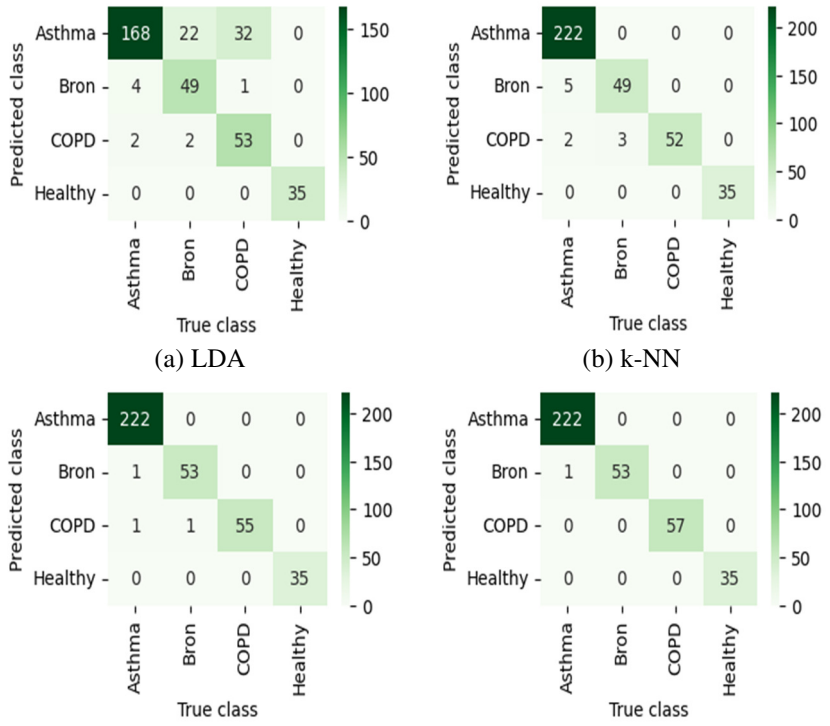
(c) DT (d) RF

Fig. 11 Confusion matrix of ICBHI dataset

Table 3 Evaluation metrics of four ML classifiers

Classifier/Metrics	Dataset	LDA	k-NN	DT	RF
Accuracy (%)	ICBHI dataset	92.2	96.9	99.3	99.72
	Hospital collected dataset	82.88	97.28	99.18	99.72
Recall (%)	ICBHI dataset	84.38	83.77	91.1	81.02
	Hospital collected dataset	90.45	95.26	96.94	100
Precision (%)	ICBHI dataset	96.06	96.06	82.07	97.62
	Hospital collected dataset	81.56	97.62	98.12	100
F1 score (%)	ICBHI dataset	88.15	87.95	84.39	99
	Hospital collected dataset	84.45	97.63	99.85	100
AUC micron average	ICBHI dataset	0.92	0.93	0.91	0.92
	Hospital collected dataset	0.93	0.69	0.82	0.97
AUC macro average	ICBHI dataset	0.91	0.94	0.82	0.95
	Hospital collected dataset	0.91	0.62	0.82	0.97

These metrics indicate the effectiveness of the RF algorithm in accurately classifying the different pulmonary diseases. Fig 12 shows the confusion matrix of four models for the hospital-collected dataset. Fig 13 shows the training time and testing time required by the four models for the hospital-collected dataset. From the figure, it is evident that LDA requires more training time compared to other models. On visual interpretation, it is observed that RF needs less time for both testing and training the dataset. Fig. 14 shows the AUC-ROC curve of four models for the hospital-collected dataset. Upon comparing the AUC-ROC curves and confusion matrix of all four classifiers, it is observed that RF outperforms the others in classifying various disease classes.



(a) LDA (b) k-NN

(c) DT (d) RF

Fig. 12 Confusion matrix of hospital collected dataset

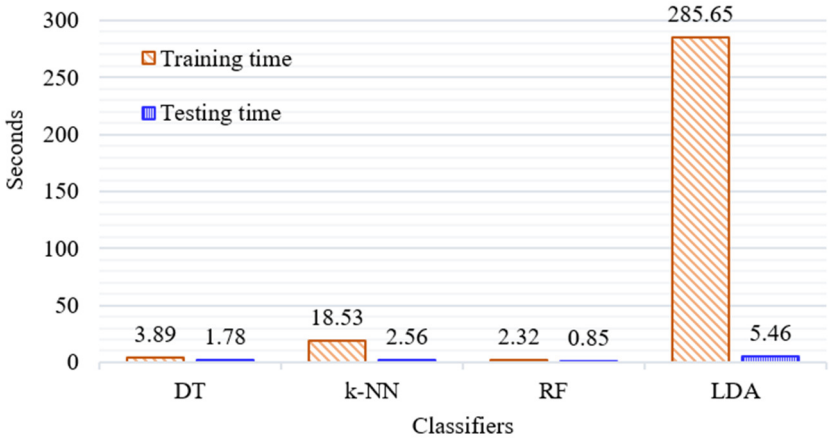
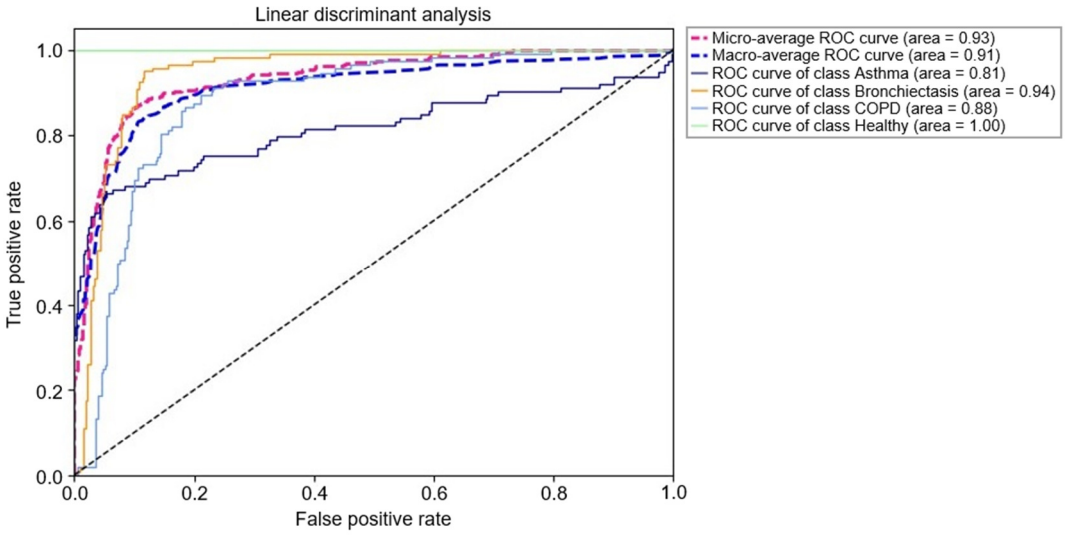
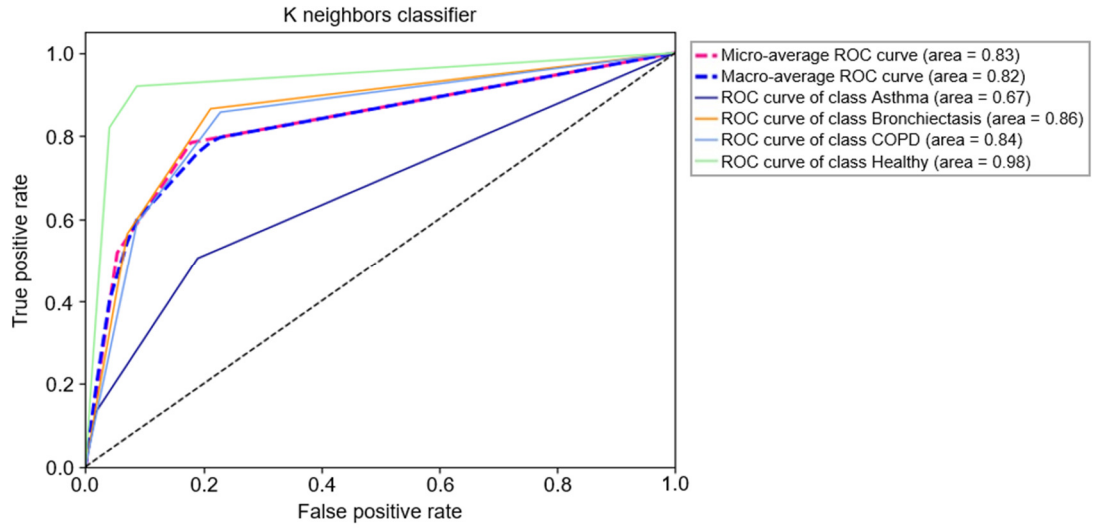


Fig. 13 Training time and testing time of four ML models

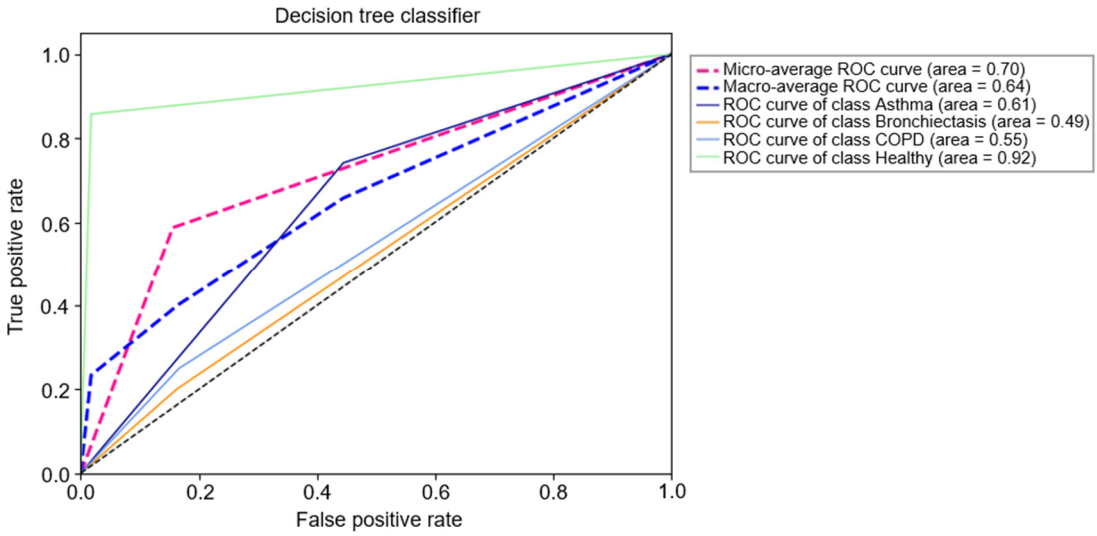


(a) LDA

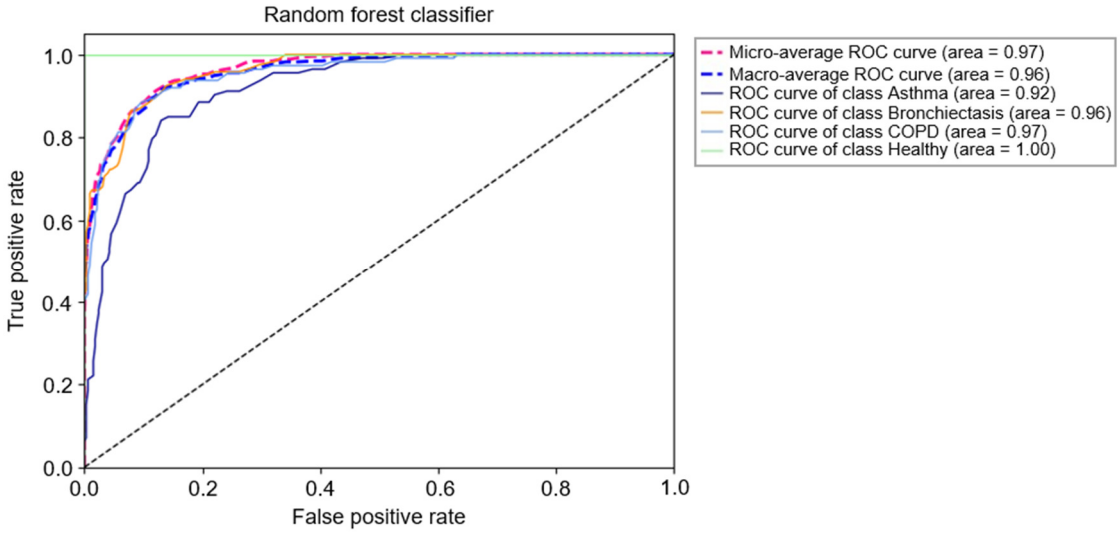
Fig. 14 AUC-ROC curve for hospital collected dataset



(b) k-NN



(c) DT



(d) RF

Fig. 14 AUC-ROC curve for hospital collected dataset (continued)

5. Discussion

This article presents a series of experiments conducted to evaluate the effectiveness of an adventitious respiratory sound classification system. The proposed approach offers a cost-effective and non-invasive technique for distinguishing between

asthma, COPD, bronchiectasis, and healthy categories based on LSs. A substantial number of data samples collected from the study participants were utilized in the analysis. To address the challenges of noise, which can significantly degrade signal quality, each respiratory sound undergoes denoising using two distinct techniques: variational mode decomposition and DWT-based denoising. These techniques yield a high SNR value of 19.47 dB, enhancing the quality of the LSs. A total of 26 cepstral features were extracted from the MFCC and GFCC coefficients, with 13 features obtained from each coefficient. These features serve as inputs to a four-ML classifier. The article compares the performance of four different classifiers, namely DT, LDA, RF, and k-NN, to determine the best classifier with the highest classification accuracy.

The existing literature on pulmonary disease classification primarily focuses on COPD, normal lung function, and asthma detection. This research gap makes it challenging to directly compare the findings of this study to those achieved using similar methodologies. Nonetheless, this study represents a novel step in the direction of screening for COPD, asthma, bronchiectasis, and normal category. The proposed technique outperforms previously published methods regarding pulmonary disease classification in terms of accuracy, precision, recall, and f1 score. Table 4 demonstrates its superior performance in classifying the feature vector generated through this pre-processing stage. Despite the challenges in directly comparing methodologies, this research contributes to advancing pulmonary disease classification and demonstrates promising results of 99.72% accuracy.

Table 4 Performance analysis of the proposed method with existing works

Reference	Classes	Features	Method	Results
Elsetrønning et al., 2020 [21]	Crackle and no crackle	Higher-order statistical, spectral features, MFCC	EMD, enhanced empirical mode decomposition (EEMD), DWT, k-NN	Accuracy: 84.38
Jung et al., 2021 [22]	Crackle and wheeze	Short Time Fourier Transform, MFCC	Depth-wise separable convolution neural network	Accuracy: 85.74
Rani et al., 2021 [23]	Bronchial, crepitation, wheezing, and normal	Chroma short-time Fourier transform (STFT), spectral centroid, spectral roll-off, zero crossing rate, MFCC	Artificial neural network (ANN)	Accuracy: 95.6
Brunese et al., 2022 [24]	Asthma, bronchiectasis, bronchiolitis, COPD, pneumonia, and lower or upper respiratory tract infection	Chromagram, root mean square, spectral centroid, bandwidth, spectral roll-off, tonnes, MFCC, zero crossing rate	Neural network algorithm	Accuracy: 98
Neili and Sundaraj, 2022 [25]	Healthy, chronic, and non-chronic	Gammatonegrams	Deep neural network-VGG16	Accuracy: 67.97
Jaffery et al., 2023 [26]	Normal, bronchiectasis, bronchiolitis	MFCC	DWT, k-NN	Accuracy: 99.3
Proposed work	COPD, asthma, bronchiectasis, healthy	MFCC and GFCC	VMD-DWT, LDA	Accuracy: 82.88 Recall: 90.45 Precision: 81.56 F1 score: 84.45
			VMD-DWT, k-NN	Accuracy: 97.28 Recall: 95.26 Precision: 97.62 F1 score: 97.63
			VMD-DWT, DT	Accuracy: 99.18 Recall: 96.94 Precision: 98.12 F1 score: 99.85
			VMD-DWT, RF	Accuracy: 99.72 Recall: 100 Precision: 100 F1 score: 100

Elsetrønning et al. [21] employed a combination of spectral, statistical, and MFCC features, utilizing the k-NN algorithm, which yielded an accuracy of 84.38%. Similarly, Jung et al. [22] trained a depth-wise separable CNN using a combination of STFT and MFCC features, achieving an accuracy of 85.74%. However, it is important to note that these works achieved higher accuracies, their focus was primarily on discriminating between the presence or absence of crackles rather than disease classification. Rani et al. [23] trained an ANN to distinguish between bronchial, crepitation, wheezing, and normal LSs. The author employed a set of time domains, spectral domain, time-frequency domain, and cepstral features in analysis, the study achieved an outcome of 95.6%. Since the study has a computational complexity more features are extracted for this work. Brunese et al. [24] implemented a neural network-based approach for multi-class disease discrimination using multiple features. This study achieved an impressive accuracy of 98%.

However, the high accuracy was obtained solely from a publicly available repository, and the performance on real-world datasets may differ. Neili and Sundaraj [25] utilized gamma tone features and a deep neural network (VGG-16) to classify three disease categories. However, their approach yielded an accuracy of only 67.97%. The accuracy can be improved if the denoising technique is applied to the signal. A considerable number of existing studies have focused on analyzing data without conducting any denoising operations. Even when LS data is recorded using standard equipment, such as an electronic stethoscope, the resulting LS signal often contains noise. Hence denoising becomes a preliminary step in pulmonary disease classification. Few studies only involved LS denoising, Jaffery et al. [26] employed the DWT technique for denoising LS. Further, the study utilized MFCC features in conjunction with the k-NN algorithm, achieving an impressive accuracy of 99.3%. However, it is worth noting that the authors conducted their simulations using publicly available datasets.

The proposed work was experimented on real-time LS data recorded from the hospital and a robust denoising technique VMD-DWT was applied to eliminate noise. The denoising technique enhances the quality of LS samples, thereby improving the classification performance. The combined cepstral features, specifically MFCC and GFCC, yield superior results compared to higher-order statistical features and wavelet features. An RF classifier results in impressive metrics of 99.72% accuracy, 100% recall, 100% precision, and a 100% f1 score for pulmonary disease classification. This means that the RF model is achieving the best performance in terms of correctly identifying positive samples (recall), providing accurate positive predictions (precision), and achieving a balanced trade-off between precision and recall (f1 score). These scores typically occur when the model is well-trained and representative of the underlying population, the model has learned the patterns in the data accurately, and there is no overlap or ambiguity in the classification boundaries. By combining two denoising methods, and combined features, the model can learn more effective patterns from LSs, leading to improved classification of pulmonary diseases.

6. Conclusion and Future Scope

This paper presents an automated computer-based system using ML to classify pulmonary disease. The LS data collected from the hospital was employed in the study. To ensure high-quality data, two denoising techniques, namely variational mode decomposition and DWT, were employed to remove noise from LSs before feature extraction. The feature vector used for classification consisted of 26 cepstral features that aimed to comprehensively represent the LSs while keeping the computational cost low. Four ML models were trained and tested, with the RF classifier achieving remarkable results. The results show an overall accuracy of 99.72%, a recall of 100%, a precision of 100%, and an f1 score of 100% for the RF classifier. The proposed method significantly improves the classification of common lung diseases and can assist pulmonologists in their clinical diagnosis. However, it is important to acknowledge certain limitations of the study. One such limitation is the presence of data imbalance, which may impact the generalizability of the findings. The proposed work did not include a statistical significance analysis of LS features. These two limitations could be considered in future research investigations for enhanced outcomes. In future work, there are plans to develop a cost-effective portable stethoscope for effective LS acquisition to enhance pulmonary disease detection.

Acknowledgment

A sincere gratitude to Dr. Anbananthan at Thanjavur Medical College and Hospital for their invaluable contributions to this research. The expertise and guidance in the field of pulmonology were instrumental in the success of this study. A sincere thanks for providing access to medical facilities and patients for data collection.

Conflicts of Interest

The authors declare no conflict of interest.

Statement of Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Statement of Informed Consent

Informed consent was obtained from all individual participants included in the study.

References

[1] S. M. Levine and D. D. Marciniuk, "Global Impact of Respiratory Disease: What Can We Do, Together, to Make a Difference?" Chest, vol. 161, no. 5, pp. 1153-1154, May 2022.

[2] B. Abera Tessema, H. D. Nemomssa, and G. Lamesgin Simegn, "Acquisition and Classification of Lung Sounds for Improving the Efficacy of Auscultation Diagnosis of Pulmonary Diseases," Medical Devices: Evidence and Research, vol. 15, pp. 89-102, 2022.

[3] T. Grzywalski, M. Piecuch, M. Szajek, A. Bręborowicz, H. Hafke-Dys, J. Kociński, et al., "Practical Implementation of Artificial Intelligence Algorithms in Pulmonary Auscultation Examination," European Journal of Pediatrics, vol. 178, no. 6, pp. 883-890, June 2019.

[4] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, et al., "An Open Access Database for the Evaluation of Respiratory Sound Classification Algorithms," Physiological Measurement, vol. 40, no. 3, article no. 035001, March 2019.

[5] Y. Feng, Y. Wang, C. Zeng, and H. Mao, "Artificial Intelligence and Machine Learning in Chronic Airway Diseases: Focus on Asthma and Chronic Obstructive Pulmonary Disease," International Journal of Medical Sciences, vol. 18, no. 13, pp. 2871-2889, 2021.

[6] Y. Shi, Y. Li, M. Cai, and X. D. Zhang, "A Lung Sound Category Recognition Method Based on Wavelet Decomposition and BP Neural Network," International Journal of Biological Sciences, vol. 15, no. 1, pp. 195-207, 2019.

[7] G. Petmezas, G. A. Cheimariotis, L. Stefanopoulos, B. Rocha, R. P. Paiva, A. K. Katsaggelos, et al., "Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function," Sensors, vol. 22, no. 3, February 2022.

[8] J. Acharya and A. Basu, "Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning," IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 3, pp. 535-544, June 2020.

[9] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A Lightweight CNN Model for Detecting Respiratory Diseases from Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 7, pp. 2595-2603, July 2021.

[10] B. Acar Demirci, Y. Koçyiğit, D. Kizilirmak, and Y. Havlucu, "Adventitious and Normal Respiratory Sound Analysis with Machine Learning Methods," Celal Bayar University Journal of Science, vol. 18, no. 2, pp. 169-180, 2022.

[11] A. Gökçen, "Computer-Aided Diagnosis System for Chronic Obstructive Pulmonary Disease Using Empirical Wavelet Transform on Auscultation Sounds," The Computer Journal, vol. 64, no. 11, pp. 1775-1783, November 2021.

[12] N. S. Haider and A. K. Behera, "Computerized Lung Sound Based Classification of Asthma and Chronic Obstructive Pulmonary Disease (COPD)," Biocybernetics and Biomedical Engineering, vol. 42, no. 1, pp. 42-59, January-March 2022.

[13] N. Asatani, T. Kamiya, S. Mabu, and S. Kido, "Classification of Respiratory Sounds Using Improved Convolutional Recurrent Neural Network," Computers & Electrical Engineering," vol. 94, article no. 107367, September 2021.
[14] A. Rizal and A. Puspitasari, "Lung Sound Classification Using Wavelet Transform and Entropy to Detect Lung Abnormality," Serbian Journal of Electrical Engineering, vol. 19, no. 1, pp. 79-98, 2022.
[15] S. Gupta, M. Agrawal, and D. Deepak, "Gammatonegram Based Triple Classification of Lung Sounds Using Deep Convolutional Neural Network with Transfer Learning," Biomedical Signal Processing and Control, vol. 70, article no. 102947, September 2021.
[16] M. Mathe, P. Mididoddi, and B. T. Krishna, "Artifact Removal Methods in EEG Recordings: A Review," Proceedings Engineering and Technology Innovation, vol. 20, pp. 35-56, January 2022.
[17] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in Audio Signal Feature Extraction Methods," Applied Acoustics, vol. 158, article no. 107020, January 2020.
[18] M. M. Zarandah, S. M. Daud, and S. S. Abu-Naser, "A Systematic Literature Review of Machine and Deep Learning-Based Detection and Classification Methods for Diseases Related to the Respiratory System," Journal of Theoretical and Applied Information Technology, vol. 101, no. 4, pp. 1273-1296, February 2023.
[19] T. A. Assegie and Y. B. Chekol, "The Performance of Machine Learning for Chronic Kidney Disease Diagnosis," Emerging Science Innovation, vol. 1, pp. 33-40, September 2023.
[20] A. Hussain, H. Malik, and M. U. Chaudhry, "Supervised Learning Based Classification of Cardiovascular Diseases," Proceedings of Engineering and Technology Innovation, vol. 20, pp. 24-34, January 2022.
[21] A. Elsetrønning, A. Rasheed, J. Bekker, and O. San, "On the Effectiveness of Signal Decomposition, Feature Extraction and Selection on Lung Sound Classification," https://doi.org/10.48550/arXiv.2012.11759, December 22, 2020.
[22] S. Y. Jung, C. H. Liao, Y. S. Wu, S. M. Yuan, and C. T. Sun, "Efficiently Classifying Lung Sounds through Depthwise Separable CNN Models with Fused STFT and MFCC Features," Diagnostics, vol. 11, no. 4, article no. 732, April 2021.
[23] S. Rani, A. Chaurasia, M. K. Dutta, V. Myska, and R. Burget, "Machine Learning Approach for Automatic Lungs Sound Diagnosis from Pulmonary Signals," 44th International Conference on Telecommunications and Signal Processing (TSP), pp. 366-371, July 2021.
[24] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "A Neural Network-Based Method for Respiratory Sound Analysis and Lung Disease Detection," Applied Sciences, vol. 12, no. 8, article no. 3877, April 2022.
[25] Z. Neili and K. Sundaraj, "Gammatonegram Based Pulmonary Pathologies Classification Using Convolutional Neural Networks," 19th International Multi-Conference on Systems, Signals & Devices (SSD), pp. 1112-1118, May 2022.
[26] S. A. F. Jaffery, S. Aziz, M. U. Khan, S. Z. H. Naqvi, M. Faraz, and A. Usman, "An Automated System for the Classification of Bronchiolitis and Bronchiectasis Diseases Using Lung Sound Analysis," International Conference on Robotics and Automation in Industry, pp. 1-6, March 2023.