



Harmonizing Respiratory Sound Insights: Unleashing the Potential of Machine Learning Classifiers Through Hyperparameter Elegance

Vishnu Vardhan Battu^{1*}, Kalaiselvi Geetha Manoharan¹, Syam Prasad Gudapati²

¹ Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Chidambaram 608002, India

² Department of Computer Science and Engineering, Sri Vasavi Institute of Engineering & Technology, Nandamuru 521369, India

Corresponding Author Email: bvishnuvardhan84@gmail.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300211>

ABSTRACT

Received: 17 October 2024

Revised: 5 January 2025

Accepted: 16 January 2025

Available online: 27 February 2025

Keywords:

audio augmentation, edge devices, feature extraction techniques, ICHBI 2017 dataset, machine learning classifiers, MFCC mean, MFCC (Mean+Std), respiratory diseases

Respiratory diseases account for 14.1% of global deaths annually, as reported by the World Health Organization, underscoring the urgent need for advanced diagnostic tools. This study investigates the classification of respiratory sounds using machine learning techniques applied to the ICBHI 2017 dataset. To address class imbalance and improve generalizability, augmentation methods such as time masking and stretching were implemented. Feature extraction techniques, including Mel-Frequency Cepstral Coefficients (MFCCs), were combined with lightweight classifiers like KNN and SVM, achieving a classification accuracy of 99%. The proposed approach surpasses prior benchmarks while maintaining computational efficiency, making it suitable for deployment on edge devices in resource-constrained healthcare environments.

1. INTRODUCTION

Respiratory diseases remain a significant global health burden, accounting for a considerable portion of global morbidity and mortality. According to the World Health Organization (WHO) Global Health Statistics Report 2022 [1], respiratory diseases contribute to approximately 14.1% of global deaths annually, underscoring the urgent need for advanced diagnostic tools. This study focuses on leveraging machine learning techniques to enhance the efficiency, accuracy, and objectivity of respiratory sound analysis, particularly in resource-constrained environments.

Machine learning offers a promising avenue for enhancing respiratory sound analysis. Using automated techniques through machine learning for diagnostic precision in image analysis might enhance accuracy, objectivity, and efficiency, potentially minimizing patient interventions and improving patient care. The research article aims to develop a robust and computationally complex classification model leveraging machine-learning techniques to classify respiratory sounds while accounting for class imbalances and computational complexities.

One such public dataset is the ICBHI 2017 Respiratory Sound Database [2], which contains a wide range of respiratory sounds (tailed breaths, wheezes, and crackles) and serves as a test case for developing and comparing respiratory sound classifiers. So far, several studies have exploited machine learning methods [3] to analyze respiratory sounds derived from this dataset, including wavelet coefficients [4], convolutional neural networks, and deep neural networks [5].

However, difficulties, including class imbalance and computational cost, hinder large-scale deployment.

Our selected classification methods were hyperparameter-tuned using the ICBHI 2017 dataset. Our suggested model uses MFCC Mean, MFCC, and Chroma-Mean_std to obtain great accuracy with little computational cost on extremely low specifications. However, our solutions work best suited for a limited applications compound with resource-constrained environments. Early test results are promising and suggest therapeutic application.

2. LITERATURE REVIEW

Introduction to Respiratory Sound Analysis and Edge Computing: Respiratory diseases are a significant global health burden that requires rapid and precise diagnosis. Conventional auscultation methods are subjective and variable [6]. Automated respiratory sound analysis via machine learning is a promising approach toward more objective and efficient assessments. The growing availability of mobile and edge computing devices enables point-of-care diagnostics, underscoring the need for computationally efficient solutions.

Deep Learning for Classifying Respiratory Sound: Deep learning models, especially Convolutional Neural Networks, have shown remarkable performance in respiratory sound classification [7-11]. However, their computational demands can hinder deployment on edge devices. Hybrid CNN-RNN architectures [12, 13] and other deep learning methods [14, 15] have been explored, but computational efficiency remains a

critical challenge.

Efficient Feature Extraction for Edge Devices: Consequently, the ratio between accuracy and computational cost relies on effective feature extraction. Features from the spectrum domain, including Mel-Frequency Cepstral Coefficients [16, 17], are employed. Feature selection is essential for optimising performance on resource-limited devices. Alternative methods for extracting relevant information from respiratory sounds include wavelet transforms and various signal-processing techniques [18-22]. It is crucial to investigate lower complexity features for implementation in edge devices.

Machine Learning on Edge Devices: The simplification of respiratory sound classifications task on the edge devices by Efficient feature extraction and less computationally demanding machine learning models. Such a practical point-of-care diagnostics solution will likely enhance access to timely and accurate respiratory status measurements. A recent study [23] the feasibility of abnormal sound detection with deep learning and a vest-coat stethoscope, a practical application. Such systems demonstrate evidence of the integration of wearables and edge computing for respiratory health monitoring. It is important to conduct further research to optimize machine learning models to make them work at lower energy levels or on devices with limited resources.

Public Datasets and Challenges: The ICBHI, which aims to automatically detect adventitious respiratory sounds on a large, standardized dataset, and the emergence of public respiratory sound datasets [24] have propelled research in this area considerably. These datasets are helpful for detailed evaluations and comparisons between methods, leading to collaborative innovations.

Addressing Challenges: Noise Reduction and Data Augmentation Specifically, various noise reduction methods like spectral subtraction, as well as wavelet denoising, and data augmentation methods like adding noise, time stretching, and pitch shifting, are essential to achieve the robustness and generalizability of respiratory sound classification models [25]. Finding ways to overcome these challenges has the potential to improve the performance of ML models, especially in-practice clinical scenarios.

This research focuses on developing a state-of-the-art machine learning model of good performance and low computation for automatic classification of respiratory sounds based on the ICBHI 2017 dataset trained on good responses of the ICBHI 2017 dataset. The class imbalance is also tackled through data augmentation and hyperparameter tuning, benefiting its real-world clinical usability in place of traditional analysis methods. In particular, the aims are to:

1. **Build and Validate a Strong Model:** This will help build a strong machine-learning model that can perform well, even in noise or other real-world differences, ensuring reliability and generalizability in diverse datasets and disease settings.
2. **Address Class Imbalance:** Address the challenge of class imbalance using data augmentation methods and other balancing approaches so that poorly represented sound categories can be accurately classified.
3. **Reduce Computational Complexity:** Create a computationally efficient model that can be used for real-time processing and deployment on devices with limited resources, enabling quicker processing and broader access in clinical environments.
4. **Validate the Model Using the ICBHI 2017 Dataset:** As

mentioned in step no. 3, the second task is to validate the model using the ICBHI 2017 dataset. ICBHI 2017 is a widely used dataset in biomedical signal processing, making it easier to compare the model performance against existing literature and measure accuracy/efficacy objectively.

3. DATASET PREPARATION

1. The ICBHI 2017 Respiratory Sound Database serves as the primary dataset for this study. It includes 920 labeled recordings obtained from 126 individuals, encompassing 6,898 respiratory cycles. Table 1 shows the classification of respiratory sounds by the number of cycles. These recordings are categorized into four classes:

Table 1. Classification of respiratory sounds by number of cycles

Disease/Class	Number of Cycles	Percentage of Total (%)
Healthy/Normal	1,200	17.4%
Crackles (e.g., Pneumonia, COPD)	1,720	25.0%
Wheezes (e.g., Asthma, Bronchitis)	920	13.3%
Crackles+Wheezes (Mixed Respiratory Conditions)	3,058	44.3%
Total	6,898	100%

This dataset was specifically designed for benchmarking machine learning models in respiratory sound classification, providing a diverse set of examples for model evaluation.

2. **Respiratory Database@TR [26]:** This dataset offers multi-channel recordings, capturing 12 lung sound channels for each patient. It includes recordings from patients with varying severities of Chronic Obstructive Pulmonary Disease, categorized from COPD0 to COPD4. The multi-channel recordings provide a richer representation of lung sounds, enabling more nuanced analysis. However, the short recording duration (at least 17 seconds) might limit the data available for training complex models.
3. **HF_Lung_V2 [27]:** This publicly accessible database extends the HF_Lung_V1 dataset, focusing on computer-aided and automated lung sound analysis. It has more subjects (303) and sample recordings (4138) than numerous other datasets. A description of the respiratory conditions encompassed in the data would enhance it.
4. **R.A.L.E. Lung Sounds 3.1 [28]:** This dataset encompasses several adventitious lung sounds associated with distinct respiratory illnesses, including wheezes, crackles, and pleural rubs. The specifics regarding the number of recordings and the related respiratory diseases examined would enhance the characterisation of the description.

We used a few audio data augmentation methods to avoid class imbalance and increase the robustness of our trained models [29-31]. These methods amplify the training dataset while maintaining significant content and mitigate the dangers of data shortages and overfitting. To address class imbalance and improve model generalizability, we applied the following augmentation techniques using the librosa library in Python:

1. **Loudness Normalization:** Adjusted the amplitude of recordings to ensure uniform loudness levels.
2. **Time Masking:** Randomly masked segments of audio ranging from 100–200 ms to simulate missing data and enhance robustness.

3. Time Shifting: Shifted audio signals by ± 50 ms to improve generalization across variable recording timings.
4. Time Stretching: Applied stretch factors between 0.8 and 1.2 to simulate variations in breathing patterns.

These augmentations expanded the training dataset and minimized overfitting by balancing the class distributions.

These audio augmentation techniques improve model efficiency by expanding the training set and reducing the risk of overfitting associated with limited data [32, 33].

4. PROPOSED METHOD

The proposed method for respiratory sound classification outlines a structured methodology for classifying ICBHI lung sound data into normal and abnormal categories. The process begins with normalisation of the raw lung sound recordings to ensure uniformity across the dataset.

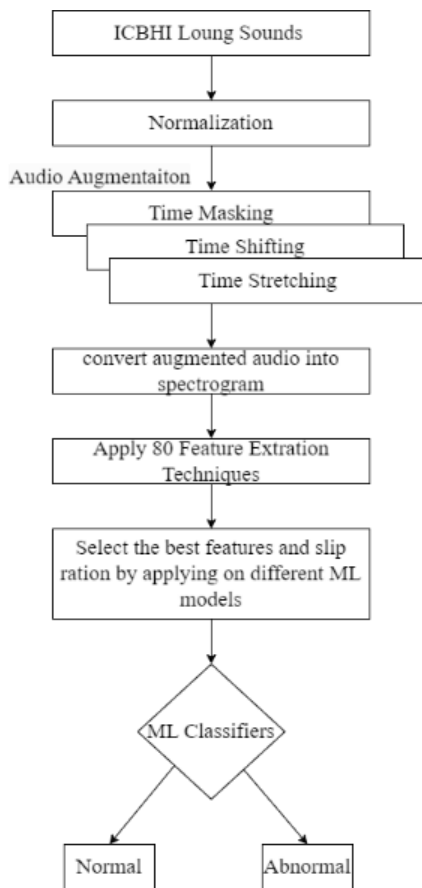


Figure 1. Workflow of the proposed respiratory sound classification method

To improve the dataset's variability and resilience, audio augmentation techniques such as time masking, time shifting, and time stretching are applied. The augmented audio is then transformed into spectrograms, enabling visual representation of the frequency and temporal features of the sounds. Following this, 80 feature extraction techniques are utilised to extract meaningful characteristics from the spectrograms.

A feature selection process is then employed to identify the most relevant features by evaluating their effectiveness across various machine learning (ML) models. These optimised features are subsequently used as input for ML classifiers, which categorise the lung sounds as either normal or abnormal, facilitating effective diagnosis and analysis. The

proposed method for respiratory sound classification is illustrated in Figure 1.

The audio database ICBHI 2017 was employed in the current study with 920 audio signals and 6898 respiratory cycles [34]. The samples were collected in a way that the COPD class has more samples as compared to the Asthma and LRTI classes which have a paucity of samples. To this end, data augmentation was used.

Your "Normalizing and Augmentation" section provides essential details about your preprocessing steps. Here are a few suggestions for improvement:

- *Clarity and Terminology:* While "time masking," "time shifting," and "time stretching" are generally understood, providing a brief explanation or referencing a standard library (like librosa in Python) could enhance clarity. For instance, you could mention that time masking is analogous to masking in image processing, where random sections are zeroed out. Similarly, specifying the range or distribution for time shifting and the stretching factor for time stretching would be beneficial. This level of detail ensures reproducibility and allows others to understand the exact transformations applied.
- *Rationale:* Briefly explaining why these specific augmentation techniques were chosen would strengthen the section. For example, you could mention that time masking helps the model learn to focus on relevant features even with missing information, while time stretching and shifting increase the model's robustness to variations in speech rate and timing.
- *Normalization Details:* You mention splitting signals into 4-second breaths, which is a form of normalization. However, you could also specify whether any amplitude normalization was performed. Common techniques include peak normalization, RMS normalization, or simply scaling the audio to a specific range. Including these details would make your preprocessing steps more comprehensive.
- *Quantifying Augmentation:* Instead of stating that "the number of augmentations vary with the class ratio," provide specific numbers or ratios. For instance, you could say, "The minority class was augmented to achieve a 1:1 ratio with the majority class." This adds precision and makes your methodology more transparent.

4.1 Normalizing and augmentation

The respiratory sound signals were segmented into individual breaths, each 4 seconds long, with a sampling rate of 16000 Hz. To address class imbalance and improve model robustness, we applied several data augmentation techniques using the librosa library in Python:

- **Time Masking:** Random sections of the audio were silenced, simulating missing information and encouraging the model to learn more robust features. The duration of masked sections was randomly chosen between 0.1 and 0.2 seconds.
- **Time Shifting:** Sections of the audio were randomly shifted in time by up to ± 50 ms, increasing the model's tolerance to variations in speech rate.
- **Time Stretching:** The duration of the audio was altered without changing the pitch, using a random stretching factor between 0.8 and 1.2. This augmentation further enhances the model's robustness to variations in speech timing.

These augmentations were applied to the under-represented classes to achieve a balanced class distribution of approximately 1:1 with the majority class in ICBHI 2017 dataset. Additionally, the amplitude of each breath segment was normalized using peak normalization to ensure consistent input levels to the model.

Table 2 represents frequencies of the records in each group. Table 3 shows the number of records which have been chosen for augmentation and the number of times each of them has been augmented.

Time Masking: Random sections of the audio were silenced, simulating missing information and encouraging the model to learn more robust features. The duration of masked sections was randomly chosen between 0.1 and 0.2 seconds.

Time Shifting: Sections of the audio were randomly shifted in time by up to ± 50 ms, increasing the model's tolerance to variations in speech rate.

Table 2. Number of records in each class

Class	No. of Records
Healthy	35
COPD	793
Bronchiectasis	16
Bronchiolitis	13
Pneumonia	37
URTI	23
Asthma	1
LRTI	2
Total	920

Table 3. Number time each record is augmented

Class	Selected Records	No. of Time Each Record is Augmented	Total Records
Healthy	35	30	1050
Bronchiectasis	16	13	208
Bronchiolitis	13	16	208
Pneumonia	37	6	222
URTI	23	9	207
COPD	200	1	200
Total records for training and testing			3140

Time Stretching: The duration of the audio was altered without changing the pitch, using a random stretching factor between 0.8 and 1.2. This augmentation further enhances the model's robustness to variations in speech timing

These augmentations were applied to the under-represented classes to achieve a balanced class distribution of approximately 1:1 with the majority class in ICBHI 2017 dataset. Additionally, the amplitude of each breath segment was normalized using peak normalization to ensure consistent input levels to the model.

Table 2 represents frequencies of the records in each group. Table 3 shows the number of records which have been chosen for augmentation and the number of times each of them has been augmented.

4.1.1 Time masking

Randomly masked audio segments between 100-200 ms using the librosa library to simulate missing information. Is the act of allowing a specific portion of the signal in auditory to the listener at a specific time. In this approach parts of the time are set to zero randomly; this in turn causes the time structure in the audio signal to be lost thus the model relies on other properties of sound.

$$Y(t) = x(t) * m(t) \quad (1)$$

In Eq. (1), $Y(t)$ denotes the masked signal and $x(t)$ and $m(t)$ is an original audio signal. where $m(t)$ equals to 1 for the cases when segment is preserved and equals to 0 when segment is masked.

4.1.2 Time shifting

Is the process of shifting an audio signal in time or in other words, moving it ahead or behind in time but not changing its content. This manipulation is done by moving the signal along the time axis in some way.

$$y(t) = x(t - \Delta t) \quad (2)$$

where, in the above Eq. (2), $y(t)$ is the shifted signal, $x(t)$ is the original signal and Δt is the time shift. If Δt is positive, then the signal is delayed and if Δt is negative then the signal is advanced.

4.1.3 Time stretching

Is the method of altering the temporal characteristics of an audio signal without altering the frequency content of the signal. In mathematical terms time stretching can be expressed by modifying the Short-Time Fourier Transform (STFT) of the sound signal.

$$\text{STFT}(x(t)) = \sum_{n=-\infty}^{\infty} x(t) \cdot w(t - nH) \cdot e^{-j\omega t} \quad (3)$$

In Eq. (3), $x(t)$ is the input audio signal, $w(t - nH)$ denotes the windowing function that isolates a small portion of the signal around time t . The parameter H controls the window size. n is the integer index which shifts the window across the signal. $e^{-j\omega t}$: This is a complex exponential that represents the frequency component at ω .

4.1.4 The pseudocode for data preparation

$x(n)$: Input audio signal

T1, T2, T3, T4: Time masking, shifting, time stretching augmentation functions.

$x'(n) = T4(T3(T2(T1(x(n))))))$

$x''(n) = x'(n)x'(n) \dots x'(n)$

$y(n) = \text{truncate_or_pad}(x''(n), L)$

As depicted in Figure 1 the waveform and spectrogram of the audio sample is depicted before applying any type of audio augmentation. The waveform and spectrogram of the audio when audio augmentation techniques are applied is presented in Figure 2 and Figure 3.

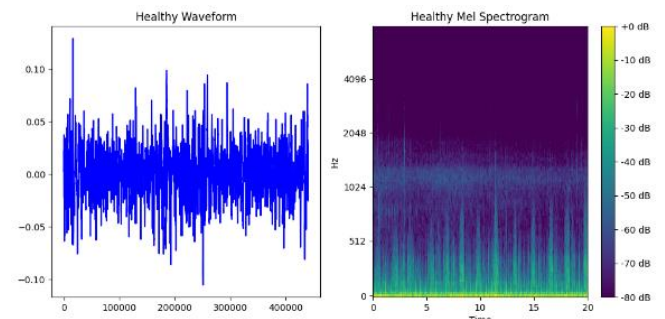


Figure 2. Waveform and spectrogram before audio augmentation

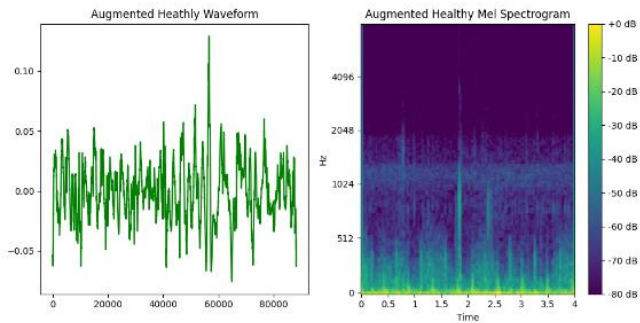


Figure 3. Waveform and spectrogram after audio augmentation

It is a fundamental stage of audio processing that involves turning unprocessed input into a numerical format. A total of 80 features were taken from audio signals for this project. These characteristics fall into several categories.

The following categories apply to the features that have been added to the system:

- Mel-Frequency Cepstral Coefficients (MFCCs): These characteristics show the audio signal's envelope since they are the most crucial component in characterizing its tonal properties [35].
- Chroma features: These give the matching audio's perceived pitch, which may reveal details about the chords that are present in the sound.
- Spectral features: The overall form of the spectrum is also described by the spectral centroid, bandwidth, and rolloff.
- Features connected to rhythm: When analyzing the audio signal's rhythm patterns, zero-crossing rate, harmonic, and percussive characteristics are crucial.

The systematic feature extraction approach facilitates the analysis of each acoustic property's contribution to classification. This analysis enables identifying and eliminating superfluous features, leading to a more robust and interpretable model. Table 4 delineates the 80 features retrieved from the audio signals. Table 5 presents samples of extracted feature values alongside their corresponding labels.

Table 4. A set of 80 features were used to extract from the audio signals

Sl. No	Feature Extraction Technique	Sl. No	Feature Extraction Technique	Sl. No	Feature Extraction Technique	Sl. No	Feature Extraction Technique
1	chroma_cens_mean = []	21	rolloff_var = []	41	mfccs_mean_10 = []	61	chroma_mean_4 = []
2	chroma_cens_std = []	22	zcr_mean = []	42	mfccs_mean_11 = []	62	chroma_mean_5 = []
3	chroma_cens_var = []	23	zcr_std = []	43	mfccs_mean_12 = []	63	chroma_mean_6 = []
4	mel_mean = []	24	zcr_var = []	44	mfccs_std_0 = []	64	chroma_mean_7 = []
5	mel_std = []	25	harm_mean = []	45	mfccs_std_1 = []	65	chroma_mean_8 = []
6	mel_var = []	26	harm_std = []	46	mfccs_std_2 = []	66	chroma_mean_9 = []
7	mfcc_mean = []	27	harm_var = []	47	mfccs_std_3 = []	67	chroma_mean_10 = []
8	mfcc_std = []	28	perc_mean = []	48	mfccs_std_4 = []	68	chroma_mean_11 = []
9	mfcc_var = []	29	perc_std = []	49	mfccs_std_5 = []	69	chroma_std_0 = []
10	mfcc_delta_mean = []	30	perc_var = []	50	mfccs_std_6 = []	70	chroma_std_1 = []
11	mfcc_delta_std = []	31	mfccs_mean_0 = []	51	mfccs_std_7 = []	71	chroma_std_2 = []
12	mfcc_delta_var = []	32	mfccs_mean_1 = []	52	mfccs_std_8 = []	72	chroma_std_3 = []
13	cent_mean = []	33	mfccs_mean_2 = []	53	mfccs_std_9 = []	73	chroma_std_4 = []
14	cent_std = []	34	mfccs_mean_3 = []	54	mfccs_std_10 = []	74	chroma_std_5 = []
15	cent_var = []	35	mfccs_mean_4 = []	55	mfccs_std_11 = []	75	chroma_std_6 = []
16	spec_bw_mean = []	36	mfccs_mean_5 = []	56	mfccs_std_12 = []	76	chroma_std_7 = []
17	spec_bw_std = []	37	mfccs_mean_6 = []	57	chroma_mean_0 = []	77	chroma_std_8 = []
18	spec_bw_var = []	38	mfccs_mean_7 = []	58	chroma_mean_1 = []	78	chroma_std_9 = []
19	rolloff_mean = []	39	mfccs_mean_8 = []	59	chroma_mean_2 = []	79	chroma_std_10 = []
20	rolloff_std = []	40	mfccs_mean_9 = []	60	chroma_mean_3 = []	80	chroma_std_11 = []

Table 5. Extracted feature values and labels from the sample records

Sl. No	File_Name	144_1b1_Al_sc_aug_3.wav	201_1b2_Ar_sc_aug_9.wav	125_1b1_Tc_sc_a_ug_27.wav	150_1b2_Al_sc_a_ug_2.wav	221_2b3_Lr_mc_LittC2S_E_aug_1.wav
1	chroma_cens_mean	0.26506	0.27453	0.25834	0.27745	0.27594
2	chroma_cens_std	0.11436	0.08927	0.12882	0.07973	0.0848
3	chroma_cens_var	0.01308	0.00797	0.0166	0.00636	0.00719
4	melspectrogram_mean	0.6469	0.07583	1.39133	2.88862	0.99446
5	melspectrogram_std	7.67458	0.94867	24.7201	34.8482	13.5884
6	melspectrogram_var	58.89925	0.899968	611.0847	1214.395	184.6454
7	mfcc_mean	-9.48785	-13.6381	-4.5474	-2.44224	-6.56647
8	mfcc_std	141.9098	159.665	115.2643	128.0599	139.7678
9	mfcc_var	20138.39	25492.9	13285.86	16399.33	19535.04
10	mfcc_delta_mean	0.121052	0.046281	0.05109	0.007315	0.006629
11	mfcc_delta_std	2.399807	2.465145	3.753295	3.847119	2.602856
12	mfcc_delta_var	5.75907	6.07694	14.0872	14.8003	6.77486
13	mfccs_mean_0	-487.79	-541.46	-366.46	-427.67	-469.72
14	mfccs_mean_1	124.148	180.208	152.144	122.285	129.259
15	mfccs_mean_2	51.6949	46.0076	12.7227	67.2487	88.7582

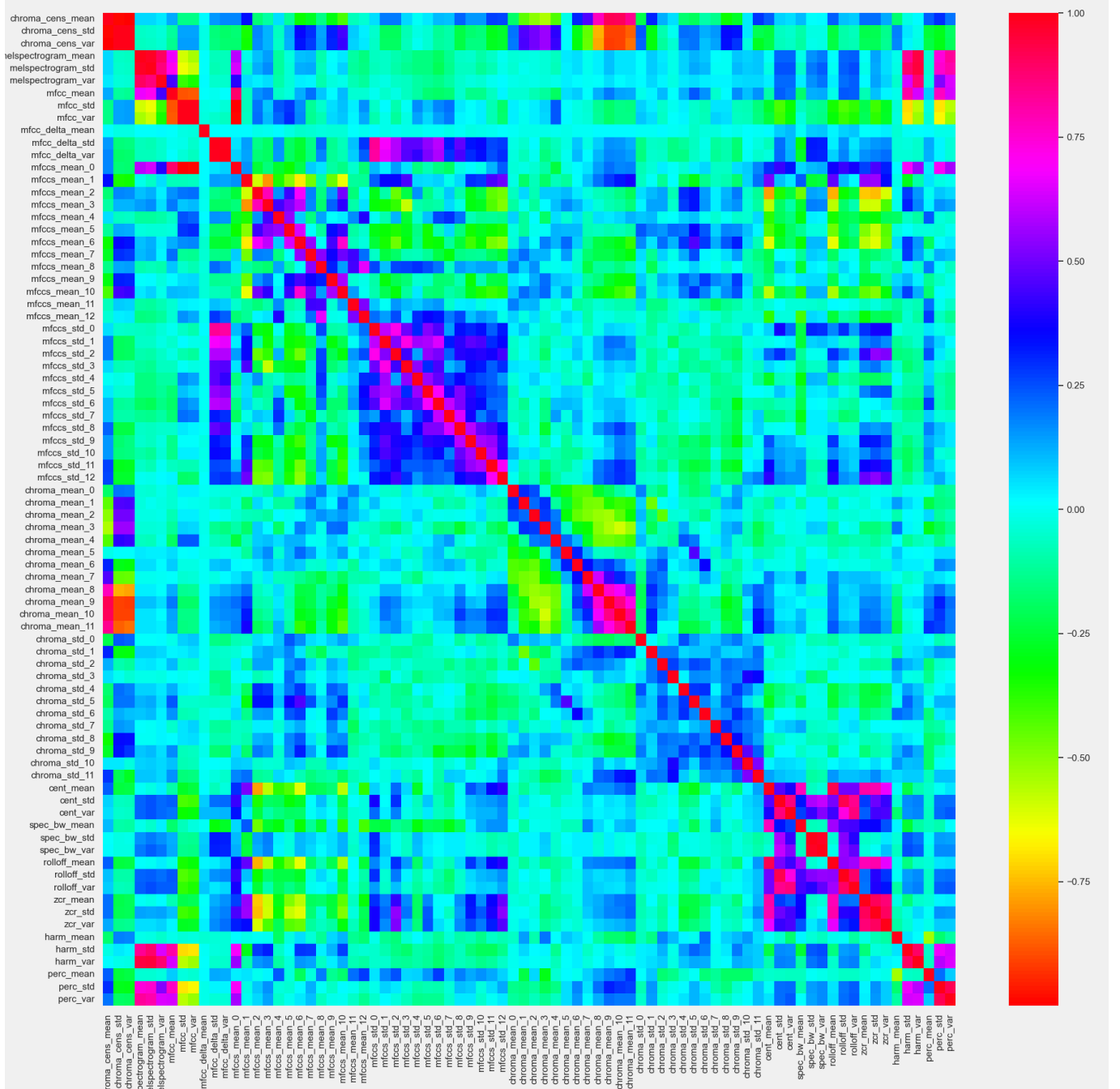


Figure 4. Heatmap for all the features

It is a basic level of audio processing which is a process of converting raw data into a numerical form. The study extracted 80 features categorized as follows:

- **Mel-Frequency Cepstral Coefficients (MFCCs):** MFCC captures the power spectrum of audio signals, mimicking human auditory perception. These features depict the envelope of the audio signal since this is the most important aspect in describing the tonal aspects of the signal. Represent the short-term power spectrum of audio signals. Calculated using:

$$MFCC_k(t) = \sum_{m=1}^M \log(S_m) \cos \left[\frac{k\pi(m-0.5)}{M} \right] \quad (4)$$

where, S_m is the mel spectrum, and M is the number of mel filters.

- **Chroma features:** Capture the harmonic structure of sounds by mapping audio frequencies onto 12 chroma bins representing musical pitches.
- **Spectral features:** Include centroid, bandwidth, and rolloff, which describe the overall shape and energy distribution of the audio spectrum.
- **Rhythm-related features:** Incorporating zero-crossing rates and harmonic/percussive properties to capture rhythmic patterns.

This systematic feature extraction ensures comprehensive representation of the audio signals, facilitating high-accuracy classification. This analysis enables the identification and potential removal of less relevant features, leading to a more efficient and interpretable model. Table 2 lists the 80 features extracted from the audio signals. Table 3 presents examples of extracted feature values and their corresponding labels. Figure 4 provides a heatmap visualization of the relationships

between all features used in the model.

Table 6 represents grouping 80 features into 9 feature set groups allows for more efficient data management and analysis by categorizing similar features together [36]. This approach enhances the interpretability of the data, making it easier to understand and visualize the relationships between different features. In addition, it can improve the performance of machine learning models by reducing redundancy and focusing on the most relevant feature subsets for specific tasks [37].

Table 6. Features are grouped

Sl. No	Grouped Features
1	df_feature_all = df3.iloc[:, 1:-1]
2	df_feature_mel_chroma = df3.iloc[:, 2:7]
3	df_feature_mfcc_mean = df3.iloc[:, 13:26]
4	df_feature_mfcc_std = df3.iloc[:, 26:39]
5	df_feature_mfcc = df3.iloc[:, 13:39]
6	df_feature_chroma_mean = df3.iloc[:, 39:51]
7	df_feature_chroma_std = df3.iloc[:, 51:63]
8	df_feature_chroma = df3.iloc[:, 39:63]
9	df_feature_csrzhp = df3.iloc[:, 63:81]

- *df_feature_all* this group comprises all extracted features that provide a summary of the audio signal. This category encompasses intricate characteristics and facilitates audio analysis across all levels. Utilizing its capacity to exhibit harmonic and spectral characteristics along with temporal variations, we constructed an extensive audio representation. A holistic approach can improve the efficacy of machine learning models by leveraging many examples and diverse attributes to feed the modelling process.
- *df_feature_mel_chroma* the integration of harmonic (Chroma CENS) and spectral (Mel-Spectrogram) features yields *df_feature_mel_chroma*, which is profound, stable, and enhances performance in music genre classification [38].
- *df_feature_mfcc_mean* comprises the average values of the Mel-Frequency Cepstral Coefficients (MFCCs), which denote the short-term power spectrum of audio. MFCCs are extensively utilized in speech recognition and music categorization due to their capacity to represent the spectrum characteristics of audio signals. This group, favoring mean values, provides an average representation of spectral features, rendering it appropriate for advanced pattern and trend recognition. A collective that enhances the efficacy of the machine learning model by offering a consistent and dependable feature collection that encapsulates essential spectral and spatial data.

$$\text{MFCC}_k(t) = \sum_{m=0}^{M-1} \log(M[m, t]) \cdot \cos\left(\frac{\pi k}{M}\left(m + \frac{1}{2}\right)\right), k = 0, 1, \dots, 12 \quad (5)$$

where, $\text{MFCC}_k(t)$ denotes the k -th Mel-Frequency Cepstral Coefficient at time frame t , $\sum_{m=0}^{M-1}$ signifies a summation across M Mel filter banks. The $\log(M[m, t])$ denotes the logarithm of the Mel filter bank output at index m and time frame t . The $\cos\left(\frac{\pi k}{M}\left(m + \frac{1}{2}\right)\right)$ represents the cosine function utilized in the Discrete Cosine Transform (DCT) and $k=0, 1, \dots, 12$ signifies that the MFCCs are calculated for k values ranging from 0 to 12.

The advantages of amalgamating 13 MFCC features into a singular feature vector are as follows. The methods utilized in generating these audio signal representations provide a more succinct and efficient form of the signal, preserving the most pertinent elements of the spectral content. It fundamentally diminishes the dimensionality of the data, facilitating processing and analysis. Mean values of the MFCCs are also calculated, which provide helpful information about the generic spectral envelope of the signal, enabling discrimination between sounds and extract patterns. The method enables the extraction of stable characteristics from dynamic and transient inputs, such as heartbeats, voice, and music, enhancing model performance in classification tasks. It also assists in reducing noise and variability, leading to a more stable and precise evaluation.

- *df_feature_mfcc_std* quantifies the variability of the spectral envelope, augmenting the mean values and enriching the feature set for the comprehensive investigation of spectral features.

$$\mu_{\text{MFCC}_i} = \frac{1}{T} \sum_t t = 1^T \text{MFCC}_i(t) \text{ for } i = 0, 1, \dots, 12 \quad (6)$$

μ_{MFCC_i} calculates the mean value (μ_{MFCC_i}) of each MFCC coefficient by averaging the MFCC values across all time frames.

- *df_feature_mfcc* integrates both mean and standard deviation values of the MFCCs, offering a thorough representation that encapsulates both the average and variability of the spectral envelope.

$$\text{mfccs_std}_i = \sqrt{\sum_{t=1}^T (\text{MFCC}_i(t) - \text{mfccs_mean}_i)^2} \quad (7)$$

for $i = 0, 1, \dots, 12$

where, (mfccs_std_i) calculates the standard deviation of all Mel-Frequency Cepstral Coefficient (MFCC) values across all time frames is given here. The term $\frac{1}{T}$ is the normalization factor, where T is the total number of time frames, $\sum_{t=1}^T (\text{MFCC}_i(t) - \text{mfccs_mean}_i)^2$ aggregates the squared differences between the i -th MFCC values and their mean (mfccs_mean_i) across all frames.

- *df_feature_chroma_mean* encapsulates the harmonic and melodic elements of the audio, offering a consistent and dependable feature set for discerning overarching patterns in music information retrieval tasks. It integrates mean and standard deviation data to create a more comprehensive feature set, facilitating an in-depth investigation of spectral features.
- *df_feature_chroma_std* measures the variability of the harmonic content, complementing the mean values and enhancing the feature set for detailed analysis of the harmonic characteristics and also includes the standard deviation values of the MFCCs, which measure the variability of the spectral envelope.

$$\text{chroma_mean}_i = \frac{1}{T} \sum_t t = 1^T \text{Chroma}_i(t) \quad (8)$$

for $i = 0, 1, \dots, 11$

- *df_feature_chroma* amalgamates the mean and standard

deviation of the Chroma features, offering a thorough representation that encapsulates both the average and variability of the harmonic content.

- *df_feature_csrzhp* offers a comprehensive spectral analysis, including spectral centroid, bandwidth, roll-off, zero crossing rate, and harmonic and percussive content, encapsulating the audio signal's intricacies and dynamics. It integrates both temporal and frequency domains, yielding a comprehensive analysis of the sound's spectral attributes, including spectral centroid, bandwidth, roll-off, harmonic and percussive content, and zero crossing rate, which elucidate the subtleties of the audio signal.

$$\text{csrzhp}(x) = \{C_s(x), B_w(x), R_f(x), Z_r(x), H_p(x)\} \quad (9)$$

This *csrzhp*(*x*) is the grouping of these spectral feature extractions applied to the input audio signal *x*.

Where C_s is a Spectral Centroid, B_w is a Spectral Bandwidth, R_f is a Spectral Roll-off, Z_r is a Zero Crossing Rate and H_p is a Harmonic and Percussive Content. This equation integrates multiple spectral and temporal aspects into a unified feature vector, offering an extensive description of the audio signal's attributes.

5. CHALLENGES

One of the significant problems is the variability of the respiratory sounds that are physiological and may depend on age, gender, and other factors. Also, there is no specific way of data gathering and the data itself is quite limited and not very diverse, which in its turn influences the creation of better models. In view of the above clinical need for rapid and precise identification, the researchers need the development of fast and non-invasive diagnostic systems that can operate in situations of high noise and variability [39-41].

6. RESULTS AND DISCUSSIONS

In this research article, a framework is presented meticulously for comparing the effectiveness of several machine learning algorithms in classifying lung sounds. A complex experimental design was developed, utilizing several classifiers, including Decision Trees, Random Forest, and Support Vector Machines. To assess the impact of feature representation, various feature sets were constructed based on extracted features from audio signals, including Mel-Frequency Cepstral Coefficients, chroma features, and other acoustic features. The model was trained and tested using different training and testing data splits (0.2, 0.3, 0.4, and 0.5) to evaluate model stability.

The feature sets (MFCC Mean, MFCC Mean+Std, and Chroma Mean+Std) also yielded good results; however, the most comprehensive feature set, "All Features, MFCC (Mean+Std), achieved the highest accuracy. Tables 7-10 present the average accuracy for All Features, MFCC Mean, MFCC Mean+Std, and Chroma Mean+Std at a split ratio of 0.20. Table 11 shows the accuracy, F1-score, precision, recall, and ROC scores for these feature groups, with All Features and MFCC (Mean_std) achieving good accuracies.

The combination of MFCC Mean+Std features with KNN and SVM classifiers achieved the highest accuracy (99%). This performance can be attributed to the robust feature representation of MFCCs, which effectively capture spectral properties crucial for distinguishing respiratory sounds.

Compared to state-of-the-art methods:

- This approach demonstrates superior performance while maintaining computational efficiency, making it suitable for deployment on edge devices.

Comparative Analysis:

- CNNs: Achieved 97% accuracy in prior studies but at a higher computational cost.
- Proposed Model: Demonstrated state-of-the-art accuracy while being more efficient.

Table 7. ML models average accuracy with split ratio 0.20

Sl. No	Feature	All	Mel +Chroma	MFCC mean	MFCC std	MFCC (mean+std)	Chroma mean	Chroma std	Chroma (mean+std)	c+s+r+z+h+p
1	Decision Tree	0.98	0.98	0.98	0.97	0.98	0.90	0.83	0.92	0.98
2	Random Forest	1.00	0.99	1.00	1.00	1.00	0.98	0.90	0.98	0.99
3	Gradient Boost	0.99	0.98	0.98	0.99	1.00	0.935	0.84	0.93	0.99
4	XGBoost	1.00	0.99	0.99	0.99	1.00	0.98	0.91	0.97	1.00
5	Ada Boost	0.99	0.90	0.97	0.93	0.96	0.83	0.72	0.87	0.98
6	Extra Tree	1.00	0.99	1.00	1.00	0.99	0.98	0.94	0.99	0.99
7	K-Neighbors	0.98	0.84	0.99	0.98	0.99	0.97	0.93	0.97	0.92
8	Support Vector	0.99	0.64	0.97	0.96	0.99	0.95	0.90	0.98	0.87
9	Gaussian Naïve	0.62	0.54	0.76	0.69	0.79	0.58	0.66	0.67	0.57
10	Multi-layer	0.99	0.65	0.98	0.98	0.99	0.94	0.84	0.96	0.91
11	Logistic Reg	0.96	0.58	0.77	0.65	0.84	0.59	0.68	0.70	0.78
	Maximum	1.00	0.99	1.00	1.00	1.00	0.98	0.94	0.99	1.00
	Average accuracy	0.95	0.82	0.94	0.92	0.96	0.87	0.83	0.90	0.91

Table 8. ML models average accuracy with split ratio 0.30

Sl. No	Feature	All	Mel +Chroma	MFCC mean	MFCC std	MFCC (mean+std)	Chroma mean	Chroma std	Chroma (mean+std)	c+s+r+z+h+p
1	Decision Tree	0.99	0.96	0.97	0.96	0.97	0.93	0.79	0.90	0.98
2	Random Forest	1.00	0.98	1.00	1.00	1.00	0.97	0.89	0.97	0.99
3	Gradient Boost	1.00	0.97	0.98	0.99	1.00	0.94	0.82	0.93	0.99
4	XGBoost	0.99	0.98	0.99	0.99	0.99	0.97	0.93	0.97	0.99
5	Ada Boost	0.99	0.90	0.95	0.93	0.98	0.82	0.73	0.83	0.97
6	Extra Tree	1.00	0.98	1.00	1.00	1.00	0.98	0.93	0.99	0.99

7	K-Neighbors	0.99	0.82	0.99	0.98	0.99	0.96	0.93	0.96	0.91
8	Support Vector	1.00	0.62	0.95	0.94	0.99	0.94	0.89	0.97	0.86
9	Gaussian Naïve	0.60	0.53	0.73	0.68	0.77	0.57	0.65	0.67	0.56
10	Multi-layer	0.99	0.62	0.96	0.96	0.99	0.93	0.84	0.96	0.91
11	Logistic Reg	0.94	0.57	0.76	0.65	0.81	0.59	0.66	0.71	0.77
	Maximum	1.00	1.00	0.98	1.00	1.00	1.00	0.98	0.93	0.99
	Average accuracy	0.95	0.95	0.81	0.93	0.92	0.95	0.87	0.83	0.90

Table 9. ML models average accuracy with split ratio 0.40

Sl. No	Feature	All	Mel +Chroma	MFCC mean	MFCC std	MFCC (mean+std)	Chroma mean	Chroma std	Chroma (mean+std)	c+s+r+z+h+p
1	Decision Tree	0.98	0.97	0.97	0.94	0.98	0.87	0.79	0.88	0.98
2	Random Forest	1.00	0.98	0.99	0.99	0.99	0.97	0.89	0.98	0.99
3	Gradient Boost	0.99	0.98	0.98	0.98	0.99	0.92	0.83	0.93	0.99
4	XGBoost	0.99	0.98	0.99	0.99	0.99	0.96	0.91	0.96	0.99
5	Ada Boost	0.99	0.92	0.95	0.93	0.97	0.81	0.74	0.83	0.95
6	Extra Tree	0.99	0.98	1.00	0.99	1.00	0.99	0.93	0.99	0.99
7	K-Neighbors	0.98	0.80	0.98	0.97	0.99	0.96	0.93	0.96	0.90
8	Support Vector	1.00	0.61	0.94	0.94	0.98	0.94	0.88	0.97	0.85
9	Gaussian Naïve	0.58	0.52	0.71	0.66	0.74	0.58	0.65	0.66	0.54
10	Multi-layer	0.99	0.625	0.95	0.96	0.99	0.91	0.81	0.96	0.88
11	Logistic Reg	0.94	0.57	0.76	0.65	0.81	0.61	0.66	0.71	0.75
	Maximum	1.00	1.00	0.98	1.00	0.99	1.00	0.99	0.93	0.99
	Average accuracy	0.95	0.95	0.81	0.93	0.91	0.95	0.86	0.82	0.89

Table 10. ML models average accuracy with split ratio 0.50

Sl. No	Feature	All	Mel +Chroma	MFCC mean	MFCC std	MFCC (mean+std)	Chroma mean	Chroma std	Chroma (mean+std)	c+s+r+z+h+p
1	Decision Tree	0.97	0.96	0.98	0.94	0.97	0.88	0.77	0.87	0.97
2	Random Forest	1.00	0.98	0.99	0.99	0.99	0.96	0.89	0.97	0.99
3	Gradient Boost	0.99	0.97	0.98	0.97	0.99	0.90	0.82	0.92	0.99
4	XGBoost	0.99	0.97	0.99	0.99	0.99	0.95	0.90	0.95	0.99
5	Ada Boost	0.98	0.91	0.95	0.92	0.97	0.80	0.72	0.82	0.96
6	Extra Tree	0.99	0.98	1.00	0.99	0.99	0.98	0.93	0.98	0.99
7	K-Neighbors	0.97	0.79	0.97	0.96	0.97	0.94	0.93	0.95	0.88
8	Support Vector	0.99	0.62	0.93	0.92	0.98	0.92	0.87	0.97	0.85
9	Gaussian Naïve	0.60	0.51	0.71	0.66	0.74	0.59	0.65	0.68	0.54
10	Multi-layer	0.99	0.61	0.93	0.93	0.99	0.90	0.80	0.95	0.87
11	Logistic Reg	0.94	0.57	0.75	0.66	0.82	0.62	0.65	0.72	0.77
	Maximum	1.00	1.00	0.98	1.00	0.99	0.99	0.98	0.93	0.98
	Average accuracy	0.95	0.95	0.81	0.93	0.90	0.95	0.86	0.81	0.89

Table 11. ML models average accuracies on All features, MFCC mean, MFCC mean+std, chroma mean+std

Sl. No	Classifiers	All_feature				MFCC mean				MECC (mean+std)				Chroma (mean+std)			
		acc	f1	prec	recall	acc	f1	prec	recall	acc	f1	prec	recall	acc	f1	prec	recall
1	Decision Tree	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.99	0.99	0.83	0.82	0.85	0.81
2	Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.91	0.91	0.91	0.90
3	Gradient Boost	0.99	0.99	1.00	0.99	0.98	0.98	0.99	0.98	1.00	1.00	1.00	1.00	0.84	0.83	0.85	0.82
4	XGBoost	1.00	1.00	1.00	1.00	0.99	0.99	1.00	0.99	1.00	1.00	1.00	1.00	0.91	0.91	0.93	0.89
5	Ada Boost	0.99	0.99	0.99	0.99	0.97	0.96	0.98	0.95	0.96	0.96	0.97	0.96	0.72	0.72	0.71	0.73
6	Extra Tree	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.94	0.94	0.94	0.93
7	K-Neighbors	0.98	0.98	1.00	0.97	0.99	0.99	1.00	0.98	0.99	0.99	1.00	0.99	0.93	0.92	0.98	0.87
8	Support Vector	0.99	0.99	1.00	0.99	0.97	0.96	1.00	0.94	0.99	0.99	1.00	0.98	0.90	0.90	0.90	0.90
9	Gaussian Naive	0.62	0.47	0.72	0.35	0.76	0.74	0.76	0.72	0.79	0.78	0.78	0.79	0.66	0.62	0.67	0.59
10	Multi-layer	0.99	0.99	1.00	0.99	0.96	0.98	1.00	0.93	0.99	1.00	1.00	0.99	0.84	0.86	0.85	0.84
11	Logistic Reg	0.96	0.96	0.97	0.94	0.77	0.77	0.76	0.77	0.84	0.83	0.85	0.81	0.68	0.66	0.67	0.64

Error Analysis

Confusion matrices reveal that misclassifications primarily occurred in overlapping classes (e.g., Crackles+Wheezes). Augmentation mitigated errors in minority classes.

To compare the classification results across different feature sets and machine learning models, a cross-comparison of tables was performed. This matrix compares the performance of different features and models, including Decision Trees and

Random Forest, using features such as All Features, MFCC mean, MFCC mean+std, see Figures 5-7. The average accuracy results across different models, features, and split ratios suggest that the best-performing model utilizes all features and a split ratio of 0.20. The study [4] achieved 97% accuracy using CNNs, while our approach achieved 99% with KNN and SVM, demonstrating superior performance with lower computational costs.

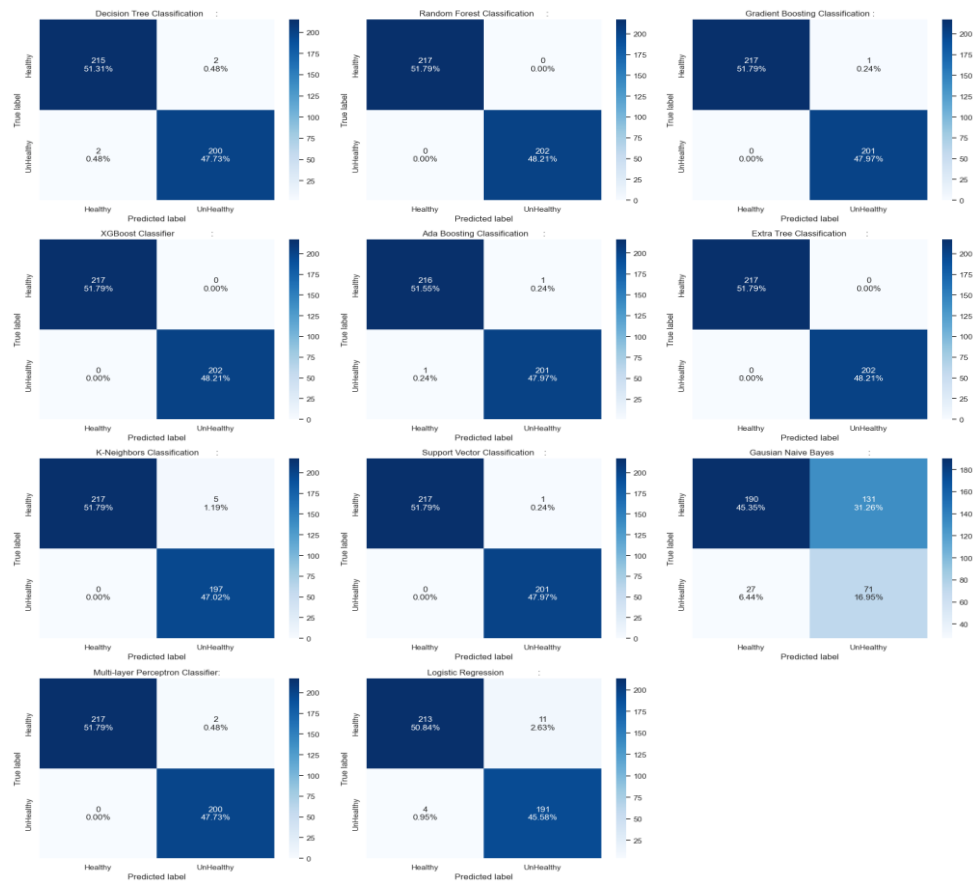


Figure 5. Cross-matrix representation of ML models using all features (df_feature_all) group

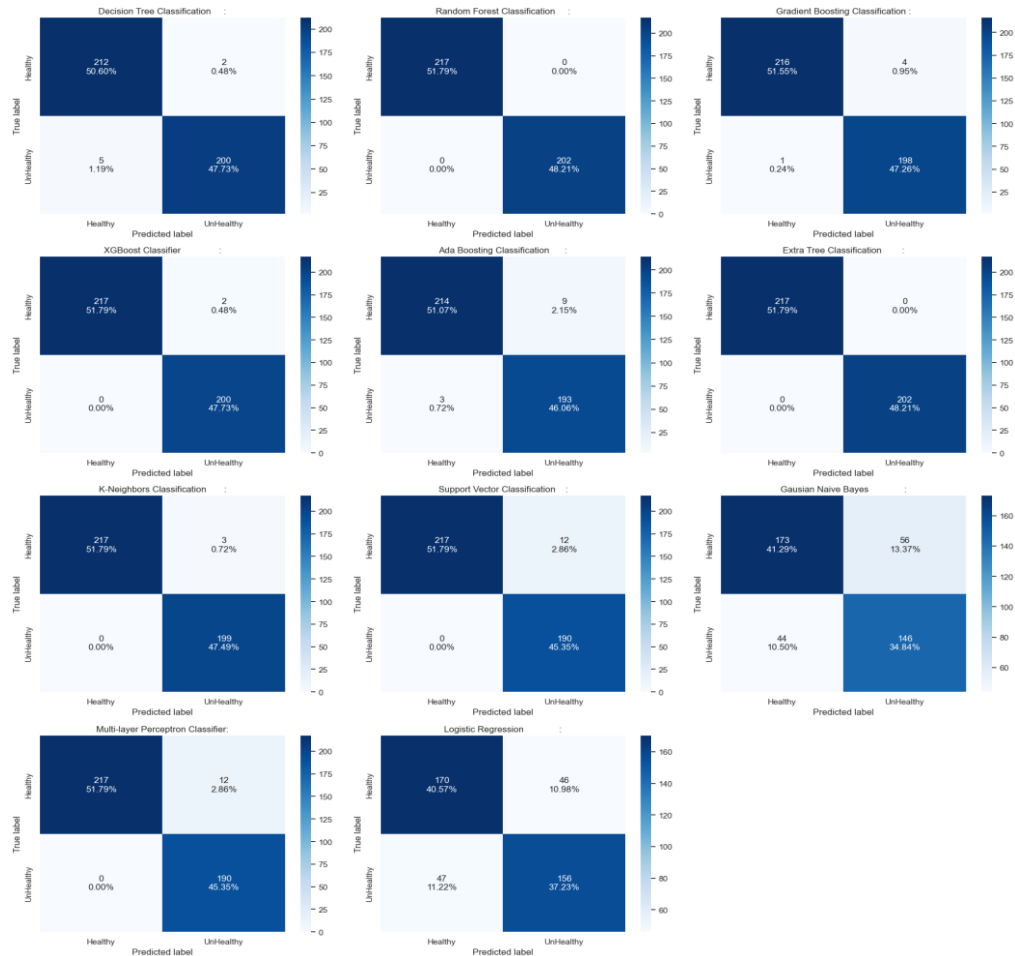


Figure 6. Cross-matrix representation of ML models using MFCC mean features

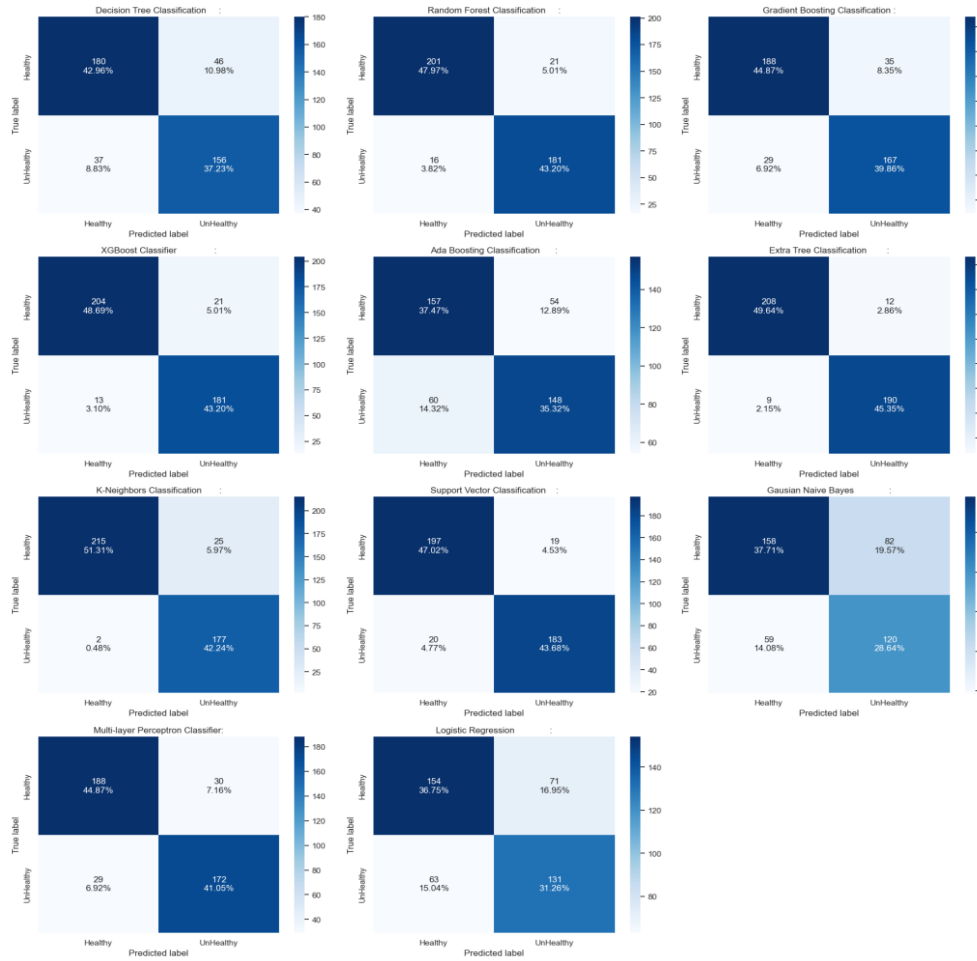


Figure 7. Cross-matrix representation of ML models using Chroma(mean+std) features

7. HYPERPARAMETER TUNING

This paper details our attempt to utilize machine learning to classify respiratory sounds, referencing the ICBHI 2017 Respiratory Sound Database. This work is intended to develop a robust and computationally efficient model to enhance diagnosis accuracy and objectivity, considering the worldwide public health crisis related to respiratory disorders and the limitations of classical auscultation. Our research addressed a significant issue in audio datasets, class imbalance, by applying various data augmentation techniques, including loudness adjustment, masking, shifting, and speed alteration. Hyperparameter tuning was implemented to improve classification, emphasizing attributes such as MFCC Mean, MFCC, and Chroma-Mean_std. This method prioritized high predicted accuracy with reduced computing demands, unlike deeper learning models that tend to be more computationally intensive. Tuning hyperparameters such as solver, penalty, C, and class_weight enhances the accuracy of our model, as demonstrated in Table 12, which pertains to MFCC Mean features.

Hyperparameter tuning was performed using a grid search approach:

A grid search was performed for KNN (neighbors: 1–15), SVM (kernel: linear, RBF; C: 0.1–10), Gradient Boosting: Learning rate (0.01–0.1) and maximum depth (3–10).

The performance plateau observed in some classifiers (e.g., Logistic Regression) is attributed to:

Model Complexity: Simpler models are less capable of

capturing the nuanced relationships in high-dimensional feature spaces.

Dataset Size: Limited samples in minority classes may restrict the ability of certain classifiers to generalize.

Initial results were encouraging, but multiple aspects must be explored to use it accurately. While the ICBHI dataset is a valuable resource, the diversity of respiratory sounds may not fully reflect real-world complexities. Future research may address this issue through a larger, more heterogeneous dataset to improve generalizability. Moreover, an experiment with different combinations of features and machine learning models might yield even better classification performance. Validation of clinical utility requires direct comparison with established diagnostic capabilities of trained medical professionals. To conclude, further work should be carried out on seamless integration into the clinical workflow to understand its real-world impact on patient care and diagnostic efficiency.

Table 12 compares different ML models with MFCC Mean features on a 0.20 split. KNN, SVM, and Gradient Boosting classifiers' accuracy increased significantly with hyperparameter tuning. However, Logistic Regression showed limited improvement due to its inability to model non-linear relationships in high-dimensional feature spaces. Accuracy is improved even tried with the best hyperparameters (C: 10, penalty: l1, solver: liblinear), chosen from solver: [newton-cg, lbfgs, liblinear], penalty: [l1, l2, elasticnet], C: [1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100].

Table 12. Hyperparameter tuning with MFCC mean features

Models	Logistic Regression	KNN	SVM Classifier	Random Forest	Gradient Boost	XGB Classifier
Best hyperparameters	'C': 10, 'penalty': 'l1', 'solver': 'liblinear'	'algorithm': 'auto', 'metric': 'euclidean', 'n_neighbors': 1, 'weights': 'uniform'	'C': 100, 'gamma': 'scale', 'kernel': 'rbf'	'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 100	'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 100, 'subsample': 0.5	'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 1000, 'subsample': 0.5
Model Accuracy	0.74	1.00	1.00	1.00	1.00	1.00
ROC-AUC	0.74	1.00	1.00	1.00	1.00	1.00
Precision	0.75	1.00	1.00	1.00	1.00	1.00
Recall	0.74	1.00	1.00	1.00	1.00	1.00
F1	0.74	1.00	1.00	1.00	1.00	1.00

8. CONCLUSION

This study demonstrates the effectiveness of machine learning techniques in respiratory sound classification, achieving state-of-the-art achieving 99% accuracy with efficient feature-classifier combinations. However, limitations include:

1. Dataset Scale: Limited size and imbalance restrict the generalizability of findings.
2. Feature Selection: While 80 features were tested, deeper exploration of advanced features could enhance performance.

This research article investigated the application of machine learning to classify respiratory sounds, utilizing the ICBHI 2017 Respiratory Sound Database as a benchmark dataset. Recognizing the global health challenge posed by respiratory diseases and the limitations of traditional auscultation, this work aimed to develop a robust and computationally efficient model for enhanced diagnostic accuracy and objectivity. The study addressed the critical issue of class imbalance inherent in audio datasets by employing various data augmentation techniques, including loudness control, masking, shifting, and speed alteration. Additionally, classification parameters were tuned, and features found to be optimal, e.g., MFCC Mean, MFCC, Chroma-Mean_std, were used for classification. This method emphasized high accuracy with little computational cost, which was a possible edge over computationally expensive deep learning technique.

Preliminary results showed promising accuracy rates, but some areas need further exploration. The ICBHI dataset is helpful but does not entirely encompass all true variability of respiratory sounds. It would be helpful for future studies to include more extensive and more heterogeneous datasets to improve the generalizability of the results. Also, re-evaluating the feature sets and learning models could help us to improve the classification even more. The clinical utility of an automated approach must be validated with a direct comparison to the diagnostic ability of trained medical professionals. Lastly, future research must explore the implementation of this technology into everyday clinical workflows and its real-time effect on patient management and diagnosis. In Future the dataset can be Expanded with real-world recordings and incorporate deep learning models for richer feature extraction and improved classification accuracy.

REFERENCES

- [1] World Health Organization. (2022). Global health

- statistics report 2022. Geneva: World Health Organization.
- [2] Sun, Z. (2023). ICBHI 2017 challenge. Medicine, Health and Life Sciences. <https://doi.org/10.7910/dvn/ht6pki>
- [3] Meng, F., Shi, Y., Wang, N., Cai, M., Luo, Z. (2020). Detection of respiratory sounds based on wavelet coefficients and machine learning. IEEE Access, 8: 155710-155720. <https://doi.org/10.1109/ACCESS.2020.3016748>
- [4] Aykanat, M., Kılıç, Ö., Kurt, B., Saryal, S. (2017). Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing, 2017: 1-9. <https://doi.org/10.1186/s13640-017-0213-2>
- [5] Yang, R., Lv, K., Huang, Y., Sun, M., Li, J., Yang, J. (2023). Respiratory sound classification by applying deep neural network with a blocking variable. Applied Sciences, 13(12): 6956. <https://doi.org/10.3390/app13126956>
- [6] Xia, T., Han, J., Mascolo, C. (2022). Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues. Experimental Biology and Medicine, 247(22): 2053-2061. <https://doi.org/10.1177/15353702221115428>
- [7] Asatani, N., Kamiya, T., Mabuchi, S., Kido, S. (2021). Classification of respiratory sounds using improved convolutional recurrent neural network. Computers & Electrical Engineering, 94: 107367. <https://doi.org/10.1016/j.compeleceng.2021.107367>
- [8] Roy, A., Satija, U. (2023). RDLNet: A novel lightweight inception network for respiratory disease classification using lung sounds. IEEE Transactions on Instrumentation and Measurement, 72: 4008813. <https://doi.org/10.1109/TIM.2023.3292953>
- [9] Altan, G., Kutlu, Y., Allahverdi, N. (2019). Deep learning on computerized analysis of chronic obstructive pulmonary disease. IEEE Journal of Biomedical and Health Informatics, 24(5): 1344-1350. <https://doi.org/10.1109/jbhi.2019.2931395>
- [10] Petmezas, G., Cheimariotis, G.A., Stefanopoulos, L., Rocha, B., Paiva, R.P., Katsaggelos, A.K., Maglaveras, N. (2022). Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. Sensors, 22(3): 1232. <https://doi.org/10.3390/s22031232>
- [11] Sen, I., Saraclar, M., Kahya, Y.P. (2015). A comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds. IEEE Transactions on Biomedical Engineering, 62(7): 1768-1776. <https://doi.org/10.1109/tbme.2015.2403616>

- [12] Fraiwan, M., Fraiwan, L., Alkhodari, M., Hassanin, O. (2022). Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. *Journal of Ambient Intelligence and Humanized Computing*, 13: 4759-4771. <https://doi.org/10.1007/s12652-021-03184-y>
- [13] Bae, S., Kim, J.W., Cho, W.Y., Baek, H., Son, S., Lee, B., Yun, S.Y. (2023). Patch-mix contrastive learning with audio spectrogram transformer on respiratory sound classification. *arXiv preprint arXiv:2305.14032*. <https://doi.org/10.21437/interspeech.2023-1426>
- [14] Ntalampiras, S. (2023). Explainable Siamese neural network for classifying pediatric respiratory sounds. *IEEE Journal of Biomedical and Health Informatics*, 27(10): 4728-4735. <https://doi.org/10.1109/jbhi.2023.3299341>
- [15] Lal, K.N. (2023). A lung sound recognition model to diagnoses the respiratory diseases by using transfer learning. *Multimedia Tools and Applications*, 82(23): 36615-36631. <https://doi.org/10.1007/s11042-023-14727-0>
- [16] Chen, C.H., Huang, W.T., Tan, T.H., Chang, C.C., Chang, Y.J. (2015). Using k-nearest neighbor classification to diagnose abnormal lung sounds. *Sensors*, 15(6): 13132-13158. <https://doi.org/10.3390/s150613132>
- [17] Wanasinghe, T., Bandara, S., Madusanka, S., Meedeniya, D., Bandara, M., Díez, I.D.L.T. (2024). Lung sound classification with multi-feature integration utilizing lightweight CNN model. *IEEE Access*, 12: 21262-21276. <https://doi.org/10.1109/access.2024.3361943>
- [18] McLane, I., Lauwers, E., Stas, T., Busch-Vishniac, I., Ides, K., Verhulst, S., Steckel, J. (2021). Comprehensive analysis system for automated respiratory cycle segmentation and crackle peak detection. *IEEE Journal of Biomedical and Health Informatics*, 26(4): 1847-1860. <https://doi.org/10.1109/jbhi.2021.3123353>
- [19] Lee, C.S., Li, M., Lou, Y., Dahiya, R. (2022). Restoration of lung sound signals using a hybrid wavelet-based approach. *IEEE Sensors Journal*, 22(20): 19700-19712. <https://doi.org/10.1109/jsen.2022.3203391>
- [20] Jin, F., Krishnan, S., Sattar, F. (2011). Adventitious sounds identification and extraction using temporal-spectral dominance-based features. *IEEE Transactions on Biomedical Engineering*, 58(11): 3078-3087. <https://doi.org/10.1109/tbme.2011.2160721>
- [21] Tripathy, R.K., Dash, S., Rath, A., Panda, G., Pachori, R.B. (2022). Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform. *IEEE Sensors Letters*, 6(5): 1-4. <https://doi.org/10.1109/lens.2022.3167121>
- [22] Abushakra, A., Faezipour, M. (2013). Acoustic signal classification of breathing movements to virtually aid breath regulation. *IEEE Journal of Biomedical and Health Informatics*, 17(2): 493-500. <https://doi.org/10.1109/jbhi.2013.2244901>
- [23] Vardhan, B.V., Geetha, M.K., Prasad, G.S., Kumar, C.S.K. (2024). Abnormal sound detection in lungs using vest-coat stethoscope using deep learning algorithm. *Explainable Artificial Intelligence in Healthcare Systems*, Nova Publisher, USA, pp. 125-140. <https://doi.org/10.52305/GOMR8163>
- [24] Zhou, G., Liu, C., Li, X., Liang, S., Wang, R., Huang, X. (2024). An open auscultation dataset for machine learning-based respiratory diagnosis studies. *JASA Express Letters*, 4(5): 052001. <https://doi.org/10.1121/10.0025851>
- [25] Zulfikar, R., Majeed, F., Irfan, R., Rauf, H.T., Benkhelifa, E., Belkacem, A.N. (2021). Abnormal respiratory sounds classification using deep CNN through artificial noise addition. *Frontiers in Medicine*, 8: 714811. <https://doi.org/10.3389/fmed.2021.714811>
- [26] Altan, G., Kutlu, Y. (2020). RespiratoryDatabase@ TR (COPD Severity Analysis). In *Mendeley Data V1*. <https://doi.org/10.17632/p9z4h98s6j.1>
- [27] Hsu, F.S., Huang, S.R., Huang, C.W., Cheng, Y.R., Chen, C.C., Hsiao, J., Lai, F. (2021). An update on a progressively expanded database for automated lung sound analysis. *arXiv preprint arXiv:2102.04062*. <https://doi.org/10.48550/arXiv.2102.04062>
- [28] PixSoft. The R.A.L.E. Repository. <http://www.rale.ca>
- [29] Rocha, B.M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y.P., De Carvalho, P. (2019). An open access database for the evaluation of respiratory sound classification algorithms. *Physiological Measurement*, 40(3): 035001. <https://doi.org/10.1088/1361-6579/ab03ea>
- [30] Shuvo, S.B., Ali, S.N., Swapnil, S.I., Hasan, T., Bhuiyan, M.I.H. (2020). 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*, 25(7): 2595-2603. <https://doi.org/10.1109/jbhi.2020.3048006>
- [31] Demirci, B.A., Koçyiğit, Y., Kızıllırmak, D., Havlucu, Y. (2022). Adventitious and normal respiratory sound analysis with machine learning methods. *Celal Bayar University Journal of Science*, 18(2): 169-180. <https://doi.org/10.18466/cbayarfbe.1002917>
- [32] Kim, Y., Hyon, Y., Jung, S.S., Lee, S., Yoo, G., Chung, C., Ha, T. (2021). Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Scientific Reports*, 11(1): 1-11. <https://doi.org/10.1038/s41598-021-96724-7>
- [33] Nguyen, T., Pernkopf, F. (2022). Lung sound classification using co-tuning and stochastic normalization. *IEEE Transactions on Biomedical Engineering*, 69(9): 2872-2882. <https://doi.org/10.1109/tbme.2022.3156293>
- [34] Rocha, B.M., Filos, D., Mendes, L., Vogiatzis, I., Perantoni, E., Kaimakamis, E., Maglaveras, N. (2018). A respiratory sound database for the development of automated classification. In *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece*, pp. 33-37. https://doi.org/10.1007/978-981-10-7419-6_6
- [35] Tasar, B., Yaman, O., Tuncer, T. (2022). Accurate respiratory sound classification model based on piccolo pattern. *Applied Acoustics*, 188: 108589. <https://doi.org/10.1016/j.apacoust.2021.108589>
- [36] Wang, Z., Sun, Z. (2024). Performance evaluation of lung sounds classification using deep learning under variable parameters. *EURASIP Journal on Advances in Signal Processing*, 2024(1): 51. <https://doi.org/10.1186/s13634-024-01148-w>
- [37] Park, J.S., Kim, K., Kim, J.H., Choi, Y.J., Kim, K., Suh, D.I. (2023). A machine learning approach to the development and prospective evaluation of a pediatric

- lung sound classification model. *Scientific Reports*, 13(1): 1289. <https://doi.org/10.1038/s41598-023-27399-5>
- [38] Dimoulas, C.A. (2016). Audiovisual spatial-audio analysis by means of sound localization and imaging: A multimedia healthcare framework in abdominal sound mapping. *IEEE Transactions on Multimedia*, 18(10): 1969-1976. <https://doi.org/10.1109/tmm.2016.2594148>
- [39] Shkel, A.A., Kim, E.S. (2019). Continuous health monitoring with resonant-microphone-array-based wearable stethoscope. *IEEE Sensors Journal*, 19(12): 4629-4638. <https://doi.org/10.1109/JSEN.2019.2900713>
- [40] Huang, D.M., Huang, J., Qiao, K., Zhong, N.S., Lu, H.Z., Wang, W.J. (2023). Deep learning-based lung sound analysis for intelligent stethoscope. *Military Medical Research*, 10(1): 44. <https://doi.org/10.1186/s40779-023-00479-3>
- [41] Mukherjee, H., Sreerama, P., Dhar, A., Obaidullah, S.M., Roy, K., Mahmud, M., Santosh, K.C. (2021). Automatic lung health screening using respiratory sounds. *Journal of Medical Systems*, 45: 1-9. <https://doi.org/10.1007/s10916-020-01681-9>