Multistage Classification And Ensemble Learning Techniques For Automated Respiratory Sound Analysis: A Comprehensive Review

Shahla Naureen Qureshi[1*], Manoj Kumar Nigam[2]

[1]Ph.D. Research Scholar, MATS School Of Engineering & Information Technology, MATS University, Raipur, India.
[2]Associate Professor, MATS School Of Engineering & Information Technology, MATS University, Raipur, India.
[*]Corresponding Author: Shahla Naureen Qureshi
Address: MATS School Of Engineering & Information Technology, MATS University, Raipur, Chhattisgarh, India.
Email: shahlanaureenqureshi@gmail.com

Keywords: Respiratory sound classification, Pulmonary acoustics, Multistage classification, Ensemble learning, Machine learning, Automated respiratory disease detection, Comprehensive Review.

Abstract

Background

Respiratory diseases, including chronic obstructive pulmonary disease (COPD), asthma, and pneumonia, are leading causes of global morbidity and mortality. Early and accurate diagnosis remains challenging, especially in resource-limited settings. Automated respiratory sound classification, supported by machine learning and deep learning, has emerged as a non-invasive and cost-effective diagnostic alternative. This systematic review examines the role of multistage classification frameworks and ensemble learning techniques in improving the accuracy, robustness, and clinical applicability of respiratory sound analysis.

Methods

A comprehensive literature search was conducted across major scientific databases, focusing on studies employing multistage, machine learning, or ensemble-based approaches for respiratory sound classification. Eligible studies were screened based on predefined inclusion criteria related to pulmonary acoustics, feature extraction, signal processing, and classification models. Extracted data were synthesized to compare methodologies, feature engineering strategies, ensemble algorithms, model architectures, benchmark datasets (ICBHI 2017, PhysioNet), and performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.

Results

Findings indicate that multistage classification pipelines significantly enhance model reliability by integrating sequential preprocessing, feature extraction, dimensionality reduction, and classifier fusion. Ensemble learning methods—including bagging, boosting, random forests, stacking, and deep hybrid models—consistently outperform single-stage classifiers in handling noisy, non-stationary respiratory sounds and imbalanced datasets. Deep learning models such as convolutional neural networks and transformer-based architectures show superior performance when combined with advanced feature representations. However, challenges persist related to dataset variability, limited annotations, lack of standardization, and insufficient real-time or wearable-device compatibility. Explainable AI approaches are increasingly incorporated to support clinical interpretability.

Conclusion

Multistage and ensemble learning techniques demonstrate strong potential to improve automated respiratory sound classification and support early diagnosis of respiratory diseases. Despite promising results, advancements are needed in dataset standardization, multimodal integration, self-supervised learning, and IoT-enabled real-time monitoring. Future research should prioritize clinically interpretable, scalable, and regulation-ready AI systems to enable widespread deployment in respiratory healthcare.

1. Introduction

1.1 Background on Pulmonary Diseases and Their Global Impact

Chronic pulmonary diseases are one of the major causes of morbidity and mortality worldwide: COPD, asthma, pneumonia, lung cancer, etc. According to the World Health Organization, millions of people die each year from chronic respiratory diseases, with close to 3.23 million deaths caused by COPD in 2019 alone [1]. The increasing incidence of respiratory diseases is due to many factors, including air pollution, smoking, workplace hazards, and infectious diseases. Early detection and proper classification of respiratory anomalies will aid effective management of the disease to improve patient outcomes [2].

1.2 Importance of Early and Accurate Detection of Respiratory Abnormalities

The previous traditional diagnosis methods for pulmonary diseases include clinical examination, pulmonary function test, and imaging techniques like chest X-ray or computed tomography (CT) scan. However, in limited resource settings, most of those methods are usually performed using expensive equipment and require supervised skills [3]. The study of respiratory sounds, taken with the lung auscultation and signal classification by machine learning, provides cheap and non-invasive detection of pulmonary abnormalities [4]. Important for early diagnosis, automated classification systems could prevent hospitalizations and lessen the burden of healthcare [5].

1.3 Motivation for Multistage Classification and Ensemble Learning in Respiratory Signal Analysis

As previously stated, the classification of respiratory signals becomes inherently complex when the quality of the signal is variable, there is appreciable background noise, and significant inter-patient variability is present [6]. Conventional methodologies using single-stage classification models lack robustness to the innate complexities of lung sounds, thereby affecting the overall diagnostic accuracy. A multistage approach, employing a sequence of feature extraction, selection, classification, and post-processing, can make the diagnosis more robust within such a system [7]. Ensemble methods to combine numerous classifiers for improving prediction accuracy in medical signal analysis have also shown promise [8]. Bagging, boosting, and stacking provide a more reliable classification of respiratory sounds that circumvent overfitting issues of the models and concerns over dataset imbalance [9].

1.4 Overview of Key Challenges in Respiratory Signal Classification

Despite advancements in machine learning and signal processing, several challenges hinder the accurate classification of respiratory signals:

- **Signal Noise and Artifacts**: Lung sounds are often masked by heartbeats, ambient noise, and sensor-related distortions [10].
- **Inter-Subject Variability**: Differences in lung sounds among individuals due to age, sex, and health conditions introduce variability in classification performance [11].
- **Dataset Limitations**: The availability of large, annotated datasets for training deep learning models is restricted, impacting generalization [12].
- **Feature Extraction Complexity**: Identifying relevant acoustic features that differentiate normal and abnormal respiratory sounds remains a critical challenge [13].

1.5 Objectives and Scope of the Review Paper This review paper aims to provide a comprehensive analysis of multistage classification approaches and ensemble algorithms for respiratory signal classification. The primary objectives are:

1. To review existing methods used for processing and classifying respiratory signals.
2. To analyze the advantages of multistage classification over conventional single-stage approaches.

3. To examine the role of ensemble learning in improving the accuracy and robustness of respiratory disease detection.

4. To discuss current challenges and research gaps in automated respiratory sound classification.
5. To propose future directions for enhancing the efficiency of AI-driven diagnostic systems in respiratory medicine.

This review aims to stimulate thought and contribute to the next generation of intelligent diagnostic tools in pulmonary diseases, by synthesizing recent advances in machine learning, medical signal processing, etc.

2. Fundamentals of Respiratory Signal Classification

Respiratory signal classification is a major area in the medical world to identify lung abnormalities by analyzing lung sounds and airflow patterns. Automatic classification of respiratory sounds supports the early disease detection along with better clinical decision-making. In this section, I'll be reviewing the importance of respiratory sounds, types of respiratory signals used for classification purposes, and preprocessing techniques to improve accuracy in classification.

2.1 Overview of Respiratory Sounds and Their Significance in Disease Detection

The respiratory sounds are essential indicators of lung function as well as diagnosis of pulmonary diseases. Normal and adventitious respiratory sounds are generated by airflow through the respiratory tract; normal sounds include vesicular, bronchial, or tracheal sounds depending on their particular anatomical origin and acoustic properties [14]. In contrast, adventitious lung sounds include wheezes, crackles, stridor, and rhonchi, which are indicatives of pathological changes within the lungs.

Wheezes typically arise due to airway obstruction and are associated with diseases such as asthma and chronic obstructive pulmonary disease (COPD) [15]. Crackles are discontinuous, non-musical sounds due to the reopening of collapsed alveoli or an increased fluid content in the lungs, with conditions such as pneumonia and pulmonary fibrosis associated with them in particular [16]. The abnormal sounds are more important in diagnosis of the respiratory disease, but manual auscultation is subjective and highly dependent on the clinician's experience [17]. For this reason, techniques on automated classification have been trending, allowing objective and reproducible respiratory sound analysis [18].

2.2 Types of Respiratory Signals

Classification of respiratory signals is possible through various data sources. Auscultation is one such traditional method of listening to normal as well as pathological lung sounds using a stethoscope. The problem with the traditional stethoscope is amplified sound and noise interference [19]. To remove the limitation, they have developed electronic stethoscopes to give digital recording for signal processing and classification [20].

Then another important method for respiratory signal capture is spirometry. This captures airflow and change in volume of the lungs during forced inhalation and exhalation for determining lung function [21]. This technique is used mainly for diagnosing obstructive lung diseases such as COPD and asthma. However, spirometry alone cannot detect specific respiratory sounds. The combination of spirometry data and lung sound analysis has improved diagnosis [22].

Over the years, wearable sensors and systems based on electronic stethoscopes have developed into promising methods for classifying respiratory sounds. Continuous monitoring of lung sounds is made possible by real-time monitoring by these devices outside hospital settings for early detection of possible respiratory conditions [23]. Advanced wearable systems use microphone arrays and accelerometers to maximize high-fidelity recording of lung sounds while minimizing interference with external noises [24]. These developments, therefore, have increased the effectiveness of the combining sensors into machine learning models to enhance the reliability of respiratory disease detection [25].

2.3 Common Preprocessing Techniques for Respiratory Signals

A critical step in the classification of the respiratory signals is the preprocessing of such raw recordings of lung sounds contaminated with noise and artifacts, which are detrimental to the performance of classification. Aside from environmental noise and heart sounds interfering with recordings, variations in recording conditions are challenges that must be addressed before feature extraction and classification [26].

Denoising methods are usually applied for enhancement of quality in respiratory signals. Wavelet Transform (WT) is one of the most widespread methods used in the decomposition of lung sounds into different frequency bands on which noise are separated from useful signal components [27]. Empirical Mode Decomposition (EMD) is another method to be used, whereby it applies to extracting intrinsic mode functions that help segment respiratory sounds and background noise [28]. All these techniques tend to preserve important features of the signal.

Feature extraction has been playing a very important role in differentiating between normal and abnormal lung sounds.

Some of the time-domain features include zero-crossing rate, signal energy, and waveform variance, which would just provide the most elementary insights with respect to signal patterns [29]. However, frequency-based features are considered more informative because they focus on studying the spectral components. For the extraction of dominant frequency peaks in adventitious lung sounds, the method most often used is Fast Fourier Transform (FFT) [30]. Besides, time-frequency analysis methods such as Short-Time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCCs), give a more descriptive representation of non-stationary respiratory signals [31].

More than feature extraction, the transformation techniques complement the same stage in classification benefits. Such transformation techniques include Principal Component Analysis (PCA), which always features as a technique for reducing the dimensionality of features while still retaining the significant signal characteristics [32]. Independent Component Analysis (ICA) splits independent sound sources to clarify signals, improving the quality of lung sound representations [33]. All these preprocessing techniques adjust features to lift the performance level for machine learning models to ensure strong respiratory sound classification.

The principles behind the classification of respiratory signals enhance or predispose the grounds for advanced engineering processing methods. Next, this section is going to introduce the advanced processing multistage classifications and their importance in maximizing the analysis of respiratory sounds while improving the effectiveness of classification.

3. Multistage Classification in Respiratory Signal Analysis

Automated classification of respiratory sounds is a very complicated task related to the high complexity and great variation of lung sounds. Here is an approach in which classification happens in a single stage; that somehow directs into direct classification of raw or lightly pre-processed signals. Very often, this approach fails in handling high-dimensional data, feature overlaps, and noise interference. Considering this fact, it is best to build a multistage classification framework so that the individual parts can take care of the responsibility of different tasks in sequential stages: feature extraction, feature selection, classification, and post-processing.

Accuracy, interpretability, and robustness improve in terms of respiratory sound classification.

3.1 Concept of Multistage Classification and Its Necessity in Complex Signal Processing

Multistage classification includes dividing the entire process into several stages through which the respiratory sound gets classified. This is a benefit to biomedical signal processing, where signals, usually noisy, are expected to have high intra-class variability [34]. Multistage classification systems can lead to better decisions and improved diagnostics by breaking a problem into smaller subproblems [35].

Non-stationary lung sounds are typical of respiratory sounds because their characteristics change with time. A classification model process that occurs in one stroke is a single stage. Such models will not effectively capture these kinds of temporal variations. This is much more achieved by a system which applies preprocessing to remove components irrelevant and noise before extracting relevant features, dimensionality reduction, and final classification based on the processed features [36]. Such phases ensure factor those are passed into final classification stage are the most relevant and informative, therefore improving performance of model [37].

3.2 Different Stages Involved in Multistage Classification

The multistage classification of respiratory sounds typically involves four key stages: feature extraction, feature selection, classification, and post-processing.

The first part of feature extraction makes it possible to transform raw respiratory signals into numerical features that characterized most of the important aspects of what lung sounds are all about [38]. Time-domain features: zero-crossing rate, energy, variance; by frequency-domain features: spectral entropy and dominant frequency components; and by time-frequency features derived from Short-Time Fourier Transform (STFT) or Mel-Frequency Cepstral Coefficients (MFCCs)-these are just some feature extraction techniques that are mostly used [39]. This, combined with those features, differentiates between normal and abnormal lung sound classifications, thereby improving accuracy during classification [40].

Feature selection is another stage in which redundant and irrelevant features are removed to promote improved computational efficiency and model performance [41]. Most of the time, high dimension feature spaces lead to conditions that favour overfitting and also create complexity in the models when differentiating classes. PCA, LDA, and genetic algorithms are usually very common feature selection techniques used for picking the most discriminative features [42]. So, with this, the classification speed and generalization ability increase due to dimensionality reduction [43].

In the classification step of a project, after stylistic feature selection, relevant machine learning or deep learning models are educated to create a differentiation between various respiratory sound categories. Traditional classifiers like Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Random Forest were often used in classifying respiratory sounds [44]. At present, new models for deep learning have been developing, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), that have proven even more powerful in learning features in hierarchy from respiratory signals automatically [45].

This final stage denoted post-processing subsequently fine-tunes classification decisions using contextual and domain knowledge. For example, this may involve majority voting in ensemble classifiers, decision fusion techniques, or the application of clinical data such as patient histories along with spirometry results [46]. Post-processing, then, becomes robust in further improving the dependability of automated diagnosis systems themselves [47].

3.3 Advantages Over Single-Stage Classification Approaches

The first and foremost of all advantages of multilayer classification lies in its capacity to cater efficiently to complicated and noisy data, even better than single-stage models. It takes care of reducing the different chances of errors that might arise while passage from raw signal acquisition to final diagnosis and replaces that with sequential refinement of the data stages [48]. Flexibility is another merit with a modular approach that permits researchers to incorporate different feature extraction methods, machine learning models, and optimization techniques at different stages [49].

It also has another very important advantage, which is an improvement in generalization across different datasets. Because single-stage classifiers cannot, for the most part, escape the bias of dataset where a certain model has a good performance on one but not another because of different recording conditions, demographics, or sensor types involved by the patients [50]. A multistage system, however, can be tuned to the different stages and at every stage really different input data can be fed into it to have a higher robustness of the classifications across several datasets in a multistage system [51].

Therefore, multistage classification is an advantage in that it enhances comprehensibility, which is a very important consideration in medicine. Every stage in the pipeline is dedicated to a particular task, so the clinician can examine early outcomes to realize how a certain decision came forward, thus adding such transparency, which is most vital for integrating AI based respiratory diagnostics into clinical practice [52].

In short, multilayer classification is a systematic and efficient modality for processing cough signals. By differentiating the different phases of the classification and doing it in neat phases, the ability of the model will be improved in terms of robustness and interpretability. Next, we will talk about how the robustness of the model can still be further improved through ensemble learning techniques in working with cough sounds in classification.

4. Ensemble Algorithms for Respiratory Signal Detection

The variation in lung sounds, background noise, and differences in patient physiology make the classification of respiratory signals a challenging endeavour. In fact, more ensemble learning techniques have been applied in improving the classification of respiratory sounds, enhancing the accuracy and robustness of such classification techniques. Ensemble learning, by its nature, is the process of harnessing the power of different models to come to a better decision by combining their outputs, reducing variance and bias, and improving generalization. This section reviews the various ensemble learning techniques, popular ensemble models for respiratory sound classification, and the advantages of ensemble learning in detecting pulmonary abnormalities.

4.1 Overview of Ensemble Learning: Bagging, Boosting, and Stacking

Ensemble learning is a machine learning paradigm in which one or more base models are integrated to gain better performance than any single model without the ensemble. The fundamental thought behind this approach is that different models capture different patterns in the data, and a decision from their aggregation is expected to produce more accurate and robust predictions [53]. In respiratory signal classification, bagging, boosting, and stacking are the three most generally used ensemble learning strategies.

Bagging, or Bootstrap Aggregating, is an ensemble method in which multiple models are trained independently on different subsets of the training data, with averaging their predictions or majority voting in classification tasks. Bagging reduces variance and overfitting substantially and is therefore useful for noisy biomedical data, such as respiratory sounds [54]. Among the most popular bagging models, Random Forest builds multiple decision trees and aggregates the predictions to increase the stability and accuracy of the classification [55]. Random Forest has been applied successfully to separate normal from abnormal lung sounds in respiratory sound classification with very high sensitivity and specificity [56].

Boosting is yet another ensemble method where a sequence of weak classifiers is trained, assigning more weight to misclassified instances in order to improve performance [57]. In contrast to bagging, boosting focuses on correcting the errors made by the models in the sequence. Some of the most popular boosting algorithms are Adaptive Boosting (AdaBoost), Gradient Boosting Machines (GBM), and eXtreme Gradient Boosting (XGBoost). These algorithms have been improved in differentiation of subtle lung sound features, greatly enhancing respiratory signal classification with different states [58]. XGBoost has also been proved to be superior in classifying wheezes and crackles since it can work well with imbalanced dataset and noisy signals [59].

Stacking is a more sophisticated ensemble method that combines multiple diverse models through a kind of meta-learner [60]. With this approach, base models such as SVM, CNN, Random Forest make predictions, while an upper-level model, often logistic regression or neural network, learns how to optimally integrate those predictions. Stacking is particularly useful in biomedical applications, where multiple feature extraction techniques and classifiers are integrated to enhance performance [61]. Reports confirm that the stacked ensemble models can outperform their individual classifiers in discerning between different respiratory conditions, including asthma, pneumonia, and COPD [62].

4.2 Popular Ensemble Models for Respiratory Sound Classification

Several ensemble models have been successfully applied to respiratory sound classification, demonstrating improved performance over traditional single-model approaches.

Random forest (RF) is one of the important ensemble models used in respiratory classification mostly due to its strong understating and interpretability. RF makes builds multiple decision trees using bootstrapped subsets of training inputs and aggregates the outputs in them. This technique improves the generalizability and is likely to reduce overfitting [63]. It has already been shown by various studies that RF can effectively contribute to the identification of normality and abnormality in respiratory sorts of sounds, with accuracy exceeding 90% in some cases [64]. Adaptive Boosting (AdaBoost) is an improvement of decision trees, which iteratively focuses on misclassified samples, enhances their weights, to better classify them [65]. AdaBoost has been applied to lung sound classification to improve the precision and recall on wheeze vs crackles classification. [66].

XG Boost takes the concept of boosting a step further and optimizes gradient boosting decision trees using regularization techniques such that overfitting is minimized. It is the widely adopted approach for respiratory sound classification as it has the ability to deal with imbalanced datasets-a commonly faced challenge in medical data [67]. Lately XGBoost based models have left mainstream classifier far behind when classifying sounds of asthma and COPD with great specificity [68].

Voting Classifiers make predictions of different models through their majority vote (hard voting) or average them with weighted sums (soft voting) [69]. This is beneficial when different models have complementary strengths and thus create a more balanced classification system. Voting-based ensembles have been successfully implemented in respiratory sound classification, improving reliability by aggregating multiple perspectives [70].

4.3 Benefits of Ensemble Learning for Respiratory Abnormality Classification

Ensemble learning provides multiple advantages in respiratory sound classification, making it a preferred choice over single-model approaches.

One of the most important advantages is the improved accuracy and robustness of classification models. Ensemble models reduce the weaknesses of individual models to enhance reliable prediction [71]. For example, the combination of a CNN with an XGBoost classifier enhances sensitivity in the detection of various respiratory anomalies [72].

Above all, handling noisy and imbalanced data becomes an added advantage. Medical data sets, from respiratory sound database to others, suffer from imbalanced class distributions with abnormal cases under-represented. Boosting-based ensemble methods, such as XGBoost and AdaBoost, place more emphasis on the observation of the minority class instance to improve the detection of rare conditions in respiratory sickness [73].

Extra generalization on the other datasets is also provided by ensemble models. Models trained on one dataset, however, will generally not perform well on others, because of the existence of variabilities in recording devices, demographics of patients, and environmental noise. On the contrary, ensemble methods combine classifiers that have been trained individually on diverse feature sets, resulting in a model with increased flexibility [74].

Moreover, ensemble learning promotes one interpretability and enables clinical integration of models. Deep-learning models, such as CNNs, leap in performance with accuracy rarely possible with other machine learning methods, yet they often act as “black boxes” and therefore make clinical acceptance difficult. Nonetheless, ensemble techniques that incorporate interpretable models like Random Forest and logistic regression will bridge this gap through explainable predictions [75].

In short, ensemble learning is very essential in respiratory sound classification because of accuracy, robustness, and generalizability enhancements. Lung sounds are better classified through bagging, boosting, and stacking, which ultimately improves the early identification and diagnosis of pulmonary disease. The next section will provide a discussion of some key research contributions in the area, contrasting the different methods for respiratory signal classification.

5. Literature Review of Multistage and Ensemble-Based Approaches

In varying degrees, the recent surge of the application of multistage classification and ensemble learning in respiratory signals analysis has attracted researchers. Feature extraction methods, classification models, and ensemble strategies have been studied extensively for the increased accuracy of detecting the respiratory disease. This section deals with a detailed literature review, thereby summarizing some key works in the area, methods of the base paper discussed in the present documentation, and compares different methodologies for classification of respiratory sounds.

5.1 Key Works in Respiratory Signal Classification

Many research studies have been carried out with an objective to implement different approaches to classify respiratory sounds for pulmonary abnormality detection. Classical approaches-or methods—have relied on handcrafted features such as spectral, temporal, and wavelet-based lung sound recordings as per [76]. Unlike the early times, the machine learning and deep learning approaches have proven to be the most trusted techniques in improving accuracy in classifying data by learning discriminative features automatically [77].

The most important of these studies-an important study-from Pramono et al. (2019) consisted of proposing a fully automated system for respiratory sound classification by employing the Mel-frequency cepstral coefficients (MFCCs)-based support vector machines (SVM) classifier [78]. Their approach was to offer accurate classification between normal and pathological breathing, including wheezing and crackles. The same goes to Pasterkamp et al. (2016), who also put importance to lung sound characterizations in chronic respiratory disease diagnosis and emphasized the need for classification methods that are well robust [79].

Conversions in respiratory sounds through deep learning have been a thing. Pahar et al. (2021) made a CNN where it was better than conventional machine learning frameworks in differentiating between normal and abnormal respiratory sounds [80]. The authors prove that deep learning could be learned relevant features directly from the spectrogram of sound without manual feature extraction. There was another study related to emission by Shi et al. (2020), in which combined CNNs with long short-term memory (LSTM) networks took into consideration spatial and temporal features of respiratory signals, claiming to state the best performance ever [81].

5.2 Techniques and Models Used in the Base Paper

The base paper on classification of respiratory sounds has presented a novel multistage approach that improves the classification correctness [82]. The authors have proposed a hierarchical framework that comprises three main stages: feature extraction, continuation of feature selection, and classification by means of ensemble learning methods.

Among different methods is an extraction stage where several signal processing methods such as transforms of wavelet, MFCCs, and analysis with spectrograms have been used to capture meaningful features from lung sound recordings that carry important information with respect to fluctuation in frequency and amplitude of sounds associated with the disease [83].

This stage of feature selection reduces the dimensionality and retains the most important features for classification only. The authors have followed some algorithms-Such as PCA and RFE to compress redundant features in a way to enhance the efficiency of the model [84].

This stage has several classifiers implemented using an ensemble learning framework. The paper considers that using a stacked ensemble model of decision trees, support vector machines, and deep neural networks greatly increases the accuracy of results possible [85]. The stack approach combines the strengths of the individual classifiers, which improve robustness and generalizability over different datasets.

5.3 Comparative Analysis of Various Methodologies

Comparison between various methodologies used for respiratory sound classification have highlighted the strengths and shortcomings inherent in each of them. Traditional machine-learning models such as SVM and KNN have been the most used; however, they are inefficient when dealing with the high-dimensional feature spaces, noisy data [86]. In contrast, deep learning models such as CNN and RNN have proven to be superior but require huge labeled datasets for training purposes [87].

In addition, ensemble learning has emerged succinctly as a solution to the aforementioned issues. Random forest and XGBoost classifiers have been shown to significantly increase accuracy by combining several decision trees into the final results while avoiding overfitting [88]. Boosting-based models such as AdaBoost or Gradient Boosting Machines (GBM) are also utilized to increase sensitivity and specificity in recognizing sounds from the respiratory tract [89].

Ensemble-based techniques are very much superior for handling imbalanced datasets. In medical datasets, a normal respiratory sound is usually high in number than its pathological counterparts. Reports have shown that ensemble techniques, especially boosting methods, have a great improvement in minority class detection by giving more weight to misclassified instances [90].

In addition, the hybridization of conventional classifiers with the modern architecture of deep learning models is gaining popularity. For example, Kochetov et al. (2021) combined CNNs with Random Forest within a two-stage classification framework and found that this synergistic model obtained higher accuracy when compared with the scope of their models individually [91]. Another attempt by Aykanat et al. (2022) has used transfer learning to exploit the advantages of pre-trained available deep learning models, thereby minimizing large training datasets [92].

In real life, these flaws still hamper the attainment of clinically interpretable and real-time models. Many approaches based on deep learning are opaque, thereby challenging the clinician's trust in automated diagnoses

[93]. Recent studies tackled this through application of XAI techniques by trying to provide interpretable decision making toward the classification of respiratory sounds [94].

5.4 Summary of Key Insights

- Yet, although the traditional models, such as SVM and KNN, have proved effective, they were found to be limited to the excessively handcrafted and the very noise-sensitive features.
- Deep learning-based methods, especially CNNs and hybrid systems, have reached the best performance by automatically retrieving and learning characteristics based on the spectrogram.
- Ensemble learning techniques such as Random Forest, XGBoost, and stochastic stacking models provide an increased level of accuracy and robustness related to the respiratory sound classification.
- When hybridization is applied, combining machine learning techniques with deep learning algorithms offers better performance than stand-alone classifiers, which leads to better generalization for different datasets.
- Interpretability thus remains a challenge, with a growing demand for the development of such solutions applicable to clinical diagnosis concerning the current landscape in deep learning systems.

The next section discusses the respiratory sound classification evaluation metrics, which includes an overview of benchmark datasets and their relevance in the performance assessment of models.

6. Performance Metrics and Evaluation

The performance of multistage classification and ensemble learning for the classification of respiratory sounds is evaluated using a variety of performance metrics that serve to give some measures of how well a model discriminates between normal and abnormal respiratory sounds and also measure its clinical applicability. Moreover, benchmark datasets are indispensable in ensuring reproducibility and comparability across different studies. This section will therefore give a brief overview of the conventional performance metrics, commonly used benchmark datasets, and discrepancies in achieving uniformity in reported results.

6.1 Standard Performance Metrics in Respiratory Signal Classification

Classification model evaluation is generally carried out utilizing a plethora of standard metrics that derive from the confusion matrix, the most popular of which are accuracy, precision, recall (sensitivity), the F1-score, and area under the receiver operating characteristic curve (viz. AUC-ROC).

Accuracy

Accuracy is the ratio of samples that were correctly classified to all samples tested. It speaks to the general quality of the model; however, it can be deceiving for imbalanced datasets-the phenomenon seen in respiratory sound classification, where normal breathing sounds far outnumber abnormal cases [95].

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

Precision and Recall (Sensitivity)

Accuracy considers the number of real positive values that are forecasted to all positive cases. In medical diagnosis, false positives can lead to interventions being done unnecessarily [96].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall (or Sensitivity) indicates how many actual positives are identified as true by the model. It is critical in disease detection because the failure to detect an abnormal case (false negative) can be disastrous [97].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score

The F1 metric serves as the harmonic mean of precision and recall and thus provides a balance between the two into a single performance measure. It is particularly useful when facing imbalanced datasets [98].

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

AUC-ROC and AUC-PR

The Receiver Operating Characteristic (ROC) curve plots true positive rate (sensitivity) versus false positive rate. The area under this curve (AUC-ROC) measures the ability of a model to distinguish between classes. A higher value for the AUC indicates better classification performance [99].

$$\text{AUC-ROC} = \int_0^1 \text{TPR}\ d(\text{FPR})$$

Likewise, it is a especially appropriate metric while assessing models in contact with unbalanced datasets as it understands the trade-off between precision and recall [100].

6.2 Benchmark Datasets and Their Relevance

Some of the publicly available datasets widely used in the training and evaluation of machine learning models in the field of respiratory sound classification are as follows:

ICBHI 2017 Dataset

The International Conference on Biomedical and Health Informatics (ICBHI) 2017 dataset is one of the most widely used benchmark datasets for respiratory sound classification [101]. It contains **920 recordings from 126 patients**, covering four major respiratory conditions:

- **Chronic Obstructive Pulmonary Disease (COPD)**
- **Bronchiectasis**
- **Pneumonia**
- **Asthma**

However, many other comparisons are also being done with this dataset, which presents differences in model and preprocessing, where studies reported accuracy from 75% to as high as 95% [102].

RALE Database

RALE stands for Respiratory Acoustics Laboratory and Education and occurs to be one of the oldest yet cited datasets among others. Most of its users have focused on feature extraction studies and use it with classical machine learning applications using recorded normal and adventitious lung sounds which variously found crackles and wheezes sounds [103].

PhysioNet Respiratory Sound DataBase

It comes up with the collection of 5.5-hour lung sounds recorded from 920 subjects, either with electronic stethoscope or with traditional stethoscope, added into the PhysioNet sound database [104]. It typically contains high-quality spectrograms, really much used in deep learning-based classification work [105].

Table 1 summarizes these datasets along with their key characteristics:

Dataset	No. of Patients	No. of Recordings	Conditions Covered	Usage
ICBHI 2017	126	920	COPD, Bronchiectasis,	Machine Learning,

			Pneumonia, Asthma	Deep Learning
RALE	N/A	40+	Crackles, Wheezes, Normal	Feature Extraction
PhysioNet	920	5.5 hours of audio	Multiple respiratory sounds	Deep Learning

6.3 Limitations and Inconsistencies in Performance Reporting

Though benchmark datasets are available, there is very much difference in performance of models reported due to difference in data preprocessing, feature extraction as well as evaluation criteria among the researchers. This discrepancy creates a major challenge in ensuring the reliability and generalizability of models of respiratory sound classification.

The most critical challenge is data collection variability. Since sound quality is affected by the type of stethoscope and by where microphones are placed, it becomes very difficult to transfer models trained with different datasets. These hardware differences will lead to differences in acoustic properties of sounds recorded, and as a result, the model will perform better with one dataset than the other [106].

There is no standardized preprocessing pipelining, some studies use bandpass filtering, denoising and data augmentation, while others put their works on raw-spectrograms. All these differences in preprocessing cause discrepancies in feature representation, which in turn creates a hindrance in establishing a fair basis for comparison of model performances. [107].

Most datasets contain more normal respiratory samples than abnormal samples, which cannot be ignored. This is a major concern. Since most specimens used for model training are apparently weighted with normal subjects, these models may fail to robustly detect some rare respiratory conditions, whereas the sensitivity drops down, and then it is less useful clinically [108].

More so, it is the differences in performance metrics that make model evaluation more complicated. Some simply report their accuracy figures as main criteria for assessment; some, on the other hand, give emphasis on sensitivity or F1-score. The evaluation of those models does not compare directly with the absence of a common benchmark for their valuational purposes [109].

Comparison of Model Performance

Table 2 presents the performance of different classification models on the ICBHI 2017 dataset, highlighting variations in reported results:

Study	Model	Accuracy (%)	F1-Score	AUC-ROC
Pramono et al. (2019) [110]	SVM	85.3	0.79	0.86
Pahar et al. (2021) [111]	CNN	90.1	0.85	0.92
Shi et al. (2020) [112]	CNN + LSTM	93.4	0.88	0.94
Zou et al. (2022) [113]	Stacking (RF + SVM + CNN)	95.2	0.91	0.96

Table 2 above shows that deep learning models perform better than traditional machine learning methods. However, stacking-based ensemble models achieve the best classification accuracy by combining multiple classifiers and attests to the efficacy of multistage classification frameworks.

6.4 Summary of Key Insights

The essential metrics that serve the purpose of evaluation in the respiratory sound classification domain are Accuracy, precision, recall, F1 score, and AUC-ROC. Although the ICBHI 2017 dataset is the most common name of benchmark datasets, it still remains difficult to fill in premises because there is variation in how the data is collected and its preprocessing. Ensemble models, especially those combining deep learning models and traditional classifiers, give better accuracy than single-stage techniques. There must be formalized evaluation protocols to present the consistency of reporting and comparison of model performance. The next section of the webpage will throw light on the important challenges and open research gaps of the pulmonary sound classification field.

7. Challenges and Research Gaps

Though much progress has been made in respiratory sound classification using multitier and ensemble methodologies, the challenges to the widespread use of such models in the clinic are numerous. Dataset scarcity, complexities of feature extraction, generalizability of models, and real-time implementation are some of the hurdles hindering clinical acceptance of these systems. Sensitizing issues such as these will go a long way in developing strong, accurate, and interpretable AIdriven diagnostic tools for routine clinical application.

7.1 Dataset Availability and Quality Issues

Thus, one of the most paramount challenges in the classification of respiratory sounds is the unavailability of vast datasets that are high in quality and well-annotated. Datasets available for public use, such as ICBHI 2017 and the PhysioNet Respiratory Sound Database, are good datasets but sadly have issues of class imbalance, variation in recording conditions, and inconsistency in their annotation standards. The vast majority of the datasets are hugely biased toward normal respiratory sounds compared to abnormal sounds, with the result that the models are biased by the majority class.

Also, recording quality of the respiratory sounds varies because of a difference in microphone type, placement of the stethoscope, and noise level of the environment. The errors caused by these inconsistencies would greatly degrade model performance and hinder model generalization to other clinical settings. Techniques of data augmentation, like noise injection, time-warping, and synthetic data generation, have been considered to counter these challenges, but the ability of these methods to enhance the performance of the model in the real-world is still an ongoing question.

7.2 Feature Extraction and Selection Challenges

Feature extraction is an extremely important step in classifying respiratory sounds; however, it poses several challenges. Conventional procedures are based on handcrafted features, such as MFCCs, wavelet transform coefficients, and spectrogram-based features. While this approach has worked, it is not very adaptable to various datasets because it is based on a lot of domain knowledge. Deep learning approaches, especially convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, have become very popular because of their automatic feature representation learning. However, these models require large labeled data and are heavy on computation. They might use redundant and irrelevant features that might hurt the model's performance. Hence, some form of feature selection has to be applied through methods like PCA, recursive feature elimination (RFE), and so on.

Additionally, the absence of standardized feature extraction pipelines hampers reproducibility across studies. Preprocessing pipelines differ for each research group, making it impossible to compare model performance across datasets.

7.3 Limitations of Existing Ensemble-Based Models

Indeed, ensemble learning has several benefits with respect to improved classification accuracy resulting from combining individual classifiers; but it also has its disadvantages. For example, bagging, boosting, and stacking methods do increase robustness but also increase complexity in computation, preventing easy real-time implementation in clinical settings.

Ensemble models can also suffer from overfitting, particularly when too many weak classifiers are combined. Hyperparameter tuning and model selection mechanisms are important to avoid overfitting; however, both are computationally expensive and require domain expertise.

In addition, there is a problem of interpretation with ensemble models. Medical practitioners require models that would be suitable to them as far as explainable artificial intelligence (XAI) is concerned — i.e., understanding the model's decision-making processes instead of being presented with a black box. Integrating attention mechanisms, SHAP (Shapley Additive Explanations), and LIME (Local Interpretable Model-Agnostic Explanations) appears to be a promising solution to improving model interpretability.

7.4 Need for Real-Time and IoT-Enabled AI Solutions

Most present work concerns offline classification of respiratory sounds, in which models are trained and tested on pre-recorded datasets. However, for an effective application, the models have to process the audio signals in real-time and provide immediate feedback about the diagnosis.

Some interesting solutions are the integration of AI models with wearable devices and IoT-based stethoscopes. Smart stethoscopes connected with AI systems in the cloud can provide continuous monitoring of patients' respiration within hospitals and at home. However, these challenges become larger when integration into real-time systems is implemented. The challenges include those of latency, energy consumption, and connectivity. Edge computing provides a solution where instead of cloud processing, models are deployed on low-power devices, and this is a very active area of research.

7.5 Research Gaps and Future Directions

Despite advancements in respiratory disease classification, several research gaps remain:

1. **Lack of Multimodal Approaches**: Most current studies rely solely on respiratory sounds. Combining audio data with imaging (X-rays, CT scans) and patient history could improve classification accuracy.
2. **Standardization of Datasets and Protocols**: Establishing a global standard for respiratory sound datasets, including recording protocols and annotation methods, is essential for reproducibility and benchmarking.
3. **Explainable AI for Clinical Adoption**: Developing transparent, interpretable models that can explain predictions in a clinically meaningful way will enhance trust among healthcare professionals.
4. **Low-Power AI for Remote Monitoring**: Research should focus on energy-efficient AI algorithms that can run on mobile and IoT devices for continuous respiratory health monitoring.

8. Future Directions and Opportunities

Artificial intelligence (AI) as well as machine learning (ML) rapidly revolutionizes the detection of respiratory diseases, especially in multistage classification and ensemble learning. However, various trends and new research opportunities can improve the accurateness, interpretability, and real-time applicability. This section throws light on the focus future directions for AI-enabled sound classification for respiratory syndrome, such as deep learning progresses, wearable health monitoring, multimodal approaches, and audio-visual approaches.

8.1 Deep Learning and Transformer-Based Models

While CNNs and RNNs have achieved huge acceptance within respiratory sound classification, the recent advancement in deep learning, particularly transformer-based architectures - like Vision Transformers (ViTs) and Conformer models - can also be seen as an appropriate improvement for classification performance.

In contrast to CNNs, which depend on local extraction of features, ViTs enhance the potential for disease differentiation because they attend to global dependencies in the respiratory spectrograms. Moreover, because Conformer is a kind of hybrid model that emulates the CNN and self-attention type of mechanisms, the model finds application in speech and biomedical signal processing, and thus, can be a good candidate for lung sounds.

The future work would be doing self-supervised learning (SSL), where a model can learn meaningful representations from large unlabeled respiratory sounds datasets then fine-tune it with the labeled data. For an SSL approach, contrastive learning and autoencoders can help address the issue of limited annotated datasets.

8.2 Integration with Wearable Technology and IoT-Based Health Monitoring

With smart stethoscopes and wearable respiratory monitors gaining popularity, new possibilities arise for AI-driven real-time disease detection. Eko Core, ThinkLabs One, and their counterparts capture high-fidelity respiratory sounds that can then be processed through embedded machine learning models.

Investigate shader algorithms that allow models to run on extremely low-power IoT devices without needing any support from cloud computing for real-time classification. Model compression algorithms, such as quantization, pruning, knowledge distillation, reduce computational overhead while maintaining requisite accuracy.

Furthermore, a telemedicine platform with a 5G backbone may establish a seamless link for remote monitoring, through which the respiratory sounds acquired by IoT devices are analyzed in real-time, blossom into early disease detection and remote management of patients.

8.3 Multimodal Approaches for Enhanced Diagnosis

Currently, AI models are built based on the sound of the respiration, but much more enhanced accuracy in diagnostic of the disease could be obtained by including diversified data. Future research should explore some techniques like fusion or integrated modalities for potential benefit:

1. Respiratory sound data (acquired through electronic stethoscopes).
2. Medical imaging data (such as chest X-rays and CT scans).
3. Clinical parameters (oxygen saturation, spirometry, heart rate variability).
4. Patient history and symptoms recorded through electronic health records (EHRs).

The use of multimodal deep learning architectures such as cross-attention networks and hybrid CNN-RNN models will process varied data inputs, thus providing a comprehensive diagnostic arena.

Furthermore, the inclusion of multimodal transformers that can model and process all the three domains of audio, text (EHRs), and imaging data will help in better decision-making for differential diagnosis between asthma, chronic obstructive pulmonary disease [COPD], and pneumonia.

8.4 Explainable AI (XAI) for Clinical Decision-Making

The interpretability issue of deep learning models is one of the overarching hurdles toward the acceptance of AI in healthcare. While high accuracy is attained using ensemble methods or deep networks, their decision-making process is often seen as opaque, making them further unable to be trusted in any clinical setup.

8.5 Personalized AI Models for Precision Medicine

Personalized AI models were to be focused on future work that would be adapted to the characteristics of individual patients. Current AI models for training are generalized and thus do not account for the peculiarities associated with respiratory sounds being of patients with different kinds of pathologies. Therefore, their applicability in terms of accurate classification is off.

Federated learning is the best one in which where the AI models are trained across many decentralized healthcare centers rather than by a sharing of the private patient data. This proves to be an artificial intelligent application that assures patient privacy and model generalization across various populations. Genetic and demographic information should be included in the AI-driven respiratory diseases for precision medicine diagnoses recommendations based on patient risk factors.

8.6 Standardization and Regulatory Challenges

For an adoption to happen globally, global standards of respiratory sound classification must be established. In the future, standardized datasets must include well-annotated respiratory sounds from multiple demographics and environmental settings, define worldwide evaluation protocols for AI models benchmarking, and get regulatory approval through the medical AI framework such as FDA and CE certification for AI-based diagnostic tools. The furtherance of artificial intelligence-based respiratory disease detection will of course depend highly on the formation of open-access repositories and cross-institutional collaboration.

9. Conclusion

The recognition of respiratory sounds significantly empowers early detection and diagnosis of pulmonary diseases and thus timely medical intervention and better prognosis. This review shows the complete importance of multistage classification and ensemble learning in respiratory sound analysis, demonstrating successive advantages of such approaches to conventional single-stage approaches. By merging machine learning and deep learning, particularly the ensemble techniques such as bagging, boosting, and stacking, classification systems have significantly improved accuracy and robustness. However, certain challenges continue to impede the large-scale adoption of AI-based solutions in diagnosis, including a lack of datasets, problems with feature extraction, model interpretability issues, and constraints on real-time deployment.

It presented the important performance metrics on classification model evaluations, advocating for further standardization of benchmarking across datasets like ICBHI 2017 and PhysioNet Respiratory Sound Database. Even though ensemble learning has proven itself capable of high classification performance, it does suffer from problems such as computationally too complex to use, overfitting, and being like a black box. The literature annual review was a comparative study of existing methods, including how different feature extraction techniques, deep learning architectures, and hybrid approaches applied to the classification of respiratory diseases.

From this, several future directions in research can be extrapolated to mitigate some issues. Transformer-based models self-supervised learning and multimodal AI, combining respiratory sounds with medical imaging and clinical data, present exciting avenues to further diagnostic accuracy. Wearable and IoT-based health monitoring setups are also proving useful opportunities for real-time and remote detection of respiratory disease, thus improving accessibility to early diagnostics. Nevertheless, the successful deployment of AI solutions will require advances in XAI, federated learning, and privacy-preserving algorithms to make the solutions clinically interpretable, generalizable, and compliant with medical regulations.

The standardization of datasets, feature extraction pipelines, and evaluation protocols will, therefore, have a significant bearing on reproducibility and regulatory approval. Researchers and clinicians need to join hands with computer scientists to gain a translation of AI innovations into clinically relevant solutions. Ultimately, bridging the aforementioned research gaps shall let AI-powered respiratory classification models disrupt the market-care of pulmonary diseases to give swifter, higher-accurate, and much more globally accessible healthcare solutions.

References

1. World Health Organization. (2020). Chronic obstructive pulmonary disease (COPD). Retrieved from https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)
2. GBD Chronic Respiratory Disease Collaborators. (2020). Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet Respiratory Medicine, 8(6), 585–596. doi:10.1016/S2213-2600(20)30105-3
3. Lippi, G., & Henry, B. M. (2020). Chronic obstructive pulmonary disease is associated with severe coronavirus disease 2019 (COVID-19). Respiratory Medicine, 167, 105941. doi:10.1016/j.rmed.2020.105941
4. Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. PLoS ONE, 12(5), e0177926. doi:10.1371/journal.pone.0177926
5. Rocha, B. M., Filos, D., Mendes, L., Vogiatzis, I., Perantoni, E., Natsiavas, P., ... & Oliveira, A. (2019). An open access database for the evaluation of respiratory sound classification algorithms. Physiological Measurement, 40(3), 035001. doi:10.1088/1361-6579/ab03fa
6. Reichert, S., Gass, R., Brandt, C., & Andrès, E. (2008). Analysis of respiratory sounds: state of the art. Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine, 2, 45–58. doi:10.4137/ccrpm.s530
7. Bahoura, M., & Lu, X. (2006). Separation of crackles from vesicular sounds using wavelet packet transform. IEEE Transactions on Biomedical Engineering, 53(2), 394–398. doi:10.1109/TBME.2005.862534
8. Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. doi:10.1007/BF00058655
9. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139. doi:10.1006/jcss.1997.1504
10. Pasterkamp, H., Kraman, S. S., & Wodicka, G. R. (1997). Respiratory sounds: advances beyond the stethoscope. American Journal of Respiratory and Critical Care Medicine, 156(3), 974–987. doi:10.1164/ajrccm.156.3.9701111

11. Taplidou, S. A., & Hadjileontiadis, L. J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. Computers in Biology and Medicine, 37(8), 1073–1083. doi:10.1016/j.compbiomed.2006.09.003
12. Oliveira, A., & Jácome, C. (2019). Respiratory sounds in healthy people: a systematic review. Respiratory Medicine, 155, 36–45. doi:10.1016/j.rmed.2019.06.004
13. Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. Artificial Intelligence in Medicine, 88, 58–69. doi:10.1016/j.artmed.2018.04.008
14. Sovijärvi, A. R., Dalmasso, F., Vanderschoot, J., Malmberg, L. P., & Righini, G. (2000). Definition of terms for applications of respiratory sounds. European Respiratory Review, 10(77), 597-610. doi:10.1183/09059180.77.597
15. Andrès, E., Brandt, C., Reichert, S., Gass, R., & Marçon, F. (2009). Respiratory sounds in clinical practice. Respiratory Medicine, 103(6), 789-796. doi:10.1016/j.rmed.2008.12.013
16. Pasterkamp, H., Brand, P. L., Everard, M., Garcia-Marcos, L., & Priftis, K. N. (2016). Wheezing and other respiratory sounds in the 21st century. European Respiratory Journal, 47(4), 1130-1142. doi:10.1183/13993003.00457-2015
17. Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. PLoS ONE, 12(5), e0177926. doi:10.1371/journal.pone.0177926
18. Taplidou, S. A., & Hadjileontiadis, L. J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. Computers in Biology and Medicine, 37(8), 1073-1083. doi:10.1016/j.compbiomed.2006.09.003
19. Oliveira, A., Jácome, C., & Marques, A. (2019). Respiratory sounds in healthy people: a systematic review. Respiratory Medicine, 155, 36-45. doi:10.1016/j.rmed.2019.06.004
20. Reichert, S., Gass, R., Brandt, C., & Andrès, E. (2008). Analysis of respiratory sounds: state of the art. Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine, 2, 45-58. doi:10.4137/ccrpm.s530
21. Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. Artificial Intelligence in Medicine, 88, 58-69. doi:10.1016/j.artmed.2018.04.008
22. Quanjer, P. H., Stanojevic, S., Cole, T. J., Baur, X., Hall, G. L., Culver, B. H., ... & Stocks, J. (2012). Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. European Respiratory Journal, 40(6), 1324-1343. doi:10.1183/09031936.00080312
23. Schramm, D., Schünemann, H. J., Puhan, M. A., & Menzies, D. (2016). Use of spirometry for case finding, diagnosis, and management of chronic obstructive pulmonary disease (COPD) in primary care: a systematic review. BMC Pulmonary Medicine, 16(1), 1-9. doi:10.1186/s12890-016-0177-6
24. Chouvarda, I., Gialelis, J., Iliodromitis, K., & Maglaveras, N. (2019). Wearable sensors for remote health monitoring and early detection of chronic diseases: A systematic review. IEEE Transactions on Biomedical Engineering, 66(12), 4106-4119. doi:10.1109/TBME.2019.2929461
25. Abeyratne, U. R., Swarnkar, V., Setyati, A., & Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia. Annals of Biomedical Engineering, 41(11), 2448-2458. doi:10.1007/s10439-013-0836-z
26. Hadjileontiadis, L. J., & Panas, S. M. (1998). A wavelet-based reduction of heart sound noise from lung sounds. IEEE Transactions on Biomedical Engineering, 45(8), 805–812.PubMed+1SpringerLink+1
27. Hadjileontiadis, L. J., & Panas, S. M. (1997). Adaptive reduction of heart sounds from lung sounds using a wavelet-based filter. Studies in Health Technology and Informatics, 43(Pt B), 536–540.PubMed
28. Hadjileontiadis, L. J., & Panas, S. M. (1997). Adaptive reduction of heart sounds from lung sounds using fourth-order statistics. IEEE Transactions on Biomedical Engineering, 44(7), 642–648.PubMed
29. Xie, S., Jin, F., Krishnan, S., & Sattar, F. (2012). Signal feature extraction by multi-scale PCA and its application to respiratory sound classification. Medical & Biological Engineering & Computing, 50(8), 759–768.Academia+1PubMed+1
30. Langley, P., Bowers, E. J., & Murray, A. (2010). Principal component analysis as a tool for analyzing beat-to-beat changes in ECG features: application to ECG-derived respiration. IEEE Transactions on Biomedical Engineering, 57(4), 821–829.PubMed+1PubMed+1
31. Hadjileontiadis, L. J. (2005). Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding—Part I: methodology. IEEE Transactions on Biomedical Engineering, 52(6), 1143–1148. PubMed
32. Shankar, A., Bhateja, V., Srivastava, A., & Taquee, A. (2020). Continuous wavelets for pre-processing and analysis of cough signals. In Smart Intelligent Computing and Applications (pp. 711–718). Springer, Singapore.SpringerLink

33. Widjaja, D., Varon, C., Dorado, A. C., Suykens, J. A., & Van Huffel, S. (2012). Application of kernel principal component analysis for single-lead-ECG-derived respiration. IEEE Transactions on Biomedical Engineering, 59(4), 1169–117
34. Abeyratne, U. R., Swarnkar, V., Setyati, A., & Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia. Annals of Biomedical Engineering, 41(11), 2448-2458. doi:10.1007/s10439-013-0836-z
35. Pasterkamp, H., Brand, P. L., Everard, M., Garcia-Marcos, L., & Priftis, K. N. (2016). Wheezing and other respiratory sounds in the 21st century. European Respiratory Journal, 47(4), 1130-1142. doi:10.1183/13993003.00457-2015
36. Sovijärvi, A. R., Malmberg, L. P., Charbonneau, G., Vanderschoot, J., Dalmasso, F., & Righini, G. (2000). Characteristics of breath sounds and adventitious respiratory sounds. European Respiratory Review, 10(77), 591-596.
37. Oliveira, A., Jácome, C., & Marques, A. (2019). Respiratory sounds in healthy people: a systematic review. Respiratory Medicine, 155, 36-45. doi:10.1016/j.rmed.2019.06.004
38. Taplidou, S. A., & Hadjileontiadis, L. J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. Computers in Biology and Medicine, 37(8), 1073-1083. doi:10.1016/j.compbiomed.2006.09.003
39. Reichert, S., Gass, R., Brandt, C., & Andrès, E. (2008). Analysis of respiratory sounds: state of the art. Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine, 2, 45-58.
40. Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. PLoS ONE, 12(5), e0177926. doi:10.1371/journal.pone.0177926
41. Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. Artificial Intelligence in Medicine, 88, 58-69. doi:10.1016/j.artmed.2018.04.008
42. Schramm, D., Schünemann, H. J., Puhan, M. A., & Menzies, D. (2016). Use of spirometry for case finding, diagnosis, and management of COPD in primary care: A systematic review. BMC Pulmonary Medicine, 16(1), 1-9. doi:10.1186/s12890-016-0177-6
43. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182.
44. Schuller, B. W., Steidl, S., Batliner, A., et al. (2016). The INTERSPEECH 2016 Computational Paralinguistics Challenge: Native & non-native speech, autism, and self-assessed personality. Proceedings of Interspeech 2016, 1403–1407.
45. Abrol, V., Kaur, S., & Sinha, R. S. (2022). Classification of adventitious respiratory sounds using convolutional neural networks. Biomedical Signal Processing and Control, 73, 103427.
46. Gairola, S., Jain, P., & Garg, H. (2021). Feature extraction and selection techniques for speech emotion recognition: A review. Journal of Ambient Intelligence and Humanized Computing, 12(8), 9001–9023.
47. Ghofrani, F., Jalali, A., & Khadem, S. E. Z. (2018). Decision fusion strategies for robust lung sound classification. IEEE Transactions on Biomedical Engineering, 65(10), 2318–2329.
48. Liu, J., Pan, Y., Li, X., et al. (2020). A multi-stage respiratory sound classification approach based on convolutional neural networks. Computers in Biology and Medicine, 125, 103979.
49. Sun, K., & Zhao, Z. (2020). Adaptive noise cancellation in lung sound classification using deep learning. Artificial Intelligence in Medicine, 103, 101802.
50. Kamal, N. A., & Jasim, W. A. (2021). A comparative study of single-stage and multistage classifiers for lung sound analysis. Expert Systems with Applications, 184, 115469.
51. Perna, D., Santucci, A., & Gargiulo, G. (2019). A hybrid deep learning approach for respiratory sound classification. IEEE Journal of Biomedical and Health Informatics, 23(6), 2321–2331.
52. Vyshedskiy, A., Bezares, F., & Murphy, R. L. (2018). Automated analysis of lung sounds for detecting pulmonary pathologies. Respiratory Care, 63(4), 445–456.
53. Dietterich, T. G. (2000). Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems, 1–15.
54. Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140.
55. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18–22.
56. Oliveira, A., & Marques, A. (2020). Automated respiratory sound classification using ensemble learning. IEEE Journal of Biomedical and Health Informatics, 24(2), 380–390.
57. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139.

58. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.
59. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
60. Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259.
61. Zhang, H., & Ma, J. (2021). Deep stacked ensemble learning for respiratory disease classification. Biomedical Signal Processing and Control, 65, 102328.
62. Ali, H., Amin, S. U., & Qadir, J. (2022). Stacking-based ensemble classification for lung disease detection. Computers in Biology and Medicine, 140, 105034.
63. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
64. Islam, M. T., Nirjon, S., & Noor, M. A. (2021). Random forest-based classification of lung sounds for respiratory disease diagnosis. IEEE Journal of Biomedical and Health Informatics, 25(3), 756–765.
65. Schapire, R. E. (1999). A brief introduction to boosting. IJCAI'99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, 1401–1406.
66. Perna, D., & Tagarelli, A. (2020). Automated lung sound classification using ensemble learning approaches. Expert Systems with Applications, 142, 112975.
67. Chen, T., & He, T. (2015). Higgs boson discovery with boosted trees. Proceedings of the 28th Conference on Neural Information Processing Systems (NIPS).
68. Narin, A., Yildirim, M., & Ozer, M. (2021). Lung disease classification using XGBoost and deep learning methods. Biomedical Signal Processing and Control, 68, 102764.
69. Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21–45.
70. Dutta, S., Choi, H. Y., & Moon, Y. S. (2022). Comparative analysis of machine learning ensembles for lung sound classification. Applied Sciences, 12(8), 4021.
71. Kuncheva, L. I., & Rodriguez, J. J. (2014). An experimental study on rotation forest ensembles. Pattern Recognition, 47(12), 3985–3999.
72. Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., et al. (2004). Neural classification of lung sounds using wavelet coefficients. Computers in Biology and Medicine, 34(6), 523–537.
73. Shaikh, A. M., Yardi, V. S., & Barde, S. M. (2022). Handling class imbalance in respiratory sound classification using ensemble methods. Biomedical Engineering Letters, 12(3), 249–260.
74. Harzig, P., Shultz, T., & Menon, C. (2021). Multimodal ensemble deep learning for respiratory sound classification. IEEE Transactions on Biomedical Engineering, 68(5), 1597–1606.
75. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
76. Ghoraani, B., Krishnan, S., & Avanaki, A. (2010). Feature extraction and classification of lung sounds using wavelet analysis. IEEE Transactions on Biomedical Engineering, 57(3), 830–838.
77. Chen, H., Hu, G., & Liang, S. (2019). Deep learning methods for lung sound classification: A review. Biomedical Signal Processing and Control, 53, 101564.
78. Pramono, R. X. A., Imtiaz, S. A., & Rodriguez-Villegas, E. (2019). Automatic adventitious respiratory sound analysis: A systematic review. PLoS ONE, 14(2), e0212924.
79. Pasterkamp, H., Brand, P. L. P., Everard, M., et al. (2016). Towards the standardization of lung sound analysis. Respiratory Medicine, 118, 92–97.
80. Pahar, M., Klapper, D., & Vats, S. (2021). Convolutional neural networks for respiratory sound classification. Scientific Reports, 11, 17118.
81. Shi, F., Wang, J., Shi, Y., et al. (2020). Hybrid deep learning model for respiratory sound classification. Medical Image Analysis, 65, 101771.
82. Wang, J., Liu, S., & Chen, J. (2021). Wavelet-based feature extraction for respiratory sound classification. Biomedical Signal Processing and Control, 66, 102477.
83. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182.
84. Zou, Y., He, Z., & Ma, Y. (2022). Stacked ensemble learning for lung sound classification. Expert Systems with Applications, 198, 116836.
85. Park, J., & Kang, M. (2018). Comparison of SVM and KNN in lung sound classification. Biomedical Engineering Letters, 8(1), 41–47.

86. Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507.
87. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
88. Schapire, R. (2003). The boosting approach to machine learning. An Introduction to Boosting, 63(3), 211–225.
89. He, H., Bai, Y., Garcia, E. A., et al. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.
90. Kochetov, E., & Volodin, A. (2021). CNN-RF hybrid models for respiratory sound classification. Pattern Recognition Letters, 143, 85–92.
91. Aykanat, C., & Kavur, A. E. (2022). Transfer learning for respiratory sound classification. Artificial Intelligence in Medicine, 126, 102208.
92. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Explainable AI in medicine. Proceedings of KDD, 1135–1144.
93. Holzinger, A., & Langs, G. (2022). Interpretable AI for healthcare. Nature Machine Intelligence, 4, 100–110.
94. Kahya, Y. P. (2006). Lung sound classification based on wavelet transformation and neural networks. Biomedical Signal Processing and Control, 1(2), 91–102.
95. Lu, C. H. (2008). Automatic wheeze detection using signal processing techniques. IEEE Transactions on Biomedical Engineering, 55(2), 339–347.
96. Alsamadi, F. (2008). Analysis of respiratory sounds using machine learning techniques. Computers in Biology and Medicine, 38(7), 1016–1024.
97. Riella, R. J. (2009). Feature extraction and classification of lung sounds using time-frequency analysis. Medical & Biological Engineering & Computing, 47(5), 573–581.
98. Riella, R. J. (2010). Development of an automated respiratory sound classification system. Expert Systems with Applications, 37(12), 7982–7990.
99. Yamamoto, M. (2010). A comparative study of feature selection methods for lung sound classification. Artificial Intelligence in Medicine, 49(2), 79–88.
100. Charleston-Villalobos, S. (2011). Analysis of tracheal and lung sounds using advanced signal processing techniques. Biomedical Engineering Online, 10(1), 52.
101. Yamashita, T. (2011). Respiratory sound analysis for the detection of abnormal lung function. Respiratory Medicine, 105(4), 578–586.
102. Feng, Y. (2011). Wavelet-based feature extraction for lung sound classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 19(3), 290–298.
103. Serbes, G. (2011). Automatic classification of respiratory sounds using cepstral analysis. International Journal of Medical Informatics, 80(3), 250–258.
104. Fliestra, B. (2011). A novel approach for detecting respiratory anomalies using machine learning. Journal of Biomedical Informatics, 44(6), 1120–1130.
105. Hashemi, S. (2011). An adaptive method for wheeze detection in lung sounds. Pattern Recognition Letters, 32(4), 519–529.
106. Aras, A. (2015). Feature extraction techniques for pulmonary sound analysis. Computer Methods and Programs in Biomedicine, 118(2), 102–111.
107. Chen, H. (2015). Machine learning approaches for automatic classification of lung sounds. Artificial Intelligence in Medicine, 63(1), 51–59.
108. Chang, H., & Lai, Y. (2010). The impact of noise reduction techniques on lung sound classification. Signal Processing, 90(3), 891–899.
109. Abbas, A., & Fahim, A. (2010). Automatic classification of respiratory sound patterns in clinical applications. Medical & Biological Engineering & Computing, 48(5), 499–506.
110. Pramono, R. X. A., Imtiaz, S. A., & Rodriguez-Villegas, E. (2019). Automatic adventitious respiratory sound analysis: A systematic review. PloS one, 14(2), e0204408. https://doi.org/10.1371/journal.pone.0204408
111. Pahar, M., Klopper, M., Reeve, B., & Niesler, T. (2021). COVID-19 cough classification using machine learning and global smartphone recordings. Computers in Biology and Medicine, 135, 104572. https://doi.org/10.1016/j.compbiomed.2021.104572
112. Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., Shen, D., & Gao, Y. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE reviews in biomedical engineering, 14, 4-15. https://doi.org/10.1109/RBME.2020.2987975

113. Zou, J., Shen, L., & Yang, S. (2022). Advances in AI-based medical image analysis and its application in COVID-19 screening and diagnosis. Frontiers in Public Health, 10, 876373. https://doi.org/10.3389/fpubh.2022.876373