ORIGINAL PAPER

An efficient lung sound classification technique based on MFCC and HDMR

Mahmud Esad Arar[1,2] · Herman Sedef[2]

Abstract

In this work, an efficient feature extraction scheme is developed for classifying the pulmonary diseases. The proposed method is hybrid which combines two important techniques that are Mel Frequency Cepstral Coefficients (MFCC) and High-Dimensional Model Representation (HDMR). MFCC is capable of imitating the human ear; therefore, it is capable of characterizing the lung sounds acquired by a stethoscope. On the other hand, HDMR performs decorrelation and denoising to the high-dimensional data. The MFCC entries establish a two-dimensional feature matrix, which is decomposed in terms of less dimensional entities by the application of HDMR. These entities are considered feature vectors that are then fed to the relevant machine learning classification algorithms and then the overall accuracies are calculated. According to the results, the proposed algorithm achieves 97.2% classification accuracy which is competitive with other existing state-of-the-art methods in the literature. HDMR also improves significantly the classification efficiency of the proposed technique. The results emphasize that HDMR can be employed as an efficient method in recognizing pulmonary disease tasks.

1 Introduction

The vital importance of diagnosis of pulmonary diseases is obvious. The respiratory sounds provide essential information about lung and pulmonary tract health. According to the World Health Organization (WHO) reports, millions of people suffer from respiratory diseases every year [1, 2].

Auscultation is a non-invasive process of listening to body sounds through a stethoscope, which is a medical device invented by René Laennec in 1816 [2, 3]. Auscultation of lungs is an easy and inexpensive way of diagnosing and monitoring the pulmonary system's health [4].

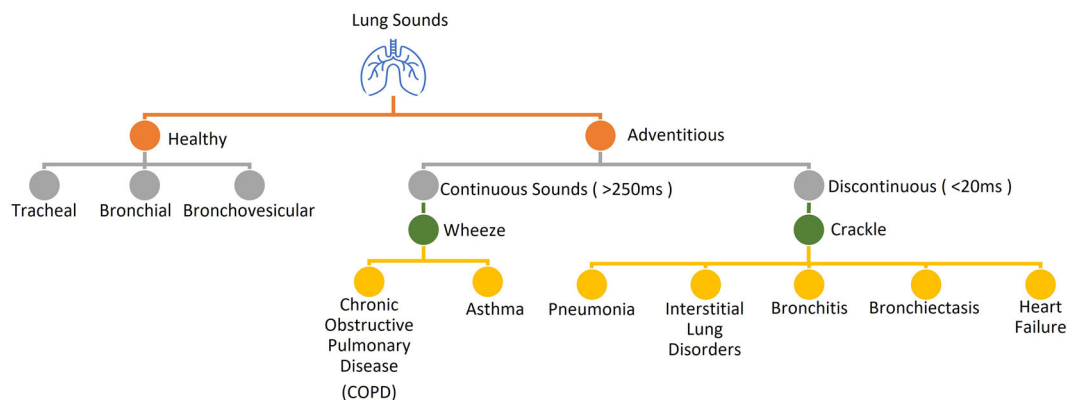Mahmud Esad Arar and Herman Sedef have contributed equally to this work.

✉ Mahmud Esad Arar
mearar@fsm.edu.tr

Herman Sedef
sedef@yildiz.edu.tr

1 Department of Biomedical Engineering, Fatih Sultan Mehmet Vakif University, Istanbul 34445, Turkey

2 Department of Electronics and Communication Engineering, Yildiz Technical University, Istanbul 34220, Turkey

Medical physicians can detect abnormal lung sounds by listening to the chest wall of subjects via manual stethoscopes and fulfilling an important role in diagnosing pulmonary system diseases. The respiratory sounds are non-stationary, cyclic but aperiodic and non-linear signals [5]. They generally range between 100 Hz and 2 kHz, whereas the human audible spectrum lies between 20 Hz and 20 kHz. Experienced medical physicians can separate (hear) normal and abnormal lung sounds with satisfactory accuracy. However, this procedure requires experience and skills that are gained through a high level of training. Surely, it is difficult to distinguish the sounds and recognize the corresponding diseases due to the lack of experience of medical physicians. Even those well-experienced ones may not be able to diagnose the disease in some difficult cases. Electronic stethoscopes and related automated classification systems present much more stable and more modest results. All research efforts aim to provide faster, more precise, and more accurate diagnoses. The purpose of automatic lung sound classification is to diagnose lung sounds, which are not easy to detect due to the conditions of both physicians and patients. Electronic hearing and recording of lung sounds and then investigating their time and frequency properties can be considered as a humble way of performing the correct diagnosis of the

Fig. 1 Types of lung sounds and their relation with pulmonary diseases

patients in an automated way [1, 6]. Lung sounds are classified into two categories as (are grouped under two headings); (i) normal (healthy) and (ii) abnormal (adventitious). Normal lung sounds are grouped as tracheal, bronchial, and bronchovesicular. On the other hand, adventitious lung sounds are mostly subdivided into two subgroups depending on their durations, which are continuous or discontinuous. Former sounds are commonly wheezes, which are musical sounds that last longer than 250 ms, while latter sounds are crackles and shorter than 20 ms (Fig. 1). Crackles are also known as non-musical and explosive sounds which are related to bronchitis, pneumonia, heart failure, COVID-19 etc. [4, 5]. Dominant frequency of crackles ranges from 200 Hz to 2 kHz, where dominant frequency of wheezes is greater than 400 Hz [1, 7].

The aim of this work is to propose a method to extract features via MFCC and HDMR methods. The MFCC is a two-dimensional feature ensemble capable of mimicking biological sounds. Because of the complexity of biological sounds and limited sensor capabilities, MFCC can contain high correlation and noise. Thus, further classification tasks can be difficult because of these drawbacks. On the other hand, the two-dimensional structure of the MFCC could lead to high computational costs during further evaluation tasks. Therefore, an efficient data decomposition technique, HDMR, is employed to extract one-dimensional decorrelated and denoised features from the MFCC matrices to reduce the computational burden and improved the distinguishability of biological signals. After the feature extraction phase, machine learning methods are employed to classify 11 types of corresponding diseases: asthma, chronic obstructive pulmonary disease (COPD), pneumonia, bronchitis, bronchiolitis, bronchiectasis, upper respiratory tract infection (URTI), lower respiratory tract infection (LRTI), heart failure, lung fibrosis, and pleural effusion, respectively. In this manuscript, it is shown that MFCC and both MFCC & HDMR feature extraction methods provide reliable classification accuracies. Nevertheless, many studies

achieved successful performance in classification and the HDMR method enables feature dimension reduction. Therefore, fast and accurate results have been achieved in contrast by exploiting high-dimensional features such as MFCC and spectrogram.

The rest of this paper is structured as follows: Related works are considered in Sect. 2. Section 3 describes the proposed method in detail, while Sect. 4 evaluates and discusses the experimental results. The manuscript concluded with a discussion of the concluding remarks.

2 Related works

In this section, we explain a brief review of classification methods for respiratory sounds that represent the vital signs of human health. The researchers have attempted various automated classification approaches to this end and have developed efficient techniques for diagnosing pulmonary abnormalities. The aim of these attempts is to assist physicians in not only monitoring, but also diagnosing the diseases when the acquired sounds are ambiguous. The automated classification methods can detect even minimal sound changes and enhance the diagnosing accuracy. The researchers also employed several Machine Learning (ML) and Deep Learning (DL) models and algorithms for distinguishing numerous diseases and categorizing different respiratory sounds. In this section, the distinctive feature extraction approaches are also mentioned, and the results of the previous studies are presented in chronological order.

Kahya et al. [8] classified healthy, restrictive, and pathological pulmonary sounds of 69 subjects by employing the autoregressive (AR) model to acquire optimal feature parameters. Multi-stage classifiers based on the K-nearest neighbor (KNN) were designed to determine the disease labels of the subjects. The average accuracy of this work was reported as 69.59%.

Bahoura [9] proposed a method in order to categorize lung sounds into two classes as normal and wheeze sounds. He submitted the comparisons of combinations of classification methods such as Gaussian mixture models (GMM), artificial neural network (ANN), and vector quantization with feature extraction methods such as MFCC, wavelet transform (WT), Fourier transform (FT) and linear predictive coding (LPC). MFCC and GMM ensemble yields the best result at 94.2% compared to other combinations.

Icer et al. [10] proposed a method, which analyzed various feature extraction methods to classify healthy, rhonchus, and crackle sounds with the help of the Support vector machine (SVM) algorithm. The corresponding feature extraction procedures were executed using three methods, namely (i) power spectrum density with Welch's method, (ii) Hilbert Huang transform and (iii) single spectrum analysis. The number of samples was 60, and the observed accuracy after the SVM adoption varies between 80 and 100%.

Palaniappan et al. [5] presented a review paper, which provides detailed information about several studies, the corresponding feature extraction methods, the classification methods, and other crucial points of the studies in the field. Various techniques were addressed such as ANN, GMM, KNN, self-organizing maps (SOMs), hidden Markov models (HMMs), genetic algorithms (GAs), fuzzy logic (FL), and classification methods of previous studies were discussed. Besides, some feature extraction methods were also depicted such as the AR model, MFFC, energy, entropy, spectral features, and wavelet.

Chen et al. [11] developed an electronic stethoscope to capture abnormal lung sounds to help physicians diagnosing the lung diseases. The MFCC method was employed for the feature extraction where the K-means algorithm was performed for reducing the amount of data for an efficient computational strategy. The KNN method was then selected to classify the corresponding respiratory sounds.

Jakovljevic et al. [12] classified the International Conference on Biomedical and Health Informatics (ICBHI) dataset employing a combination of GMM and HMM. The output labels were represented as healthy sounds, wheeze, and crackle classes.

Aykanat et al. [13] designed and produced an electronic stethoscope. The physicians recorded 17,930 lung sounds from 1630 subjects. MFCC features were classified via an SVM algorithm where the spectrogram images of lung sounds were categorized using a convolutional neural network (CNN) algorithm. Four different combinations of the dataset were classified, and the results of the SVM and the CNN algorithms were compared.

Bardou et al. [14] performed several classification methods such as SVM, KNN, GMM and CNN to categorize the lung sounds exploiting local binary pattern (LBP) characteristics extracted from the spectrograms and the MFCC

features. The CNN classifier achieved 95.56% accuracy, where the others were: (i) MFCC-SVM with 91.12%, (ii) LBP-SVM with 71.21%, (iii) MFCC-CNN with 91.67%, and (iv) LBP-CNN with 80% accuracies, respectively.

Demir et al. [15] conducted a study based on the ICBHI dataset. They adopted two deep learning-based approaches for lung sound classification. In the first approach, recordings of lung sounds were classified using an SVM method, where CNN model was utilized to extract the features. In the second approach, the spectrogram images and CNN models were employed for the classification task. The accuracies of these approaches were 65.5 and 63.09%, respectively.

Fraiwan et al. [16] employed ensemble machine learning algorithms for detecting a wide range of pulmonary diseases. In this study, models combining boosted and bagged types of decision trees and linear discriminants were compared to baseline models such as SVM, KNN, linear discriminant analysis (LDA), and basic decision trees (DT). The best result was achieved by the boosted decision trees. This work combines two datasets, which are ICBHI dataset and King Abdullah University Hospital (KAUH) dataset, respectively. The authors adopted three entropy-based algorithms for feature extraction, which are Shannon, logarithmic energy, and spectral entropy.

Rocha et al. [17] claimed that event durations have significant effects on lung sound classification performance. They utilized traditional and deep learning-based approaches to categorize adventitious respiratory sounds. Five classifiers were employed in the study, which are LDA, linear SVM, Gaussian SVM, boosted trees and CNN algorithm. The CNN algorithm takes the spectrogram images as inputs. The other classifiers were fed with other features extractions from the spectrograms and several acoustic tools. The researchers compared lung sounds in three tasks, which are (i) wheeze vs. others, (ii) crackles vs. others, (iii) wheeze vs. crackles vs. others. When the durations of the recordings were fixed, they achieved more than 90% accuracy. On the other hand, varying durations yielded accuracies lower than 90%.

A dataset and a corresponding study were proposed by Fraiwan et al. [1] after preprocessing steps and filtrations, they employed CNN and Bidirectional long short-term memory (BiLSTM) units to classify six classes. They achieved an average accuracy of 99.62%. KAUH [18] and ICBHI [17, 19, 20] datasets were combined for the lung sounds of 203 patients selected out of 238 subjects.

Engin et al. [21] attempted to classify 400 respiratory cycles of 94 subjects with the aim to achieve high accuracy with low-dimensional features by utilizing the sequential forward selection (SFS) method. MFCC, time domain features, frequency domain features and linear predictive coding (LPC) methods were employed to extract features. The LDA, KNN, SVM, and naive Bayes (NB) classification algorithms

were exploited to test the performance of the features and 90.63% accuracy was achieved by KNN algorithm.

Kwon et al. [22] suggested a novel feature extraction method named shifted delta coefficients in lower-subspace (SDC-L), which enhances feature extraction procedure by decreasing hyperspectral dimension. SDC-L classified lung sounds of ICBHI dataset. The scientists measured the performance of this method through three machine learning algorithms which were SVM, KNN, and random forest (RF). Moreover, two deep learning methods used are multilayer perceptron (MLP) and CNN. Finally, they evaluated the performance of a new method with a hybrid deep learning method that was a combination of CNN with LSTM. The researchers compared the work with previous studies and observed competitive results.

Petmezas et al. [2] categorized healthy and pathological lung sounds into four classes as normal, crackles, wheezes, and both wheezes and crackles by a combination of CNN and long short-term memory (LSTM) deep learning algorithms. They also used the focal loss function to solve the imbalance problem of the dataset. In the beginning, researchers preprocessed the data to eliminate imbalance situations by resampling, sample padding, and cropping data. The duration of lung sounds was fixed at 2.7 s and resampled at 4 kHz. The short-time Fourier transform (STFT) was applied for feature extraction. The accuracies of the study were 76.39% for 10-fold cross-validation, 74.57% for leave-one-out cross-validation and 73.69% for 60/40 splitting strategies.

Saldanha et al. [23] analyzed the effect on data augmentation of datasets by variational autoencoders (VAE). They utilized three different variational autoencoders which are the multilayer perceptron VAE, the convolutional VAE, and the conditional VAE. The aim of this study was to synthesize respiratory sounds based on the imbalanced ICBHI dataset to enhance the classification performance.

Furthermore, the first study of automated lung sound classification was proposed by Cohen et al. [24]. It was an early sample of statistical classification and an AR method. In this work, provided feature extraction methods were based on time and frequency domain features. In another early sample of automated lung sound classification, Sankur et al. [25] compared KNN (nonparametric) and quadratic (parametric) classifiers with the 6th-order AR model for the feature extraction method. In addition, Kandaswamy et al. [26] adopted wavelet transform for feature extraction and employed ANN to resolve the correct class labels. Moreover, the researchers [27] employed time-frequency and scale analysis methods to extract the high-dimensional feature sets. The extracted feature sets were then classified via three machine learning algorithms, which were SVM, KNN, and MLP, respectively. In another study, Messner et al. [28] utilized spectrogram features to compare the performance of different ANN algorithms such as MLP, Bidirectional Gated Recurrent

Neural Network (BiGRNN), and Convolutional Bidirectional Gated Recurrent Neural Network (ConvBiGRNN). In their study, researchers [3] tackled with crackles, wheezes, rhonchi, and normal respiratory sounds using deep learning algorithms and SVM. The corresponding features were extracted employing VGG16 ANN architecture fed with Mel-spectrograms. In study [7], the researchers compared the classification performance of STFT vs MFCC feature extraction methods. The performances of the classification methods were evaluated by a new algorithm named depthwise separable convolutional neural network (DSCNN).

3 The proposed method

3.1 Preprocessing

To assess the proposed method, two public datasets are employed from King Abdullah University Hospital (KAUH) and International Conference on Biomedical Health Informatics (ICBHI) 2017, respectively. Although we provide extensive information on these datasets in Sect. 4, it is worth to depict the corresponding preprocessing steps at this stage. Since our aim is to develop a feature extraction model for both datasets, we initially resampled each recording signal at 4 kHz frequency. Resampling at a frequency of 4 kHz is obligatory for the ICBHI dataset because it includes three distinct sample frequencies (44.1, 10, and 4 kHz). This resampling is necessary because it corresponds to the lowest frequency within the dataset while ensuring a satisfactory resolution. It should be noted that our study also incorporated another dataset named KAUH, which was originally sampled at 4 kHz. A visual representation of the comprehensive setup of the proposed method is illustrated in Fig. 2, which shows the detailed configuration of the system [1, 2, 16, 17]. Then, we cropped each signal to 5 s since this duration is sufficient for a respiratory cycle of an adult individual. Therefore, equal-sized 1256 sample signals were obtained for harnessing the subsequent feature extraction strategy.

3.2 Mel frequency Cepstral coefficients

The filterbank-based Mel Frequency Cepstral Coefficient (MFCC) method is one of the most widely used feature extraction approaches for audio signal processing [5, 14, 29]. The MFCC captures important information about the spectral content of a signal and is exploited in a variety of applications such as speech recognition, speaker identification, music, and respiratory sound classification [30–32].

MFCC is a well-established technique because it mimics human hearing system regarding its sensitivity to small changes in lower frequencies [22]. While the Hertz stands for the unit of linear frequencies, the Mel is the unit for the
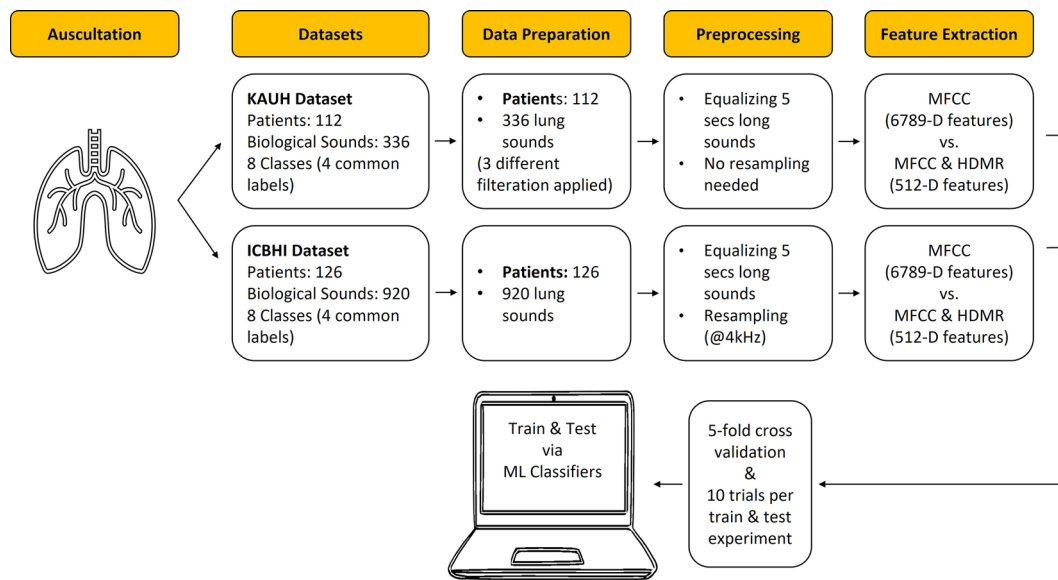
Fig. 2 Flowchart of the proposed method based on MFCC and MFCC+HDMR feature extraction techniques
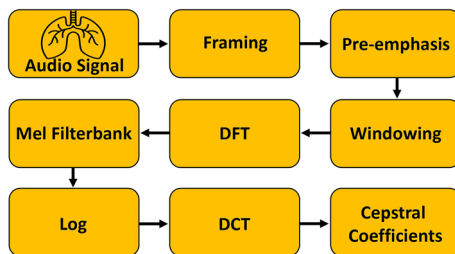

Fig. 3 Determination process of Mel frequency cepstral coefficients

Mel frequency scale. The mathematical relation between two units is defined as follows: [7, 11]

$$f_{\text{Mel}} = 2595 \log \left(1 + \frac{f}{700} \right) \tag{1}$$

The MFCC method contains the signal windowing and performs the Discrete Fourier Transform (DFT) initially. Then, the log function is applied to the transformed signals and the resulting frequencies are converted to the Mel scale. The MFCC process is finalized by adopting the Discrete Cosine Transform (DCT). Therefore, a 2-D mathematical entity can be considered as a matrix for each 1-D respiratory signal (Fig. 3).

In our experiments, we initially cropped each audio signal to 5 s and resampled it at 4 kHz. The MFCC frame size and corresponding frame shift were selected as 30 and 20 ms, respectively. Thus, 120 samples were generated. We also selected the Hamming window and employed 40 filters of the Mel filterbank.

3.3 High-dimensional model representation

High-dimensional model representation (HDMR) is a feature extraction method for multidimensional data. HDMR depends on a divide-and-conquer philosophy and enables to re-express a high-dimensional entity in terms of lower-dimensional components [33]. If the Mel Frequency Cepstral Coefficient matrix for a specific respiratory sound signal is denoted as $\mathbf{M}$ and assuming that $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$, then the HDMR expansion of $\mathbf{M}$ is given as follows:

$$\mathbf{M} = m_0 \, \mathbf{s_1} \mathbf{s_2}^T + \mathbf{m_1} \mathbf{s_2}^T + \mathbf{s_1} \mathbf{m_2}^T + \mathbf{m_{1,2}} \tag{2}$$

where m_0 is a scalar, $\mathbf{m_1}$, $\mathbf{m_2}$, $\mathbf{s_1}$ and $\mathbf{s_2}$ are vectors, while $\mathbf{m_{1,2}}$ stands for a matrix. In Eq. (2), m_0 is named the scalar or the constant HDMR term for the matrix $\mathbf{M}$. However, $\mathbf{m_1}$ and $\mathbf{m_2}$ are called 1-D HDMR terms; hence, they are vectors. On the right-hand side of Eq. (2), $\mathbf{s_1}$ and $\mathbf{s_2}$ are defined as

$$\mathbf{s_1} = \underbrace{[1 \ \cdots \ 1]^T}_{n_1} \quad \mathbf{s_2} = \underbrace{[1 \ \cdots \ 1]^T}_{n_2} \tag{3}$$

where they provide the algebraic dimensional convenience due to the outer product definition for the vectors.

The constant HDMR term m_0 is computed by using the following formula

$$m_0 = \frac{1}{n_1 n_2} \mathbf{s_1}^T \mathbf{M} \mathbf{s_2}. \tag{4}$$

In Eq. (4), it is obvious that m_0 is the average of the entries of the $\mathbf{M}$ matrix. Thus, it is convenient to say that m_0 stands for the roughest depiction of the data under consideration.

The first of 1-D HDMR terms for the matrix $\mathbf{M}$, that is $\mathbf{m_1}$, is determined as

$$\mathbf{m_1} = \frac{1}{n_2}\mathbf{M}\,\mathbf{s_2} - m_0\,\mathbf{s_1}. \tag{5}$$

According to Eq. (5), $\mathbf{m_1}$ of size n_1 includes the average of the columns of $\mathbf{M}$ and these mean values are shifted by m_0. That means $\mathbf{m_1}$ is a vector in the average direction of the columns belonging to the Mel frequency cepstral coefficient matrix. Somehow, the contribution of this entity to the matrix $\mathbf{M}$ is decorrelated from the other direction which is spanned by the rows of $\mathbf{M}$.

Similarly, the 1-D HDMR component which is responsible for the interpretation of the second dimension of $\mathbf{M}$ is determined as follows:

$$\mathbf{m_2} = \frac{1}{n_1}\mathbf{M}^T\mathbf{s_1} - m_0\,\mathbf{s_2}. \tag{6}$$

Here, $\mathbf{m_2}$ is a vector, has n_2 number of entries, and characterizes the second dimension of the Mel frequency cepstral coefficient matrix under consideration.

On the other hand, the 2-D HDMR term in Eq. (2), $\mathbf{m_{1,2}}$ is treated as the residual term and can be computed by subtracting the first three outer products from $\mathbf{M}$. The residual HDMR term is also considered noisy in various applications [33–35]. Therefore, we do not intend to exploit this term for further analysis.

3.4 Feature extraction and classification

As stated above, the aim of the paper is to classify pulmonary diseases by using acquired respiratory sounds. To this end, we employ several classification techniques from Machine Learning literature. In order to achieve high accuracy rates in machine learning classifiers, efficient features usage is necessary. The structure and the size of the features are also crucial since these entities directly affect the training and testing durations. To consider these properties, we propose a two-step feature extraction scheme. Initially, we generate MFCC matrix of size 498×14 for each respiratory signal. Since these matrices are 2-D and have a high amount of entry values, they are irrelevant to be exploited as features in Machine Learning classification methods. Therefore, we apply an efficient data reduction and feature extraction technique, HDMR enables to reduce the dimension of the data and extract effective features. Thus, we assess the 1-D HDMR components, $\mathbf{m_1}$ and $\mathbf{m_2}$ vectors, whose explicit formulas are provided in Eqs. (5) and (6), respectively, to generate efficient features in the following expression.

$$\mathbf{v} = \begin{bmatrix} \mathbf{m_1}^T & \mathbf{m_2}^T \end{bmatrix}^T \tag{7}$$

The $\mathbf{v}$ features are vectors of size 512 and convenient to be fed to any suitable classification algorithm.

To evaluate the performance of the proposed approach, the extracted features are exploited in several well-known classification algorithms, which are the Decision Trees (DT) [36], Support Vector Machines (SVM) [13–15], K-Nearest Neighbors (KNN) [10, 11, 27, 37], Ensemble Classifiers [16] and Kernel-based Classifiers [38]. The variations of the depicted classification methods are considered to enhance also the convincibility of the proposed feature extraction scheme.

## 4 Experimental results

4.1 The datasets

In this work, two datasets are utilized to conduct the relevant experiments. These datasets are ICBHI and KAUH [17–20]. The datasets include 238 subjects and 1256 audio recordings in total. Both datasets are public and free to access.

4.1.1 The ICBHI dataset

ICBHI is a public respiratory sound dataset released for a scientific challenge at the International Conference on Biomedical and Health Informatics (ICBHI) in 2017. This dataset has been very popular among the scientists, and research papers have been published in this regard. The dataset was acquired in Portugal and Greece, and it contains 5.5 h of recordings in 920 annotated audio files from 126 subjects. The dataset is composed of 8 classes in terms of diseases which are healthy, asthma, bronchiectasis, bronchiolitis, chronic obstructive pulmonary disease (COPD), lower respiratory tract infection (LRTI), upper respiratory tract infection (URTI), and pneumonia. The dataset is represented in Fig. 4. On the other hand, the dataset is separated into four classes in terms of lung sounds such as healthy, crackles, wheezes, and both crackles and wheezes. The sampling frequencies of the recorded wav files are 4 kHz (90), 44.1 kHz (824), and 10 kHz (6). The duration of the recordings varies between 10 and 90 s [19, 20]. ICBHI dataset is imbalanced containing 7 diseases, which lacks some adventitious samples including rhonchi.

4.1.2 The KAUH dataset

The KAUH dataset was recorded at King Abdullah University Hospital (KAUH), Jordan, and acquired from 112 subjects. The dataset was augmented by applying 3 different filters: Bell mode, Diaphragm mode, and extended mode filters. Thus, the KAUH dataset contains 336 pulmonary record signals. These signals are labeled per 8 different diseases: asthma, bronchiolitis, bronchitis, COPD, heart failure, pneu-
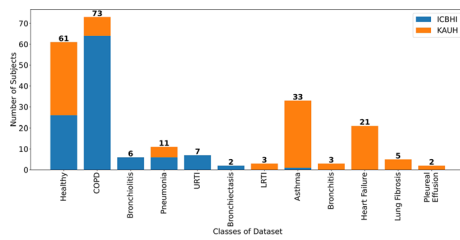
Fig. 4 Number of subjects belonging to healthy and disease classes

monia, lung fibrosis, and pleural effusion. The dataset is represented in Fig. 4. The sampling frequency of the audio signals is 4 kHz and the duration is 5 s, which are considered enough for a complete cycle of respiration [1, 18].

4.2 Classification performance

To demonstrate the efficiency of the proposed feature extraction methods based on MFCC and HDMR, several experiments are conducted. In these experiments, we employ both KAUH dataset and its combination with ICBHI data. Then, we perform feature extraction techniques on these datasets. These feature extraction schemes depend on MFCC and HDMR. The MFCC is a widely used tool in signal processing and can be considered as a standard feature extraction technique for digital signals. Therefore, MFCC were effectively used in other lung sound classification tasks in the related literature. However, HDMR is considered quite a new technique for the field. By acknowledging the decorrelation and the denoising capabilities of HDMR, we applied this new technique to the matrix, which contains the MFCC features. As a result of HDMR application, the 2-D feature structure is reduced to 1-D, which enables to decrease the size of the corresponding feature space and improves computational efficiency. To address the HDMR effect in the proposed approach, we organize several experiments for the abovementioned datasets. Accordingly, we exploit 14 distinct

classification algorithms to highlight the performance of both MFCC and MFCC+HDMR feature extraction approaches. These 14 classification methods are referred to as well-known classification algorithms in machine learning. The employed algorithms cover tree algorithms, variations of SVM methods, various KNN techniques, the subspace discriminant and logistic regression methods.

In each classification experiment, we use fivefold cross-validation to optimize the related hyperparameters and avoid overfitting the corresponding classification method. In kernel SVM, the Radial Basis Function (RBF) kernel, which is an effective tool for classifying nonseparable high-dimensional features, is adopted. In KNN-based algorithms, the number of closest neighbors is selected as 1 for the fine KNN, whereas this value is set as 10 for the Medium KNN, Cubic KNN, and Weighted KNN algorithms. In all experiments, the dataset was randomly split into 80% training data and 20% test data. To provide stable and accurate results, each experiment was repeated independently, and the average results are presented in Fig. 5.

It is obvious from this figure that HDMR improves consistently the feature extraction capability of MFCC for KAUH dataset. One can easily observe that the yellow bars are higher than the blue bars for all ML classification techniques. Moreover, application of HDMR along with MFCC as the feature extractor manages to boost the overall classification accuracy from 70.6 to 97.2% using fine KNN classifier. Besides, the MFCC+HDMR combination increases the accuracy from 53.6 to 94.1% and from 32.8 to 86.7% using weighted KNN and SVM Kernel methods. Similar results are encountered also for KAUH+ICBHI dataset. The HDMR utilization along with MFCC improves the accuracy from 77.9 to 85.1%, from 71.3 to 83.7%, from 78.2 to 82.6% and 66.4 to 79.4% by exploiting fine KNN, weighted KNN, Bagged trees and SVM kernel methods, respectively.

Additionally, comparison of our approach is also shown with existing techniques in literature, essentially the same

Fig. 5 Overall accuracy results for different ML classification algorithms employed for considered dataset and feature extraction scheme setups
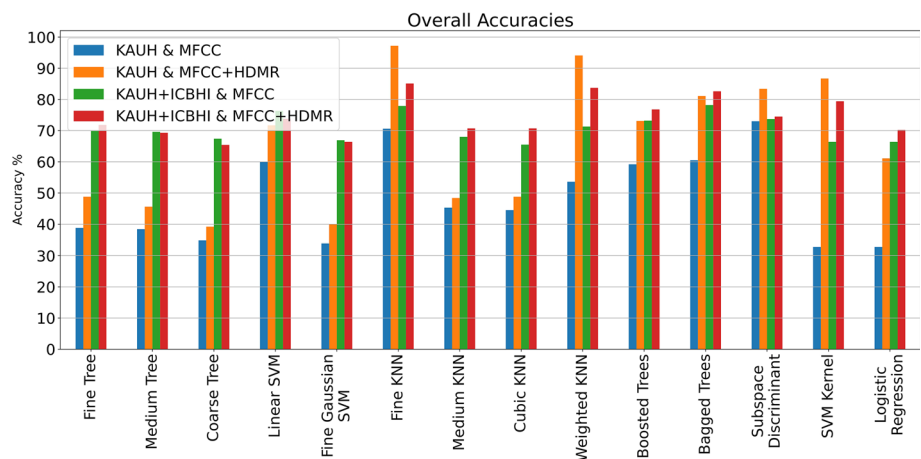
datasets are employed. To this end, we provide accuracy results from 7 up-to-date references along our method. The results are provided in Table 1, which covers the results from the works, which perform their corresponding approaches on ICBHI and KAUH+ICBHI datasets. The proposed approach suggests MFCC and HDMR feature combination extraction method and fed relevant machine learning classification algorithms with 1-D features. On the other hand, the other techniques in Table 1 exploit several deep learning concepts such as CNN, LSTM, BiLSTM and some pre-trained deep learning architectures, which enable optimization of the feature weights and perform classification concurrently. Deep learning techniques are talented on improving the overall accuracy in classification tasks, but they suffer from high computational burden during the training process. They also require a high amount of data samples to optimize the weight entities and specific hardware installments. Our method depends on a feature extraction scheme, which is capable of mimicking the lung sound characteristics, decorrelating, denoising and reducing the feature space size synchronously. Then, the features are organized as 1-D vectors and given to machine learning algorithms for the training process. According to Table 1, the proposed method achieves competitive accuracy performances in comparison with the state-of-the-art techniques.

5 Concluding remarks

The aim of this study is to develop an efficient feature extraction technique for classifying various pulmonary diseases accurately. To this end, the hybrid scheme is proposed for the combination of the MFCC and HDMR methodologies. MFCC is capable of imitating the human ear, which enables to characterize the acquired lung sound through stethoscopes. On the other hand, HDMR is qualified on decorrelation and denoising the given data while performing the data reduction. MFCC of a signal is two-dimensional in a matrix form. We apply HDMR to this matrix and generate efficient 1-D vector features. The extracted features are then delivered to the classification algorithms for high overall accuracy achievements.

The utilization of HDMR in lung sound classification is the most important contribution of the present work. According to the results obtained, HDMR empowers the classification accuracy established by MFCC only. It also enables data reduction, which is crucial for the subsequent machine learning tasks. We believe that these aspects of HDMR address its capacity and emphasize that it can be employed as an efficient feature extractor for biological sounds classification tasks.

Table 1 Overall accuracy comparison of the proposed method with other state-of-the art techniques employing ICBHI and KAUH datasets

References	Dataset	Subjects	Recordings	Features	Classifiers	Accuracy (%)
[1]	ICBHI+KAUH	213	1483	Spatial & Temporal	CNN + BiLSTM	99.62
[2]	ICBHI	126	6898	Spectrogram	CNN + LSTM with FL	76.39
[15]	ICBHI	126	6898	Spectrogram	CNN & SVM	65.50
[16]	ICBHI+KAUH	215	1484	Entropy features	Ensemble	98.27
[17]	ICBHI	126	10,775	MFCC & STFT	LDA & SVM & RUSBoost & CNN	99.60
[22]	ICBHI	126	6898	MFCC & SDC & SDC-L	SVM & KNN & RF & MLP & CNN & CNN + LSTM	95.00
[23]	ICBHI	126	6898	MFCC (Data Aug. via VAE)	ANN & CNN & LSTM ResNet-50 & Efficient Net B0	97.00
Proposed method	ICBHI+KAUH	238	1256	MFCC + HDMR	DT & KNN & SVM & Ensemble & Kernel	97.20

ANN artificial neural network, *BiLSTM* bidirectional long short-term memory, *CNN* convolutional neural network, *DT* decision tree, *FL* focal loss, *LDA* linear discriminant analysis, *RF* random forest, *RUS* random under-sampling, *SDC* shifted Delta Cepstral coefficient, *SDC-L* shifted Delta Cepstral coefficient in lower-subspace, *STFT* short-time Fourier transform, *VAE* variational autoencoders

Declarations

Conflict of interest The authors declare no competing interests.

Ethical Approval This declaration is not applicable.

References

1. Fraiwan, M., Fraiwan, L., Alkhodari, M., Hassanin, O.: Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. J. Ambient. Intell. Humaniz. Comput. **13**, 4759–4771 (2021). https://doi.org/10.1007/s12652-021-03184-y

2. Petmezas, G., et al.: Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. Sensors (2022). https://doi.org/10.3390/s22031232

3. Kim, Y., et al.: Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. Sci. Rep. **11**(1), 1–11 (2021). https://doi.org/10.1038/s41598-021-96724-7

4. Sarkar, M., Madabhavi, I., Niranjan, N., Dogra, M.: Auscultation of the respiratory system. Ann. Thorac. Med. **10**(3), 158 (2015). https://doi.org/10.4103/1817-1737.160831

5. Palaniappan, R., Sundaraj, K., Sundaraj, S.: Artificial intelligence techniques used in respiratory sound analysis—a systematic review. Biomedizinische Technik/Biomed. Eng. **59**(1), 7–18 (2014). https://doi.org/10.1515/bmt-2013-0074

6. Pramono, R.X.A., Bowyer, S., Rodriguez-Villegas, E.: Automatic adventitious respiratory sound analysis: a systematic review. PLoS ONE **12**(5), e0177926 (2017). https://doi.org/10.1371/journal.pone.0177926

7. Jung, S.Y., Liao, C.H., Wu, Y.S., Yuan, S.M., Sun, C.T.: Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features. Diagnostics **11**(4), 732 (2021). https://doi.org/10.3390/diagnostics11040732

8. Kahya, Y.P., Guler, E.C., Sahin, S.: Respiratory disease diagnosis using lung sounds **5**, 2051–2053 (1997). https://doi.org/10.1109/IEMBS.1997.758751

9. Bahoura, M.: Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Comput. Biol. Med. **39**(9), 824–843 (2009). https://doi.org/10.1016/j.compbiomed.2009.06.011

10. Içer, S., Gengeç, Ş: Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. Digit. Signal Process. **28**, 18–27 (2014). https://doi.org/10.1016/j.dsp.2014.02.001

11. Chen, C.H., Huang, W.T., Tan, T.H., Chang, C.C., Chang, Y.J.: Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors **15**(6), 13132–13158 (2015). https://doi.org/10.3390/s150613132

12. Jakovljević, N., Lončar-Turukalo, T.: Hidden Markov model based respiratory sound classification. **66**, 39–43 (2018). https://doi.org/10.1007/978-981-10-7419-6_7

13. Aykanat, M., Kilic, O., Kurt, B., Saryal, S.: Classification of lung sounds using convolutional neural networks. EURASIP J. Image Video Process. **2017**(1), 65 (2017). https://doi.org/10.1186/s13640-017-0213-2

14. Bardou, D., Zhang, K., Ahmad, S.M.: Lung sounds classification using convolutional neural networks. Artif. Intell. Med. **88**, 58–69 (2018). https://doi.org/10.1016/j.artmed.2018.04.008

15. Demir, F., Sengur, A., Bajaj, V.: Convolutional neural networks based efficient approach for classification of lung diseases. Health Inf. Sci. Syst. **8**(1), 1–8 (2019). https://doi.org/10.1007/s13755-019-0091-3

16. Fraiwan, L., et al.: Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. Biocybern. Biomed. Eng. **41**(1), 1–14 (2021). https://doi.org/10.1016/j.bbe.2020.11.003

17. Rocha, B.M., Pessoa, D., Marques, A., Carvalho, P., Paiva, R.P.: Automatic classification of adventitious respiratory sounds: a (un)solved problem? Sensors **21**(1), 57 (2021). https://doi.org/10.3390/s21010057

18. Fraiwan, M., Fraiwan, L., Khassawneh, B., Ibnian, A.: A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. Data Brief **35**, 106913 (2021). https://doi.org/10.1016/j.dib.2021.106913

19. Rocha, B.M., et al.: A respiratory sound database for the development of automated classification 33–37 (2018). https://doi.org/10.1007/978-981-10-7419-6_6

20. Rocha, B.M., et al.: An open access database for the evaluation of respiratory sound classification algorithms. Physiol. Meas. **40**(3), 035001 (2019). https://doi.org/10.1088/1361-6579/ab03ea

21. Engin, M.A., Aras, S., Gangal, A.: Extraction of low-dimensional features for single-channel common lung sound classification. Med. Biol. Eng. Comput. **60**(6), 1555–1568 (2022). https://doi.org/10.1007/s11517-022-02552-w

22. Kwon, A.M., Kang, K.: A temporal dependency feature in lower dimension for lung sound signal classification. Sci. Rep. **12**(1), 1–11 (2022). https://doi.org/10.1038/s41598-022-11726-3

23. Saldanha, J., et al.: Data augmentation using variational autoencoders for improvement of respiratory disease classification. PLoS ONE **17**(8), e0266467 (2022). https://doi.org/10.1371/journal.pone.0266467

24. Cohen, A., Landsberg, D.: Analysis and automatic classification of breath sounds. IEEE Trans. Biomed. Eng. **9**, 585–590 (1984)

25. Sankur, B., Kahya, Y.P., Güler, E.Ç., Engin, T.: Comparison of AR-based algorithms for respiratory sounds classification. Comput. Biol. Med. **24**(1), 67–76 (1994). https://doi.org/10.1016/0010-4825(94)90038-8

26. Kandaswamy, A., Kumar, C.S., Ramanathan, R.P., Jayaraman, S., Malmurugan, N.: Neural classification of lung sounds using wavelet coefficients. Comput. Biol. Med. **34**(6), 523–537 (2004). https://doi.org/10.1016/S0010-4825(03)00092-1

27. Serbes, G., Sakar, C.O., Kahya, Y.P., Aydin, N.: Pulmonary crackle detection using time-frequency and time-scale analysis. Digit. Signal Process. **23**(3), 1012–1021 (2013). https://doi.org/10.1016/j.dsp.2012.12.009

28. Messner, E., et al.: Multi-channel lung sound classification with convolutional recurrent neural networks. Comput. Biol. Med. **122**, 103831 (2020). https://doi.org/10.1016/j.compbiomed.2020.103831

29. Turner, C., Joseph, A.: A wavelet packet and MEL-frequency cepstral coefficients-based feature extraction method for speaker identification. Procedia Comput. Sci. **61**, 416–421 (2015). https://doi.org/10.1016/j.procs.2015.09.177

30. Rabiner, L.R., Schafer, R.W.: Theory and Applications of Digital Speech Processing. Pearson (2010)

31. Li, F., Cox, T.: Digital Signal Processing in Audio and Acoustical Engineering. CRC Press (2019)

32. Lopez-Ruiz, R.: From Natural to Artificial Intelligence. Inte-chOpen (2018). https://doi.org/10.5772/intechopen.71252

33. Tuna, S., Özay, E.K., Tunga, B., Gürvit, E., Tunga, M.A.: An efficient feature extraction approach for hyperspectral images using wavelet high dimensional model representation. Int. J. Remote Sens. **43**(19–24), 6899–6920 (2022). https://doi.org/10.1080/01431161.2022.2147036

34. Tuna, S., et al.: Iterative enhanced multivariance products representation for effective compression of hyperspectral images. IEEE Trans. Geosci. Remote Sens. **59**(11), 9569–9584 (2021). https://doi.org/10.1109/TGRS.2020.3031016

35. Manzhos, S., Ihara, M.: Optimization of hyperparameters of Gaussian process regression with the help of a low-order high-dimensional model representation: application to a potential energy surface. J. Math. Chem. **61**, 7–20 (2023). https://doi.org/10.1007/s10910-022-01407-x

36. Kingsford, C., Salzberg, S.L.: What are decision trees? Nat. Biotechnol. **26**(9), 1011–1013 (2008). https://doi.org/10.1038/nbt0908-1011

37. Uddin, S., Haque, I., Lu, H., Moni, M.A., Gide, E.: Comparative performance analysis of k-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci. Rep. **12**(6256), 1–11 (2022). https://doi.org/10.1038/s41598-022-10358-x

38. Camps-Valls, G., Bruzzone, L.: Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **43**(6), 1351–1362 (2005). https://doi.org/10.1109/TGRS.2005.846154