IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 May 2024, accepted 15 June 2024, date of publication 21 June 2024, date of current version 28 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3417183

RESEARCH ARTICLE

Enhancing Lung Acoustic Signals Classification With Eigenvectors-Based and Traditional Augmentation Methods

NASEEM BABU[1], DAYANANDA PRUTHVIRAJA[2], (Senior Member, IEEE), AND JIMSON MATHEW[1], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, Bihar 801106, India
[2]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

Corresponding author: Dayananda Pruthviraja (dayananda.p@manipal.edu)

ABSTRACT Identifying lung sound signal patterns is essential for detecting and monitoring respiratory diseases. Existing approaches for analyzing respiratory sounds need domain specialists. Therefore, an accurate and automated lung sound classification tool is required. In this paper, we have developed an automatic diagnostic system to classify these signals. It can support healthcare systems in low-resource environments with limited resources and a shortage of qualified medical professionals. This paper presents an eigenvectors-based data augmentation method to enhance the detection rate of automatic diagnostic systems. This proposed method provides noise-free data samples with the principal components that capture the most significant variations in the data. In the classification process, various machine learning-based classifiers are employed along with spectrogram-based features.

INDEX TERMS Lung sound signals, classification, eigenvectors, classifier models, respiratory disease, feature extraction.

I. INTRODUCTION

Respiratory diseases are a global health concern and have become the third leading cause of death worldwide. According to the Global Impact of Respiratory Disease report [1], the COVID-19 pandemic has resulted in the loss of more than 5.7 million lives in just 24 months, and most of them suffered from respiratory diseases. However, these respiratory diseases ranked among the top 10 global causes of death. According to the World Health Organization, approximately 3.2 million people die yearly due to five major respiratory diseases, including lung cancer, tuberculosis, chronic obstructive pulmonary disease (COPD), acute lower respiratory tract infection (LRTI), and asthma. Respiratory diseases have a profound impact on both individuals and healthcare systems. Therefore, early detection, diagnosis, and treatment of respiratory diseases are essential to reduce these deadly disease's impacts. The conventional diagnosis process used by healthcare professionals and general practitioners involves the subjective assessment of the respiratory system through chest auscultation. This process requires listening to lung sounds to determine whether respiratory sounds are normal or adventitious (like stridor, wheezing, rhonchi, and crackles), which are critical clinical problems. These Adventitious sounds can be distinguished from normal ones by considering their energy, frequency, intensity, pitch, timbre, and musicality [2]. Hence, assessing lung sounds is essential for recognizing a wide range of respiratory diseases and distinguishing between their chronic and non-chronic aspects [3], [4]. The automated detection of respiratory diseases has been the subject of comprehensive primary research using a range of machine learning and deep learning methods [5], [6], [7], [8]. Auscultation is an easy, cost-effective, and non-invasive procedure. In this procedure, medical professionals use a stethoscope to listen to chest sounds and identify lung abnormalities.

Several studies about automated respiratory disease detection have been investigated to address these issues [9], [10]

The associate editor coordinating the review of this manuscript and approving it for publication was Zihuai Lin.

with various feature extraction techniques have been used in this context, encompassing statistical features, wavelet coefficients [11], entropy, and spectrogram-based features [12]. In [13] used a neural network-based framework for respiratory disease detection, and in [14] used hidden Markov models for detecting cough in continuous sound signals. Shuvo et al. [15] introduce a lightweight convolutional neural network (CNN) framework for the detection of respiratory diseases based on individual breath cycles. This architecture utilizes a combination of scalogram-based features extracted from lung sounds. In [16], a hybrid convolutional and recurrent neural network-based model for classifying respiratory sounds with Mel-spectrograms is presented. A patient-specific model-tuning framework was also introduced to screen respiratory patients and create individualized classification models. In [17] analyzing breathing patterns recorded during non-invasive forced oscillation lung function tests across six groups. Detection of wheezes with wavelet using higher-order spectral features involves the application of the continuous wavelet transform (CWT) presented in [18], and Monge-Álvarez et al. [19] using Hu moments as a feature vector for cough detection. In [20] introduces a machine-learning framework for robust cough detection in audio data suitable for deployment in mobile scenarios. Sankur et al. [21] detecting pathological and healthy conditions using autoregressive (AR) models. In [22], a contrastive learning method was introduced to integrate extra out-of-class data into the model, and the classifying of respiratory sounds with an incremental supervised neural network presented in [23]. Azarbarzin and Moussavi [24] proposed an automated and unsupervised extraction of snore sounds from respiratory sound signals. Respiratory disease detection has been addressed with machine learning classifiers in [25], including binary class classification (Normal and Abnormal) or (fake and real detection) [26]. In [27] proposed a CNN-based deep learning model for ternary class classification. The author expanded the work by combining advanced deep learning models, including recurrent neural network (RNN) with Mel-frequency cepstral coefficient (MFCC) features. In [27], a six-class abnormality classification was performed employing an RNN architecture coupled with substantial preprocessing. In [28], the DenseNet169 CNN model was used with optimized preprocessing techniques for respiratory sound classification. In [10], ResNet18 with focal loss was used. The authors of [9] introduced a CNN framework combined with MFCCs feature vector. Additionally, in [29], a short-time Fourier transform with ResNet18 was used and in [30], a two-level ensemble model combining machine learning classifiers and spectrogram features. Mussell [31] provided a case report on respiratory abnormalities detection in pneumothorax and pleural via sound visualization and quantification. The respiratory sound signal datasets often suffer from imbalances and limited sizes. To address these issues, state-of-the-art methods utilize traditional data augmentation techniques, including noise addition, time shifting, time stretching, and pitch shifting [32], [33]. However, it's important to note that while these augmentation techniques aim to increase dataset size, they may also introduce distortions that impact the signal quality, resulting in a noisier representation of the original signal. This paper presents a new eigenvectors-based data augmentation method for respiratory sound signal classification to preserve the original signal quality. This method aims to improve classification accuracy and efficiently tackle challenges associated with imbalanced and small datasets.

The major research contributions (RC) of this paper can be summarized as follows:

- ***RC01:*** This paper introduced a new eigenvectors-based data augmentation method to enhance respiratory sound signals' data quality and classification performance.
- ***RC02:*** The performance of the proposed method is evaluated using a range of spectrogram-based features. A comparative analysis is conducted, comparing the proposed method to traditional approaches across various performance metrics using machine learning classifiers.

The rest of the paper is structured as follows: In Section II, discuss the datasets, and in Section III, discuss data preprocessing and feature extraction. Section IV discusses traditional and proposed augmentation methods and comparisons among those. Section V summarizes the framework used in this paper. Section VI presents a detailed performance analysis. Finally, Section VII provides a conclusion and future directions for the paper.

II. DATASETS

This paper examines the performance of various machine learning classifiers, employing both traditional and proposed methods, on two publicly available respiratory sound datasets: SPRSound 2022 [34], [35] and ICBHI 2017 [36].

A. SPRSOUND 2022

The SPRSound 2022 dataset comprises 2,683 samples containing 9,089 recorded respiratory sound signals from 292 individuals, resulting in a total duration of 8.2 hours. These recordings include male and female patients aged one month to 18 years, and the signals with low signal quality were identified as Poor Quality [10], [28], [29], [30], while the high-quality respiratory sound signals were categorized as Normal, Continuous Adventitious Sounds (CAS), Discontinuous Adventitious Sound (DAS), and CAS & DAS. The CAS sounds are continuous abnormal breath sounds, such as wheezes and stridor. These sounds usually have a distinct, repetitive pattern and are often associated with obstructed airways. DAS are irregular, non-continuous sounds like crackles, which can be fine or coarse. These recorded respiratory sound signals are divided into two categories event and record level. The respiratory cycles at the event level have the labels such as Normal, Fine Crackle, Coarse Crackle, Wheeze, Rhonchi, Stridor, and Wheeze & Crackle. The number of data samples for each event is as follows: Normal (6,887), Fine Crackle (1,167), Coarse Crackle (66), Wheeze (865), Rhonchi (53), Stridor (17), and

Wheeze & Crackle (34) [9]. The sound signals at the record level have labels such as Normal, Poor Quality, CAS, DAS, and CAS & DAS records, with sample counts of 1,785, 187, 233, 347, and 131, respectively.

B. ICBHI 2017

The ICBHI 2017 dataset contains 920 data samples collected from 126 patients, categorized into Normal, Crackle, Wheeze, and both (Crackle & Wheeze). The number of respiratory sound signals is 6,898, with 1,864 cycles including both Crackle and Wheeze, 886 cycles having both Crackle & Wheeze simultaneously, and the remaining sounds labeled as Normal [36].

III. PREPROCESSING AND FEATURE EXTRACTION

A. PREPROCESSING

1) DENOISING AND RESAMPLING

During the auscultation process, respiratory sound signals are often contaminated with background noise, making it crucial to remove this noise to extract meaningful information. The recorded respiratory sound signals frequency ranges from $50Hz$ to $2500Hz$. A 5^{th} order band-pass filter is used to remove background noise and maintain the purity of the signal. This filter has lower and upper-frequency cutoffs set at $50Hz$ and $2500Hz$, respectively [34]. The recorded respiratory sound signals are segmented according to the start and end times of the respiratory cycles. These segmented respiratory sound signals have a fixed duration of five seconds and are sampled at a rate of $8000Hz$ [9].

B. FEATURES EXTRACTION

Feature extraction is essential for capturing crucial information from raw data. However, selecting the most appropriate features can be a challenging task. This study used Short-Time Fourier Transform (STFT) [37], Mel Frequency Cepstral Coefficients (MFCCs) [38], [39], Mel Spectrogram, and Log Mel Spectrogram [40] to extract features for respiratory sound signals. The choice of STFT is because of its ability to provide a comprehensive time-frequency representation, capturing the dynamic changes in respiratory sounds over time; MFCCs are included due to their widespread success in audio processing, particularly in capturing features that mimic human auditory perception, which is crucial for distinguishing between various respiratory conditions. The Mel Spectrogram was selected to align with the human auditory system's frequency sensitivity, offering a detailed view of energy distribution across key frequency bands relevant to respiratory sounds, and the Log Mel Spectrogram was employed to enhance the detection of slight variations in sound intensity, leveraging the logarithmic scale to represent the loudness perception of human hearing better. The choice of these features is because they can capture both the temporal and spectral characteristics of respiratory sound signals, thereby improving the accuracy and robustness of our proposed method.

1) SHORT-TIME FOURIER TRANSFORM (STFT)

It takes a short-duration signal segment, computing its Fourier transform, and converting the signal from the time to the frequency domain [37]. STFT of a discrete-time signal $x\,(n)$ is described as:

$$Y\,[r,k] = \sum_{n} x\,[n]\,w\,[n-r]\,e^{-\frac{j2\pi nk}{f_c}} \tag{1}$$

where $Y\,[r,k]$ is a discrete function of time and frequency, $w\,[n]$ is finite duration window, and f_c is the number of discrete frequency channels.

2) MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCCS)

It is the most successfully used representation in the speech and speaker recognition field [38], [39]. Discrete Fourier transform (DFT) is employed to compute the spectrum $Z\,[p]$, and the magnitude of $Z\,[p]$ is weighted by the sequence of the F filter frequency. The real cepstral connected with the resultant energies $E\,[p]$ is guided to as the MFCC and is calculated for the provided frame as:

$$c\,[q] = \sum_{p=0}^{F-1} \log\left(E\,[p]\right) \cos\left(q\,(p+0.5)\,\frac{\pi}{F}\right) \tag{2}$$

where $q = 0, 1, 2, \ldots, F-1$, the first element $q\,[0]$ is discarded because it denotes the mean value of the analyzed frame signal.

3) MEL-SPECTROGRAM

It is used to convert the frequencies into a Mel scale [40]. Convert f hertz into m mels with

$$\begin{aligned} m &= 2595 \cdot \log_{10}\left(1+\frac{f}{700}\right) = 1127\ln\left(1+\frac{f}{700}\right) \\ f &= 700\left(10^{\frac{m}{2595}}-1\right) = 700\left(e^{\frac{m}{1127}}-1\right) \end{aligned} \tag{3}$$

4) LOG-MEL SPECTROGRAM

After normalizing the Mel spectrogram, we scaled the signal to the filter bank value and converted the dB unit. This function is denoted by the Log-Mel spectrogram [40]. The formula is given as follows:

$$10log_{10}\left(\frac{s}{ref}\right) \tag{4}$$

IV. AUGMENTATION METHODS

Augmentation is a process that generates synthetic data by making minor modifications to the original data samples, thereby expanding the dataset size. In this study, augmentation is performed in two ways: the first method utilizes traditional techniques such as noise addition, time shifting, time stretching, and pitch shifting, while the second method is based on eigenvectors.

A. TRADITIONAL METHODS

The traditional methods for augmentation involved minor modifications to the original data samples. The augmentation

is directly applied to 1D respiratory sound signals by adding Gaussian noise, performing time shifting and stretching, and altering the pitch of the signals. The augmented samples were combined with the original data samples, after which feature extraction was performed. The preprocessing, augmentation and feature extraction process are shown in Figure 1 for augmented data samples; similarly, the features are extracted for original data samples.

1) NOISE ADDITION

In this process, Gaussian noise is introduced into the data samples. The nature of this noise is random, but it aligns with the principles of Gaussian distribution [33]. The probability density function for this distribution is given as follows:

$$N(\chi : \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{5}$$

The random noise (*noise*) is calculated from the $N(\chi : \mu, \sigma)$ and added to the original data samples with some multiplication factor p. This paper keeps this p value at 0.002; suppose an original discrete-time signal $x(n)$ after adding the noise, the new discrete time signal $y(n)$.

$$\begin{aligned} y(n) &= x(n) + p \times noise \\ y(n) &= x(n) + 0.002 \times noise \end{aligned} \tag{6}$$

2) TIME SHIFTING

The time-shifting of the discrete time signal results in time advance or time delay. Suppose we have a discrete-time signal $x(n)$, and we want to shift this signal left or right by n_0 units [32]. The signal is shifted to the right if n_0 is positive and represents the delayed signal; conversely, if n_0 is negative, the signal is shifted to the left, shifting the advance signal in the time. The new audio signal $y(n)$ can be mathematically expressed as:

$$y(n) = x(n \pm_0) \tag{7}$$

Here, the audio signal is shifted to the right with $n_0 = 1600$, and this is a delayed version of the original audio signal.

$$y(n) = x(n - 1600) \tag{8}$$

3) TIME STRETCHING

Time stretching is a process that allows for the modification of the duration or speed of an audio signal without affecting the original audio signal. This paper uses a stretching factor of 0.98, which means that if the stretching factor exceeds one, it accelerates the audio. If it falls below one, it decelerates the audio signal while maintaining fidelity. It can be mathematically expressed as:

$$t_{new} = \frac{t_{old}}{v} \tag{9}$$

t_{new} is the time duration of the output signal, t_{old} is the time duration of the input signal, and v is the relative speed of the audio signal (how many times slower or speed up the audio signal, here v is set as 0.98).

4) PITCH SHIFTING

This process alters the pitch of an audio signal without changing its speed. The adjustment is made using a step value ranging from -5 to 5. In this paper, a step value of 2 is used [32], [33].

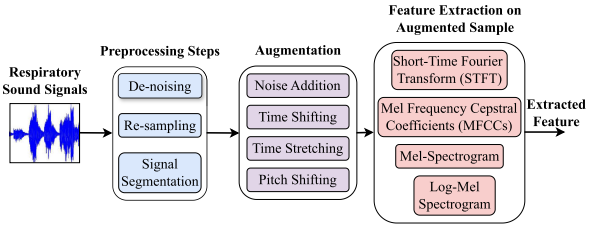

FIGURE 1. The preprocessed 1D signal is used for the data augmentation, and then types of features are extracted.

B. PROPOSED METHOD

The proposed method presents eigenvector-based data augmentation using principal component analysis (PCA) for respiratory sound signals classification. Although the proposed method in this paper is well-established in the literature for dimensionality reduction [41], [42], [43], [44] [45], [46], [47]. In this study, PCA eigenvectors are utilized to augment data samples. To increase the dataset size and improve the performance and robustness of classifier models. This approach offers a new perspective on enhancing data quality. At first, these respiratory sound signal is converted into a feature vector using Mel Frequency Cepstral Coefficients (MFCCs), Short-Time Fourier Transform (STFT), Log-Mel Spectrogram, and Mel Spectrogram. The dimension of the feature vector is represented as $A \times B$, where A is the number of time frames, and B is the number of features (e.g., frequency components). This feature vector is passed to the PCA to identify the principal components or eigenvectors l that capture more variations in the data and remove the fewer variance components (noisy components) from the feature matrix. Selecting only a subset of the principal components (those explaining most of the variance in the data) and reconstructing the feature matrix with these selected components using inverse PCA transformation. Now, consider this reconstructed feature vector as a new data sample in the transformed domain and add it to the original feature vector. Figure 2 shows the proposed approach's workflow.

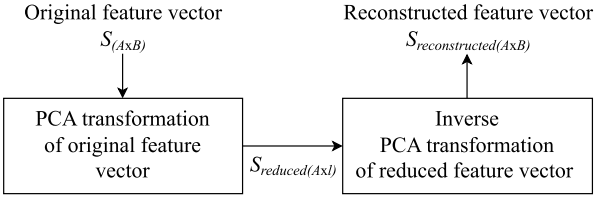

FIGURE 2. Before applying PCA to the 1D respiratory sound signal, we first transformed it into a 2D vector $S_{(A\times B)}$ using various features.

Let's consider calculating MFCCs for a respiratory signal, and it generates a feature vector S with a $A \times B$ size.

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1B} \\ S_{21} & S_{22} & \dots & S_{2B} \\ \vdots & \vdots & \ddots & \vdots \\ S_{A1} & S_{A2} & \dots & S_{AB} \end{bmatrix} \tag{10}$$

mean (μ_j) calculation for a specific column j is:

$$\mu_j = \frac{1}{A} \sum_{i=1}^{A} S_{ij} \tag{11}$$

standard deviation (σ_j) calculation for a specific column j is:

$$\sigma_j = \sqrt{\frac{1}{A} \sum_{i=1}^{A} (S_{ij} - \mu_j)^2} \tag{12}$$

Center the data by subtracting the mean from each data point for every feature:

$$S_{centered} = \begin{bmatrix} \frac{(S_{11}-\mu_1)}{\sigma_1} & \frac{(S_{12}-\mu_2)}{\sigma_2} & \dots & \frac{(S_{1B}-\mu_B)}{\sigma_B} \\ \frac{(S_{21}-\mu_1)}{\sigma_1} & \frac{(S_{22}-\mu_2)}{\sigma_2} & \dots & \frac{(S_{2B}-\mu_B)}{\sigma_B} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(S_{A1}-\mu_1)}{\sigma_1} & \frac{(S_{A2}-\mu_2)}{\sigma_2} & \dots & \frac{(S_{AB}-\mu_B)}{\sigma_B} \end{bmatrix} \tag{13}$$

Calculate the covariance matrix (C_{cov}) of the centered data:

$$C_{cov} = \frac{1}{2} S_{centered}^T \cdot S_{centered} \tag{14}$$

eigenvalue decomposition on the C_{cov} to obtain the eigenvalues (λ_i) and eigenvectors (V_i)

$$C_{cov} \cdot V_i = \lambda_i \cdot V_i \tag{15}$$

sort all the eigenvalues in descending order ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_B$). The elbow method is used to identify the eigenvector number l that captures the most significant variance while having essential information.

To obtain the reduced-dimensional representation, project the centered data onto the selected l eigenvectors:

$$S_{reduced} = S_{centered} \cdot V_l \tag{16}$$

where $S_{reduced}$ is the reduced data matrix, $S_{centered}$ is the centered data matrix, and V_l contains the top l eigenvectors. To reconstruct the data, multiply the reduced data ($S_{reduced}$) by the selected eigenvectors (V_l^T) scale it back by multiplying by (σ_j), and add back the mean (μ_j) that was subtracted in the first step for every feature j:

$$S_{reconstructed} = (S_{reduced} \cdot V_l^T) \cdot \sigma_j + \mu_j \tag{17}$$

$S_{reconstructed(A \times B)}$ have the same shape as the original data matrix $S_{(A \times B)}$. Add the reconstructed data samples $S_{reconstructed(A \times B)}$ to the final feature vector list to create more data samples in the transformed domain. Figure 3 illustrates the feature vector and its reconstruction.

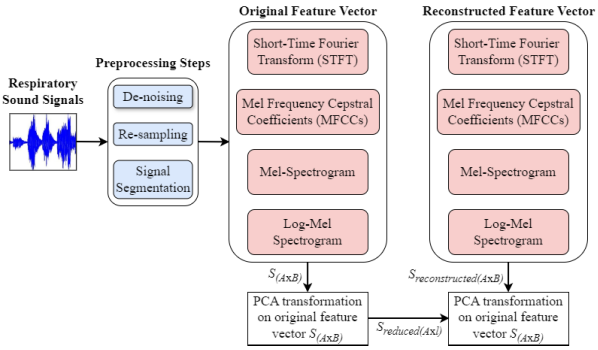

FIGURE 3. Represents the PCA and inverse PCA transformation for preprocessed respiratory sound signal in the transformed domain.

C. PROPOSED METHOD AND TRADITIONAL METHODS

Sound signals have unique characteristics, including frequency patterns and temporal dependencies. The augmented data samples generated with the help of eigenvectors are noise-free, enhancing signal quality. In contrast, augmentation using traditional methods introduces noise and may not closely resemble the original sample. Figure 5 illustrates the impact of both the proposed and traditional methods. Notably, the spectrogram generated by the proposed method, shown in Figure 4(b), is noise-free and closely resembles the original sound signal spectrogram in Figure 4(a). In comparison, traditional methods exhibit noise distribution and a loss of the original signal's characteristics, as shown in Figures 4(c, d, e, and f).

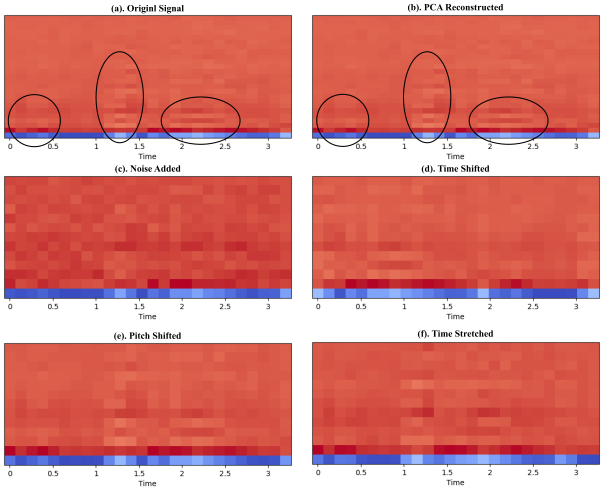

FIGURE 4. The figure illustrates the influence of the eigenvector-based augmentation and traditional methods.

V. PROPOSED FRAMEWORK

The proposed framework consists of several stages. In the initial stage, raw respiratory sound signals undergo preprocessing, as detailed in the preprocessing section above. This preprocessing step ensures the signals are clean and suitable for further analysis. The next stage involves data

augmentation and feature extraction. Augmentation is performed using both traditional methods and eigenvector-based techniques to enhance signal quality. The extracted features are used as input for classifier models. In the final stage, we rigorously evaluate the performance of these classifier models to determine their effectiveness. Figure 5 illustrates the architecture of the proposed framework, showcasing each stage of the process in detail.

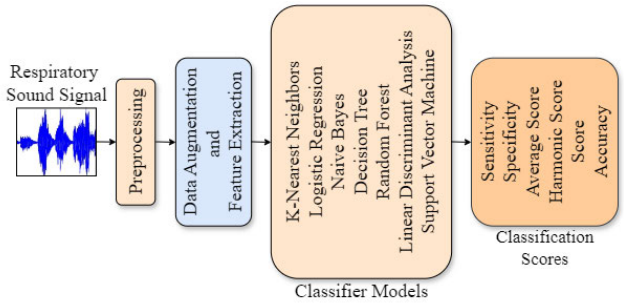

FIGURE 5. The proposed framework illustrates the process from raw signal preprocessing to the classification of respiratory sound signals.

VI. PERFORMANCE ANALYSIS AND RESULTS

This section analyzes the classification performance of various classifier models across different scenarios. We examine how each model performs under varying conditions and datasets, providing a detailed comparison to identify the most effective approaches for classifying respiratory sound signals.

A. CLASSIFICATION TASKS

The respiratory sound signals are broadly divided into two levels: record and event level. Record level refers to the entire recording of respiratory sound, which may include several breathing cycles and different types of adventitious sounds. Analysis at this level aims to capture global features of the respiratory condition. At the same time, the event level focuses on specific segments within the recording, such as individual breaths or particular types of adventitious sounds (e.g., wheezes and crackles). Analysis at this level aims to identify and characterize discrete events within the respiratory cycle. In both the datasets SPRSound 2022 and ICBHI 2017, these record and event-level respiratory sound signals are further categorized into two sub-categories such as Task11 and Task12 for event-level classification and Task21 and Task22 for record-level classification. A detailed description of each task corresponding to each dataset is provided in Table 1.

B. CLASSIFIER MODELS

The machine learning classifier models, including support vector machine (SVM) [39], [48], logistic regression (LR) [48], k-nearest neighbor (KNN) [48], naive Bayes (NB) [48], random forest (RF) [48], decision tree (DT) [48], and linear discriminant analysis (LDA) [48] were trained and evaluated with various performance measure metrics. These models were used to evaluate the effectiveness of the proposed method and its impact on classification results.

TABLE 1. Information regarding the classification tasks.

Tasks	*Task Category*	*Diseases*
SPRSound Dataset		
Task11	Event Level	Normal, and Adventitious
Task12		Normal, Wheeze, Rhonchi, Coarse Crackle, Fine Crackle, Stridor, and Wheeze & Crackle
Task21	Record Level	Normal, Adventitious, and Poor Quality
Task22		Normal, CAS, DAS, CAS & DAS, and Poor Quality
ICBHI Dataset		
Task11	Event Level	Healthy, and Unhealthy
Task12		Healthy, Crackle, Wheeze, and Wheeze & Crackle
Task21	Record Level	Healthy, and Adventitious
Task22		Healthy, COPD, Pneumonia, URTI, LRTI, Bronchiectasis, and Bronchiolitis

C. EVALUATION METRICS

The proposed framework is evaluated with various matrics considered as Sensitivity, Specificity, Average Score, Harmonic Score, and Score [34], [35], explored by state-of-the-art methods. Consider Task22 on the SPRSound 2022 dataset [34] to gain an understanding of these metrics. In this instance, the letters CD, PQ, D, C, and N stand for the corresponding numbers of *CAS & DAS*, *Poor Quality*, *DAS*, *CAS*, and *Normal* disease, while the CD_{cd}, PQ_{pq}, D_d, C_c, and N_n denote the classification outcomes. The total number of samples represents by CD_t, PQ_t, D_t, C_t, and N_t (5)-class classification).

1) SENSITIVITY

It is defined as the ratio of correctly detected *Adventitious* samples and the total number of *Adventitious* samples, and it is calculated as follows:

$$Sensitivity\ (S_e) = \frac{CD_{cd} + PQ_{pq} + D_d + C_c}{CD_t + PQ_t + D_t + C_t} \tag{18}$$

2) SPECIFICITY

It is defined as the ratio of correctly detected *Normal* samples and the total number of *Normal* samples.

$$Specificity\ (S_p) = \frac{N_n}{N_t} \tag{19}$$

3) AVERAGE SCORE

Is the arithmetic mean of sensitivity and specificity.

$$Average\ Score\ (AS) = \frac{(S_e + S_p)}{2} \tag{20}$$

4) HARMONIC SCORE

It is the harmonic mean of sensitivity and specificity.

$$Harmonic\ Score\ (HS) = \frac{2 * S_e * S_p}{(S_e + S_p)} \tag{21}$$

5) SCORE

Score is defined as the average of the average score (AS) and harmonic score (HS).

$$Score = \frac{(AS + HS)}{2} \tag{22}$$

D. SCORE ANALYSIS

The *Score*, considered an essential performance measure metric, is used to evaluate the classifier's performance with spectrogram-based features. Notably, the decision tree (DT) classifier is better within the traditional methods, achieving the highest classification score when using the STFT feature vector for Task11, Task12, Task21, and Task22, as illustrated in Figure 6. In contrast, the proposed approach showed good results with the decision tree classifier using the MFCC feature vector for Task11, Task12, and Task21 and the STFT feature for Task22, as shown in Figure 7. Both Figures 6 and 7 show detailed information about the performance of all classifiers with the extracted features.

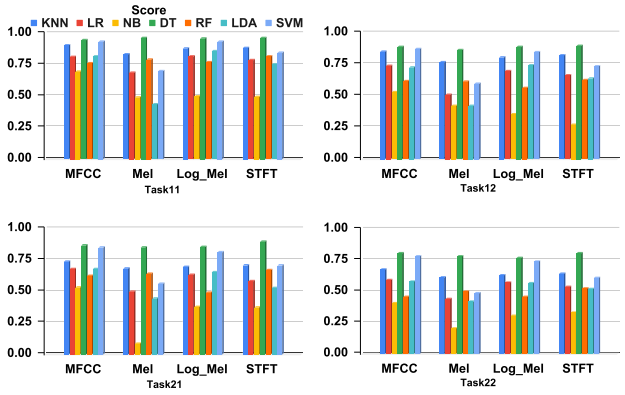

FIGURE 6. The classification score obtained with traditional methods.

E. SCORE COMPARISON

A detailed score comparison between the proposed and traditional methods is shown in Figure 8, which indicates the effect of extracted features on classifier performance. Notably, the decision tree classifier outperformed others in Task11 and Task12 when using MFCC and STFT features in the proposed approach. Similarly, for Task21 and Task22, the decision tree classifier exhibited superior performance when using MFCC features in the proposed method.

F. ANALYSIS WITH CONFUSION MATRIX

The confusion matrix is constructed for each task to evaluate class-level performance and ensure the proposed approach's effectiveness across different classes using the SPRSound

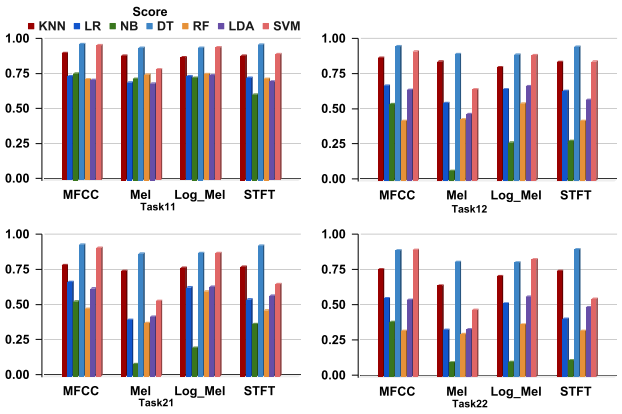

FIGURE 7. The classification score achieved using the proposed approach.

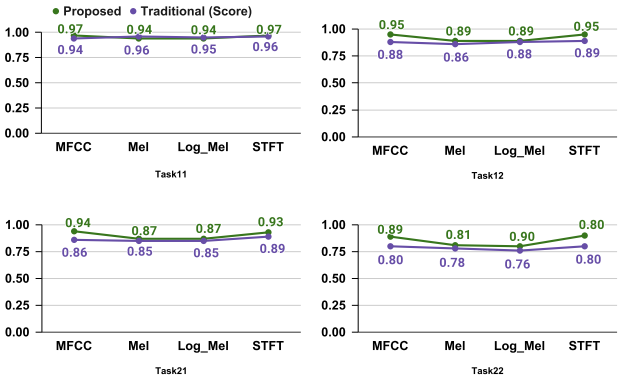

FIGURE 8. An analysis of the scores achieved by the proposed approach compared to traditional methods.

TABLE 2. Comparative analysis of classification scores using different methods.

Methods	*Task*	$S_e(\%)$	$S_p(\%)$	$AS(\%)$	$HS(\%)$	$Score(\%)$
SPRSound Dataset						
Proposed	Task11	98.13	95.25	96.69	96.67	96.68
	Task12	96.63	94.16	95.39	95.37	95.39
	Task21	95.41	93.34	94.37	94.36	94.36
	Task22	89.41	90.84	90.84	90.84	90.12
Traditional	Task11	80.54	95.80	88.17	87.50	87.83
	Task12	77.05	95.72	86.38	85.37	85.88
	Task21	75.60	91.46	83.53	82.77	83.15
	Task22	68.09	90.33	79.21	77.64	78.43
ICBHI Dataset						
Proposed	Task11	80.34	81.06	80.70	80.69	81.90
	Task12	83.11	74.38	78.74	78.50	78.62
	Task21	97.74	90.47	94.10	93.96	95.70
	Task22	96.00	100	98.00	98.00	98.00
Traditional	Task11	80.34	81.06	80.70	80.69	80.70
	Task12	73.00	80.00	76.50	76.30	76.40
	Task21	94.23	95.48	94.85	94.85	94.85
	Task22	84.12	96.63	90.37	89.94	90.15

dataset. The y-axis represents true labels in the matrix, while the x-axis represents predicted labels. The diagonal values in the confusion matrix indicate the total of correctly predicted samples. Analysis of the confusion matrix reveals that the proposed approach consistently delivers good results for all classes, as depicted in Figure 9. Since the dataset used is imbalanced, the Normal class has more data samples than

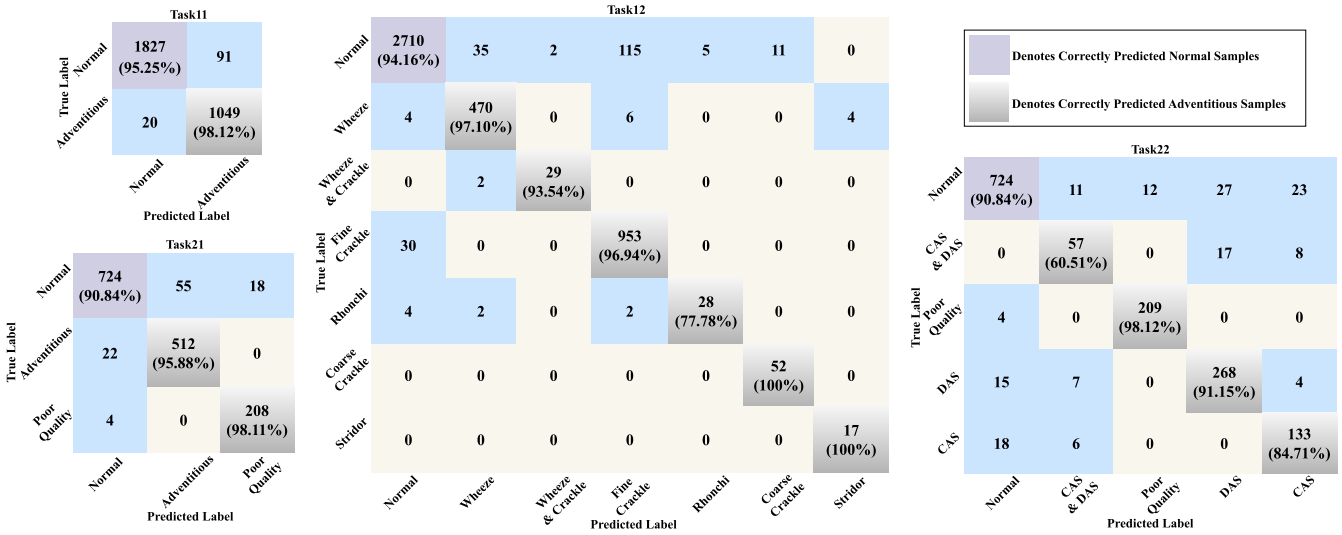

FIGURE 9. Performance analysis of the proposed approach with confusion matrix in each case, such as two-class classification (Task11), ternary classification (Task21), and multi-class classification (Task12 and Task22).

TABLE 3. Comparison with existing work.

Task	*Approach*	$S_e(\%)$	$S_p(\%)$	$AS(\%)$	$HS(\%)$	$Score(\%)$
Task11	Machine Learning Classifiers [34]	67.66	83.62	75.64	74.80	75.22
	Two-Level Ensemble Model [30]	85.30	90.20	87.70	87.70	87.70
	DenseNet169 CNN Model [28]	85.00	93.20	89.10	88.90	89.00
	ResNet18 with STFT [29]	89.00	90.00	89.00	89.00	89.00
	ResNet18 with Focal Loss [10]	94.90	88.40	91.65	89.99	91.60
	2D CNN Model with MFCC [9]	90.50	99.00	94.00	94.60	94.30
	Proposed	**98.13**	**95.25**	**96.7**	**96.66**	**96.68**
Task12	Machine Learning Classifiers [34]	45.95	82.35	64.15	58.99	61.57
	Two-Level Ensemble Model [30]	66.10	83.40	74.80	73.80	74.30
	ResNet18 with STFT [29]	68.00	94.00	81.00	79.00	80.00
	ResNet18 with Focal Loss [10]	83.70	88.40	86.05	87.20	86.00
	DenseNet169 CNN Model [28]	88.70	93.10	90.90	90.90	90.90
	2D CNN Model with MFCC [9]	92.00	97.70	94.85	94.76	94.80
	Proposed	**96.63**	**94.16**	**95.4**	**95.38**	**95.40**
Task21	Machine Learning Classifiers [34]	47.11	68.26	57.68	55.74	56.71
	Two-Level Ensemble Model [30]	77.70	65.20	71.40	70.90	71.10
	ResNet18 with STFT [29]	77.00	66.00	72.00	71.00	71.50
	ResNet18 with Focal Loss [10]	76.10	80.00	78.05	79.01	78.00
	DenseNet169 CNN Model [28]	74.50	94.30	84.40	83.20	83.80
	2D CNN Model with MFCC [9]	90.00	97.60	93.80	93.64	93.70
	Proposed	**95.41**	**93.34**	**94.37**	**94.36**	**94.36**
Task22	Machine Learning Classifiers [34]	20.66	67.43	44.04	31.63	37.84
	ResNet18 with STFT [29]	23.00	86.00	54.00	36.00	45.00
	Two-Level Ensemble Model [30]	38.40	73.30	55.90	50.40	53.10
	ResNet18 with Focal Loss [10]	52.20	80.00	66.10	72.38	64.60
	DenseNet169 CNN Model [28]	47.40	95.40	71.40	63.30	67.30
	2D CNN Model with MFCC [9]	80.00	95.90	87.95	87.23	87.60
	Proposed	**89.41**	**90.84**	**90.12**	**90.12**	**90.12**

other classes. The confusion matrix for Task11 and Task12 shows the event level classification. Task11 has two classes (Normal and Adventitious); the model can correctly predict 95.25% for the Normal class and for the Adventitious class with 98.12%. The confusion matrix for Task 12 has seven classes (Normal, Wheeze, Wheeze & Crackle, Fine Crackle, Rhonchi, Coarse Crackle, and Stridor) and denotes the lowest classification score of 77.78% for Rhonchi and the highest classification score of 100% for Stridor. The confusion matrix for Task21 and Task22 is the record level classification in which Task21 shows three classes (Normal, Adventitious, and Poor Quality) classifications with a classification score of 90.84%, 95.88%, and 98.11%. The confusion matrix for Task22 has five classes (Normal, CAS & DAS, Poor Quality, DAS, and CAS), with the lowest classification score of 60.51% for CAS & DAS and the highest of 98.12% for Poor Quality.

G. ABLATION STUDY

In this section, we evaluated the performance of the proposed approach using two datasets and compared it to traditional methods, as shown in Table 2. The results demonstrate that the scores achieved with the proposed approach are consistently higher than those obtained with existing methods. This comparison highlights the effectiveness and robustness of our approach in improving classification accuracy and reliability.

In summary, the classification results obtained from our proposed approach, using various feature extraction methods and classifiers, show significant improvement. Notably, the decision tree (DT) classifier with MFCC feature vectors excelled in Task11, Task12, Task21, and Task22 on the SPRSound 2022 dataset. Similarly, in the case of the ICBHI dataset, the decision tree (DT) classifier with MFCC feature vectors outperformed other methods. The proposed method achieved scores for different tasks, including 96.7% for binary classification Task11, 95.4% for multi-class Task12, 94.36% for ternary class Task21, and 90.12% for multi-class Task22. Compared to existing approaches [9], [10], [28], [29], [30], [34], these results represent a significant improvement of 2.4%, 0.6%, 0.66%, and 2.4%, respectively. Detailed comparisons are presented in TABLE 3.

VII. CONCLUSION

This paper presents an innovative eigenvectors-based data augmentation method for respiratory sound signal classification, leveraging principal component analysis (PCA) on spectrogram-based features. We conducted a series of experiments using machine learning classifiers to validate our approach. The comparative analysis against traditional methods demonstrated that our proposed method significantly enhances classification scores and effectiveness. One of the key challenges in our method is selecting the optimal number of eigenvectors. By employing the elbow method, we were

able to determine the number of eigenvectors that capture the most significant variance while ensuring the retention of essential information.

As part of our future research endeavors, we have already initiated efforts toward real-time implementation of our proposed approach. Our objective is to provide empirical data and practical insights that demonstrate the potential of our approach in improving healthcare delivery in underserved regions. We aim to assess our system's effectiveness and reliability in real-world conditions.

REFERENCES

[1] S. M. Levine and D. D. Marciniuk, "Global impact of respiratory disease: What can we do, together, to make a difference?" *Chest*, vol. 161, no. 5, pp. 1153–1154, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0012369222000496

[2] A. Bohadana, G. Izbicki, and S. Kraman, "Fundamentals of lung auscultation," *New England J. Med.*, vol. 370, pp. 744–751, Feb. 2014.

[3] S. A. H. Tabatabaei, P. Fischer, H. Schneider, U. Koehler, V. Gross, and K. Sohrabi, "Methods for adventitious respiratory sound analyzing applications based on smartphones: A survey," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 98–115, 2021.

[4] A. K. Majumder and S. K. Chowdhury, "Recording and preliminary analysis of respiratory sounds from tuberculosis patients," *Med. Biol. Eng. Comput.*, vol. 19, no. 5, pp. 561–564, Sep. 1981.

[5] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177926.

[6] A. H. Sfayyih, N. Sulaiman, and A. H. Sabry, "A review on lung disease recognition by acoustic signal analysis with deep learning networks," *J. Big Data*, vol. 10, no. 1, p. 101, Jun. 2023.

[7] Y. Kim, Y. Hyon, S. S. Jung, S. Lee, G. Yoo, C. Chung, and T. Ha, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Sci. Rep.*, vol. 11, no. 1, p. 17186, Aug. 2021.

[8] S. B. Sangle and C. J. Gaikwad, "COVID-19 respiratory sound signal detection using HOS-based linear frequency cepstral coefficients and deep learning," *Circuits, Syst., Signal Process.*, vol. 43, no. 1, pp. 331–347, Jan. 2024.

[9] N. Babu, J. Kumari, J. Mathew, U. Satija, and A. Mondal, "Multiclass categorisation of respiratory sound signals using neural network," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 228–232.

[10] J. Li, X. Wang, X. Wang, S. Qiao, and Y. Zhou, "Improving the ResNet-based respiratory sound classification systems with focal loss," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 223–227.

[11] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Comput. Biol. Med.*, vol. 39, no. 9, pp. 824–843, Sep. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482509001267

[12] S. Jayalakshmy and G. F. Sudha, "Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101809.

[13] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, "CNN-MoE based framework for classification of respiratory anomalies and lung disease detection," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 2938–2947, Aug. 2021.

[14] S. Matos, S. S. Birring, I. D. Pavord, and D. H. Evans, "Detection of cough signals in continuous audio recordings using hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1078–1083, Jun. 2006.

[15] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2595–2603, Jul. 2021.

[16] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 535–544, Jun. 2020.

[17] C. M. Ionescu, J. A. Tenreiro Machado, and R. De Keyser, "Analysis of the respiratory dynamics during normal breathing by means of pseudophase plots and pressure–volume loops," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 1, pp. 53–62, Jan. 2013.

[18] S. A. Taplidou and L. J. Hadjileontiadis, "Analysis of wheezes using wavelet higher order spectral features," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1596–1610, Jul. 2010.

[19] J. Monge-Álvarez, C. Hoyos-Barceló, P. Lesso, and P. Casaseca-de-la-Higuera, "Robust detection of audio-cough events using local hu moments," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 184–196, Jan. 2019.

[20] J. Monge-Álvarez, C. Hoyos-Barceló, L. M. San-José-Revuelta, and P. Casaseca-de-la-Higuera, "A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2319–2330, Aug. 2019.

[21] B. Sankur, Y. P. Kahya, E. Ç. Güler, and T. Engin, "Comparison of AR-based algorithms for respiratory sounds classification," *Comput. Biol. Med.*, vol. 24, no. 1, pp. 67–76, Jan. 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010482594900388

[22] W. Song and J. Han, "Patch-level contrastive embedding learning for respiratory sound classification," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104338. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422007923

[23] Z. Dokur, "Respiratory sound classification by using an incremental supervised neural network," *Pattern Anal. Appl.*, vol. 12, no. 4, pp. 309–319, Dec. 2009.

[24] A. Azarbarzin and Z. M. K. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1156–1162, May 2011.

[25] B. Tasar, O. Yaman, and T. Tuncer, "Accurate respiratory sound classification model based on piccolo pattern," *Appl. Acoust.*, vol. 188, Jan. 2022, Art. no. 108589. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X21006836

[26] N. Babu, P. Kumar, J. Mathew, and U. Satija, "Exploration of bonafide and spoofed audio classification using machine learning models," in *Proc. IEEE 19th India Council Int. Conf. (INDICON)*, 2022, pp. 1–6.

[27] D. Perna, "Convolutional neural networks learning from respiratory data," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 2109–2113.

[28] W.-B. Ma, X.-Y. Deng, Y. Yang, and W.-C. Fang, "An effective lung sound classification system for respiratory disease diagnosis using DenseNet CNN model with sound pre-processing engine," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2022, pp. 218–222.

[29] Z. Chen, H. Wang, C.-H. Yeh, and X. Liu, "2D CNN," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 233–237.

[30] L. Zhang, Y. Zhu, S. Tu, and L. Xu, "A feature polymerized based two-level ensemble model for respiratory sound classification," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2022, pp. 238–242.

[31] M. J. Mussell, "The need for standards in recording and analysing respiratory sounds," *Med. Biol. Eng. Comput.*, vol. 30, no. 2, pp. 129–139, Mar. 1992.

[32] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 279–283, Mar. 2017.

[33] S. Wei, S. Zou, F. Liao, and W. Lang, "A comparison on data augmentation methods based on deep learning for audio classification," *J. Phys., Conf. Ser.*, vol. 1453, no. 1, Jan. 2020, Art. no. 012085.

[34] Q. Zhang, "SPRSound: open-source SJTU paediatric respiratory sound database," *IEEE Trans. Biomed. Circuits Syst.*, vol. 16, no. 5, pp. 867–881, Oct. 2022.

[35] Q. Zhang, "Grand challenge on respiratory sound classification for SPRSound dataset," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2022, pp. 213–217.

[36] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. Pedro Paiva, I. Chouvarda, and P. de Carvalho, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Meas.*, vol. 40, no. 3, Mar. 2019, Art. no. 035001.

[37] M. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 1, pp. 55–69, Feb. 1980.

[38] C. K. On, P. M. Pandiyan, S. Yaacob, and A. Saudi, "Mel-frequency cepstral coefficient analysis in speech recognition," in *Proc. Int. Conf. Comput. Informat.*, Jun. 2006, pp. 1–5.
[39] R. Palaniappan and K. Sundaraj, "Respiratory sound classification using cepstral features and support vector machine," in *Proc. IEEE Recent Adv. Intell. Comput. Syst. (RAICS)*, Dec. 2013, pp. 132–136.
[40] Y. Choi, H. Choi, H. Lee, S. Lee, and H. Lee, "Lightweight skip connections with efficient feature stacking for respiratory sound classification," *IEEE Access*, vol. 10, pp. 53027–53042, 2022.
[41] M. Ringnér, "What is principal component analysis?" *Nature Biotechnol.*, vol. 26, no. 3, pp. 303–304, Mar. 2008.
[42] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *J. Signal Inf. Process.*, vol. 4, no. 3, p. 173, 2013.
[43] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, Jul. 2010.
[44] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Comput. Geosci.*, vol. 19, no. 3, pp. 303–342, 1993.
[45] S. M. Holland, "Principal components analysis (PCA)," Dept. Geol., Univ. Georgia, Athens, GA, USA, Tech. Rep., 2008, pp. 2501–30602.
[46] S. Xie, F. Jin, S. Krishnan, and F. Sattar, "Signal feature extraction by multi-scale PCA and its application to respiratory sound classification," *Med. Biol. Eng. Comput.*, vol. 50, no. 7, pp. 759–768, Jul. 2012.
[47] R. Remesan, M. Bray, and J. Mathew, "Application of PCA and clustering methods in input selection of hybrid runoff models," *J. Environ. Informat.*, pp. 137–152, 2018.
[48] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 204–218, 2021.

NASEEM BABU received the B.Tech. degree in information technology from AKTU University, Lucknow, Uttar Pradesh, India, and the M.Tech. degree in computer engineering from the National Institute of Technology, Jaipur, India. He is currently pursuing the Ph.D. degree in computer science and engineering with Indian Institute of Technology Patna, India. His research interests include audio signal processing, brain–computer interfaces, machine learning, deep learning, and respiratory sound signals.

DAYANANDA PRUTHVIRAJA (Senior Member, IEEE) received the B.E., M.Tech., and Ph.D. degrees in computer science and engineering from Visvesvaraya Technological University, Belagavi, India. Since 2023, he has been a Professor and the Head of the Information Technology Department, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India. In the previous assignment, he was a Professor and the Head of the Department of Information Science and Engineering, JSSATE, Bengaluru. He has published many articles in national and international journals in the field of deep learning and machine learning. He has a few research grants and conducted consultancy work. His research interests include image processing and information retrieval, machine learning, and deep learning and its applications.

JIMSON MATHEW (Senior Member, IEEE) received the master's degree in computer engineering from Nanyang Technological University (NTU), Singapore, and the Ph.D. degree in computer engineering from the University of Bristol, Bristol, U.K. Throughout his career, he has held positions at various prestigious institutions, including the Centre for Wireless Communications, National University of Singapore; Bell Laboratories Research, Lucent Technologies North Ryde, Australia; the KTH Royal Institute of Technology, Stockholm, Sweden; and the Department of Computer Science, University of Bristol. He was the Head of the Department of Computer Science and Engineering, Indian Institute of Technology Patna (IIT Patna), India. He is currently a Professor with the Computer Science and Engineering Department, IIT Patna. He has made significant contributions to the field of computer engineering and has a strong academic portfolio. He holds multiple patents, has coauthored three books, and has published more than 100 papers in renowned international journals and conferences. His expertise and contributions have had a notable impact on computer engineering. His research interests include fault-tolerant computing, hardware security, large-scale integration design, and design automation. He is a member of the Institution of Engineering and Technology (IET).