Improvise approach for respiratory pathologies classification with multilayer convolutional neural networks

Saumya Borwankar[1] · Jai Prakash Verma[1] · Rachna Jain[2] · Anand Nayyar[3]

Abstract

Every respiratory-related checkup includes audio samples collected from the individual, collected through different tools (sonograph, stethoscope). This audio is analyzed to identify pathology, which requires time and effort. The research work proposed in this paper aims at easing the task with deep learning by the diagnosis of lung-related pathologies using Convolutional Neural Network (CNN) with the help of transformed features from the audio samples. International Conference on Biomedical and Health Informatics (ICBHI) corpus dataset was used for lung sound. Here a novel approach is proposed to pre-process the data and pass it through a newly proposed CNN architecture. The combination of pre-processing steps MFCC, Melspectrogram, and Chroma CENS with CNN improvise the performance of the proposed system, which helps to make an accurate diagnosis of lung sounds. The comparative analysis shows how the proposed approach performs better with previous state-of-the-art research approaches. It also shows that there is no need for a wheeze or a crackle to be present in the lung sound to carry out the classification of respiratory pathologies.

Keywords CNN (Convolutional neural network) · MFCC (Mel-frequency cepstral coefficients) · Melspectrogram · CENS (Chroma energy normalized statistics) · Respiratory pathologies classification

✉ Anand Nayyar
anandnayyar@duytan.edu.vn

Saumya Borwankar
17bec095@nirmauni.ac.in

Jai Prakash Verma
jaiprakash.verma@nirmauni.ac.in

Rachna Jain
rachnajain@bpitindia.com

[1] Institute of Technology, Nirma University, Ahmedabad, Gujarat, India

[2] IT department, Bhagwan Parshuram Institute of Technology, New Delhi, India

[3] Graduate School, Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam

1 Introduction

The diagnosis is more related to audio or visual information gathered from different methods. Medicine specialist diagnoses the data. Becoming a medicine specialist takes skill, experience, and time. The World Health Organization (WHO) gave a statistic [2] based on the available physician per 1000 population. Around 45% of WHO member states have less than an average number of physicians that is 1 per 1000. With these statistics, the diagnosis at a non-compressible time with the physicians being overbooked can lead to a misdiagnosis. So there is a need to find ways to help medicine specialists save their time and make an accurate diagnosis. Automatic diagnosis of the disease using computers can help achieve this goal.

Every respiratory-related diagnosis is based on an audio sample taken from the patient by a specialist from various parts of the body using a specific tool (sonography, stethoscope). It helps with the diagnosis of the disease and hearing sounds (normal, wheeze, crackle, etc.) [43][63]. It shows us that the audio sample helps in the classification of respiratory diseases. Heart-related disorders are one of the leading causes of death around the globe [1]. There have been many studies published about audio analysis of heart-related problems. A review paper [17] on signal processing techniques concludes that the gathering, analyzing, and classification of heart sounds for better diagnosis is required. As Machine Learning and Deep Learning techniques deploy to help in the medical sector, the diagnosis has become more accurate and saves a lot of time [16][10]. This recent approach seems viable to help with the detection and improve the performance of systems aimed at diagnosing the disease at an early stage. Also, this could help doctors and nursing staff prepare reliable and faster diagnosis reports [68]

Breath sounds may be reduced and expiration may be longer in COPD patients [70]. Patients with COPD, particularly those with chronic bronchitis, may hear coarse crackles at the start of inspiration [20]. These crackles have a "popping" sound to them, vary in quantity and timing, and can be heard in any part of the lungs [35]. These early inspiratory crackles can also be heard during expiration, and coughing can make these noises go away [35]. The passage of boluses of gas through an intermittently obstructed airway causes these harsh crackles [20]. During inspiration, normal breath sounds are characterised by a low noise. These sounds are hardly discernible during expiration [64]. There are no distinct peaks and the sound is not melodic [56]. The lobar and segmental airways create the inspiratory component of the sound, whereas the more proximal regions produce the expiratory component. Normal lung sounds are considered to be caused by air turbulence [56].

The International Conference on Biomedical Health Informatics (ICBHI) had organized a challenge aimed at developing an automatic identification of wheezes and crackles in respiratory audio samples. For this task, a corpus was built [66] which consists of audio samples of seven respiratory pathologies namely asthma, chronic obstructive pulmonary disease (COPD), lower respiratory tract infection (LRTI), pneumonia, bronchiolitis, upper respiratory tract infection (URTI) and bronchiectasis along with audio samples from healthy people. The detailed description of the selected dataset discussed in the dataset section. The leading cause of pathologies in the corpus is smoking. Other then the smoking some are genetic while others are by environmental factors [22].

Chronic Obstructive Pulmonary Disease (COPD) is very similar to asthma as it causes the same effects like shortness of breath and cough. So the detection of COPD is difficult [72]. It is mainly by smoking. All these symptoms can also misinterpret as old age symptoms.

Upper Respiratory Tract Infection (URTI) disease is common around the time of fall and winter and can happen at any given time. It is mainly by a virus [38]. The symptoms are often confused with pneumonia [22]. Pneumonia infected people mostly recover in a little time, sometimes it is difficult for certain people, and it can cause death in some cases. So the detection of the disease is necessary.

Bronchiectasis is considered as chronic illness. In this condition, the airway of the lungs can widen and damages the air passage and allows mucus and bacteria to build up in the lungs. It causes the blockage of airways. This condition can often be confused with bronchiolitis or common cold [77]. But there is a little difference between the two, bronchiolitis is mostly affected by young children and can cure. On the other hand, bronchiectasis is a chronic disease.

Lower Respiratory tract Infections (LRTI) disease has symptoms that depend on the advancement of the disease. At the early stages, symptoms are similar to bronchiolitis and bronchiectasis. The later stages can cause a pneumonia-like situation.

In recent times the advancements in technologies and the availability of data for the research of such conditions. The Deep Learning algorithm, Convolutional Neural Network(CNN) has helped in the field of medical and biological data for the segmentation and classification [39][71], brain tumor classification [58], prostate [40] to name a few. Here in this paper, we propose a new way of preprocessing the audio files with CNN architecture to classify the pathologies. The contribution of this work is to suggest a new approach to classify audio samples. A Deep Learning approach using a convolution neural network has used in [59] that was able to achieve an accuracy of 83%. A new architecture of the convolutional neural network proposed along with an approach at preprocessing data to compare the results with the previous. We have used a matrix of MFCC features to fed to the 1D CNN as the 1D CNN will be able to capture all the important features. As it has a kernel size of 1 so no important feature will be left behind. The main target of this work aims towards:

1.1 Contribution

Based on all the discussion mentioned above, we come up with a solution that can overcome the limitation of the previous research work and give a better and more accurate result.

1. Novel approach incorporating feature transformation by combining pre-preprocessing steps MFCC, Melspectogram and Chroma CENS with CNN.
2. Incorporation of future deep learning feature extraction methods like d-vectors, i-vectors for better audio signal representations.
3. Analysis on ICBHI dataset (chronic, non-chronic and normal classes) with the comparison of three features (MFCC, Melspectrogram and ChromaCENS).

The proposed research has numerous practical applications in field of medical science, along with operating at low cost solution, if deployed at a hospital level. It has the potential to work in a semi-supervised way which can make healthcare more accessible as well as cost effective. This work helps incorporate maximum information from the audio as it is dependent on three features as compared to one. Deploying this service can help common public make use of the technology for free, they would just require to get their respiratory sound from an instrument and use the deployed service to get a diagnosis.

1.2 Paper organization

The paper divides into seven sections in which Section 2 shows detailed information about previous approaches and related work in the field of respiratory sound classification. Section 3 shows a description of the dataset used for the evaluation of the proposed approach. Section 4 presents information on the implementation scenario and description of the machine on which the algorithm carried out. Section 5 describes the execution and the implementation scenario. Section 6 discusses the results and experiments and has a comparison table that shows us the performance of different approaches. Section 7 presents the conclusion and future work for the proposed research work.

2 Related work

There are various studies that deal with this challenge from the perspective of audio processing. Pasterkamp et al. [57] played a major role in reporting the types and characteristics of respiratory sound. The presence of wheezes and crackles has helped many research work to classify different categories of diseases. It is also not necessary to consider the difference in gender or age in this task of respiratory analysis [24][18]. There respiratory sound samples are subject to noise such as artefacts and heart sounds [53].

Many previous studies on respiratory pathology analysis have been conducted using machine learning and various signal processing algorithms [52]. A review [62] was published in 2007 highlighted the researchers study to identify markers such as wheezes and crackles. With these markers being identified this problem was transformed into a classification problem [12] and machine learning can help with this task. Islam et al. [30] were able to detect asthma by the use of lung sound samples with the help of the wheeze. They used the samples collected from 60 individuals with half of the asthma. The authors collected the data from the back of the individuals. The classification was done with the help of Artificial Neural Network (ANN) and also with Support Vector Machine (SVM). Best accuracy found with SVM that was around 93%. Other research articles based on the diagnosis of wheezes and crackle [25] used configurations of neural networks, and they were able to obtain an accuracy of 93% for crackles and 91.7% for wheezes. Bardou et al. [7] used the dataset with seven classes stridor, squeak, polyphonic wheeze, monophonic wheeze, fine crackle, coarse crackle, and the normal. The classification reached the best results with the help of CNN's.

Yunseo et al. [36] dealt with something similar. The number of patients with respiratory diseases in Seoul on a daily basis was gathered. The daily number of patients treated for respiratory disorders per 10,000 residents was predicted using meteorological and air-pollution parameters. To determine the relevance of feature selection, we used a relief-based feature selection method. Two alternative prediction models were developed using the gradient boosting and Gaussian process regression (GPR) approaches.

Mahmoud et al. [4] shared their findings via averaging the predictions of many models trained and assessed individually on distinct sound kinds, simple binary classifiers may reach an AUC of 96.4 percent and an accuracy of 96 percent. Finally, the goal of this research is to highlight the relevance of the human voice, as well as other respiratory noises, in sound-based COVID-19 diagnosis.

Eu Sun Lee et al. [37] used a machine learning approach as an AI tool to examine the associations between weather and air pollution factors and respiratory illness patients

attending EDs, based on the consistently reported and systemized data registry of national emergency medical facilities. This study used data from three days before to a visit, as well as the daily temperature difference and other information, to estimate the exact values of meteorological conditions and their amount of effect.

Chen et al. [13] proposed ResNet architecture for the classification of an Optimized S-Transform based (OST) feature map of wheezes, crackles, and normal sounds. The RGB maps of the feature map fed into the ResNet architecture for classification. The results compared with ResNet Short Term Fourier Transform (STFT) and ResNet S-transform (ResNet-ST) with better results carried out by the ResNet-OST. Jakovljevic et al. in [31] had come up with an algorithm to classify audio samples into wheezes, both wheezes with crackles and crackles with normal from ICBHI corpus [66]. The methodology they followed consisted of noise suppression using spectral subtraction after which features were extracted, then the Hidden Markov Models are used to classify and obtain 39.56% average of specificity and sensitivity. Perna el at. [59] used CNN to classify the different pathologies present in the ICBHI corpus [66]. The MFCC features extracted from the audio that later fed into the 2D CNN architecture for classification. They had compared the effects of different activation functions, and the two techniques SMOTE sampling and RUS technique. They were able to get an accuracy of 83%, F1 score of 0.84 and recall of 0.82. Prerna et al. [60] introduced the use of a Long Short Term Memory (LSTM) based approach for the classification of respiratory pathologies. The authors had used MFCC features to classify the audio sample and reached an accuracy of about 99%. The normal and abnormal classes were considered. Further, the LSTM was used to classify three i.e. chronic, non-chronic, and normal and four classes i.e. normal, wheeze, crackle, and both where they were able to get an accuracy of about 98% and 74% respectively, and F1 score of 0.91 and recall of 0.82 when three classes were taken. Garcia et al. [21] proposed a CNN architecture considering the mel-spectrogram image as a feature. The melspectrograms were classified into three classes i.e. chronic, non-chronic and normal with the help of a CNN model. They were able to reach an accuracy of about 99% on the three classes, F1 score of 0.900 and recall of 0.986.

Guler et al. [26] studied power spectral density features of the respiratory sounds and they were classified into crackles, wheeze and normal respiratory sounds. Electret microphones had been used to record 129 subject for their respiratory sound. For the classification genetic algorithm (GA)-based Artificial neural networks and artificial neural network (ANN) were used. Accuracy of 81-91% and 83-93% were calculated for ANN and GA-based ANN respectively.

Alsmadi et al. [3] proposed an autoregressive model for the problem of classification of respiratory sound. ECM-77B microphone was used to record 43 subjects for their respiratory sounds and for the classification they used k-nearest neighbour algorithm (k-nn) which led to a recognition rate of 96%. Dockur et al. [15] proposed an incremental supervised neural network for the task of classification of respiratory sound. For the evaluation they had 18 subjects and the power spectrum features were used. After which a grow and learn (GAL) network was used for the incremental supervise network and their approach was compared to previous approaches. Sankar et al. [67] studied about a feedforward neural network approach for this task. The features that they used were respiratory rate, energy index, strength of dominant frequency and dominant frequency. Electret microphone was used on size subjects to record their respiratory sounds which led to classification accuracy of 98.7%.

Hashemi et al. [29] highlighted the use of wavelet based features along with multilayer perceptron network for the task of this classification. Electric stethoscope were used to record 140 subjects for their respiratory sounds and the approach led to a recognition rate of 89.28%. Flietstra et al. [19] proposed support vector machine (SVM) for the this problem. STG 16 lung sound analyser was used to record 257 subjects for their respiratory sounds. They were able to get mean classification accuracy of 84% with this approach. Palaniappan et al. [54] presented a comparative study about the used of SVM and k-nn for the task of classification of respiratory sounds. They evaluated their approach on R.A.L.E. [52] dataset with the use of MFCC features along with one way ANOVA test [44] and were able to reach classification accuracy of about 92.1% and 98.26% for SVM and k-nn respectively.

Gadge et al. [50] studied the analysis of respiratory sounds and developed a MATLAB based tool. The tool helped filter heart sounds from acoustic pulmonary sounds. Mhetre et al. [47] developed a tool for the plotting and viewing of spectrogram of respiratory sounds. Lin et al. [41] proposed a neural network architecture to classify wheeze and normal breath sounds using average truncate method. ECM microphone was used to record 58 subjects for their respiratory sounds. They reported a specificity of 1 and sensitivity of 0.946. Umeki et al. [73] proposed a hidden markov model (HMM) for this task of classification of normal and abnormal respiratory sounds with the help of respiratory rate from breath sounds. They were able to reach classification accuracy of about 83.7%. Maruf et al. [45] presented SVM, gaussian mixture model (GMM) and ANN for the task of classification of normal and crackle respiratory sound with the help of spatial temporal features. They reported classification accuracy of 92.6%, 85.3%, and 97.56% for SVM, ANN and GMM respectively. Lin et al. [42] developed a recognition system capable of detecting wheeze with the help of backward-propogating neural network based on spectrogram features. The proposed approach was able to reach a specificity of 1 and sensitivity of 0.946.

Yadav et al. [75] described a machine learning-based method for. Use the classification system to determine whether a pulmonary signal is natural or abnormal, pulmonary acoustic sounds were documented. They extracted wavelet coefficients which were fed into a supervised machine learning classifier called SVM, which efficiently classifies pulmonary sounds as natural or abnormal. When it comes to classifying pulmonary tones, the experimental findings indicate an accuracy of 92.30 percent.

Goudarzi et al. [23] proposed a novel approach with the help of recurrent fuzzy function. The dataset used had 31 asthma, 27 COPD and 25 normal patients. using a 10-fold cross validation an accuracy of less than 80% was reported. Naves et al. [49] used higher order statistical features which were extracted from respiratory sounds. Although the number of samples used for this research work were limited to come to a conclusion. They used two classifiers Naive Bayes and k-nn, along with this Fisher's discriminant analysis for dimension reduction and they were able to reach an accuracy of about 94.6%. Palaniappan et al. [51] presented a reliable classifier with the help of SVM based on parametric feature extraction technique. They were able to reach an accuracy of 89.68% and 88.72% for MFCC and AR coefficients respectively. Chambers et al. [8] proposed a way to look at a trend for a patient rather than focusing only on audio content for every respiratory cycle. Their macro level approach gave an accuracy of about 85%. Chinazunwa et al. [74] proposed a mobile application to classify respiratory sounds using machine learning methods. They were able to reach an accuracy of about 88.9%, 75.8% and 86.7% with k-nn, SVM and random forest.

Murat et al. [6] proposed a convolution neural network for this task and compared his approach to another SVM approach. The classification was done for several classes such as rhonchus, rale, normal, singular respiratory sounds. The maximum accuracy presented was 86% and 86% for SVM and CNN respectively for healthy vs pathological classification. Chen et al. [11] proposed a classification of rale, rhoncus, wheeze and normal sounds with the use of MFCC and k-nn as a classifier. They were able to reach an accuracy of about 93.2% with data from 140 subjects. Aras et al. [5] used data from 27 subjects and were able to get an accuracy of about 96% with the classes being rale, rhoncus and normal sounds with the help of k-nn and MFCC, LFCC being the features. Serbes et al. [69] used data from 26 patients and classified crackles with an accuracy of 97.2% with the help of SVm as a classifier and wavelet transform as a feature extraction method. Yamashita et al. [76] proposed a HMM based approach to classify normal and emphysema sounds. For their work they used data from 114 patients. They were able to reach a maximum accuracy of 88.7% with segmentation of audio samples. Feng et al. [32] used data from 21 subjects to classify normal and abnormal sounds with the help of temporal-spectral dominance spectrogram as a feature extraction method and k-nn as a classifier. They were able to reach an accuracy score of 92.4%.

Charleston et al. [9] proposed the use of AR model features with multilayer perceptron as a classifier with data from 27 subject. They were able to reach an accuracy score of 75% and 93% for the classes of normal and abnormal sounds. Yamamoto et al. [46] used raw data from 114 subjects and with the help of this data he was able to make a predictor with 84.2% accuracy for normal and abnormal sounds and HMM as a classifier. Kahya et al. [33] proposed the use of k-nn with wavelet transform features extracted from sounds of 20 different subjects. They were able to reach an accuracy score of 46%. Riella et al. [65] proposed the use of FFT and STFT features along with multilayer perceptron to reach an accuracy of 92.8%. They were able to classify just wheeze sounds. Kandaswamy et al. [34] proposed the use of multilayer perceptron with wavelet transform and STFT as features from sounds. They were able to reach an accuracy of 94.2% for lung sounds. From our literature survey we were able to figure out that the classification was being varied across different features and classifiers, to eradicate this our proposed approach aims at including three features together for the classification along with convolutional neural network. Nishi et al. [27] proposed a SVM classifier using MFCC features for the task of detecting COPD from lung sounds. They were able to reach an accuracy of about 98.2%. A comparative analysis shows in Table 1 for better understanding of approaches and the classification done by different authors and their methodologies along with the classification classes. Along with the advantages and limitations of the previous approaches.

3 Dataset details

The International Conference on Biomedical and Health Informatics (ICBHI) corpus [66] consists of a total of 5.5 hours of recording which contain 6898 respiratory cycles of 126 patients. The categories are Chronic Obstructive Pulmonary Disease (COPD), Upper Respiratory Tract Infection (URTI), Healthy, Pneumonia, Bronchiectasis, and Bronchiolitis. The number of audio files present in the corpus is given in Table 1.

The respiratory cycles are annotated by experts in their domain, to provide the presence of respiratory pathologies. The annotation format is done in a style that includes the last part of the respiratory cycle as well as the beginning, presence of crackles and wheezes. The

Table 1 A comparative analysis for better understanding of approaches and the classification done by researchers

Ref	Objective	Methodology	1	2	3	4	5	6	7	8	Pros	Cons
Islam et al. [30]	To classify lung sound based on wheeze	MFCC - ANN, SVM	✓							✓	Whole lung sound cycle used for the classification	Multi-class classification not looked at
Jakovljevi et al. [31]	Classify audio into normal, crackle, wheeze, and both crackle and wheeze	MFCC - HMM	✓	✓	✓	✓	✓	✓	✓	✓	Use of ensemble of classifiers.	Author mentions the use of advance noise suppressing techniques which were not looked at.
Bardou et al. [7]	Classify audio into lung sounds of wheeze, normal, monophonic, coarse, fine, polyphonic,crackle, squawk, and stridor sounds.	MFCC - SVM, KNN, GMM Spectrogram - CNN	✓								LBP feature classification and introduction of CNN	Larger CNN topologies need to be made
Chen et al. [13]	Classify audio in wheeze crackles and normal	Spectrogram - CNN		✓	✓	✓	✓	✓	✓	✓	Different spectral mapping techniques used	Multi-class classification and real-time detection not looked at.
Guler et al. [25]	The different neural structures of PSD values is classified into crackle, wheeze, and normal.	Welch method		✓							Hybridization of genetic algorithm with neural networks, the Welch with PSD estimation calculations methods used.	Work needs to be done in proper model selection
Prema et al. [60]	Classify audio into chronic, non-chronic, wheezes, crackle and normal	MFCC-RNN		✓	✓	✓	✓	✓	✓	✓	The use of LSTM to enhance the performance	Author mentions the use of multiple deep learning architectures for future work
Garcia et al. [21]	Classify audio into chronic, non-chronic, wheezes, crackle and normal	Melspectrogram-CNN		✓	✓	✓	✓	✓	✓	✓	The use of melspectrogram and CNN model for classification	Use of variable length audio and the detection of parts of melspectrogram emphasizing detection
Feng et al. [32]	Classify sounds into normal and abnormal	Spectrogram-knn		✓							Use of temporal-spectral dominance along with k-nn improved performance.	The representation of time-frequency has to be looked at from high complexity point of view.

Table 1 (continued)

Ref	Objective	Methodology	1	2	3	4	5	6	7	8	Pros	Cons
Serbes et al. [69]	Classify crackles from audio	Wavelet transform - SVM	✓								Preprocessing helps remove no-information containing regions with DTCWT.	Effect of windowing for time-frequency and mother wavelet function has not been looked at.
Guler et al. [26]	Classify audio into normal, wheeze and crackles	Welch -Genetic algorithm ANN	✓								Use of welch method helped in preprocessing	More networks need to be looked at for better performance.
Nishi et al. [27]	Classify COPD from audio	MFCC-SVM	✓			✓					The linear predictive coefficient and median frequency are found to be best biomarkers for COPD detection	Considering multi-centered data along with COPD subjects.
Chen et al. [11]	Classify rale, rhoncus, wheeze and normal sounds	MFCC-knn	✓								Approach provides sound frame identification results to grade the lung sounds	Different variations of ML algorithms were not looked at.
Hashemi et al. [29]	Classify wheeze from respiratory sounds.	Wavelet transform- MLP	✓								They were able to demonstrate that higher wavelet do not produce better results.	Computation cost was increased.
Yamashita et al. [76]	Classify normal and emphysema	Segmentation-HMM	✓								They were able to get robust results between two likelihoods.	Multi class classification was not looked at.
Prema et al. [59]	Classify audio into chronic, non-chronic diseases and normal	MFCC - CNN	✓	✓	✓	✓	✓	✓	✓	✓	The use of convolutional variational autoencoders for augmenting the classes with fewer samples.	Classification based on variable length.

1:normal, 2:Bronchiectasis, 3:Bronchiolitis,4:COPD. 5:Pneumonia, 6:LRTI, 7:URTI, 8:Asthma.

audio recorded were gathered using various equipment with a ranging duration of 1s to 90 s, and the average duration being 2.7 s with a standard deviation of 1.17 s and the median being about 2.54 s. The data is obtained from 7 different locations on the patient's chest like right and left anterior, right and left lateral, trachea, and right and left posterior.

For comparison purposes, the dataset has been divided into 3 parts Chronic, Non-Chronic, and Healthy. The Chronic part contains COPD, Bronchiectasis, and Asthma, while Non-Chronic contains Bronchiolitis, URTI, LRTI, and Pneumonia. This will help us in the proper evaluation of our proposed method with respect to the previous approach in [59].

4 Proposed research work

For our proposed approach highlighted in (Fig. 1) we have first augmented the data so for the augmentation of the data, we have used a delay function that adds the delayed audio with the original audio. The delay function adds a delay of 250ms, 500ms, 750ms and 1000ms, so we are able to produce 4 different versions of one single audio sample. Next, we have used 3 types of features namely Mel-frequency cepstral coefficients (MFCCs), Melspectrogram features and Chroma energy normalized statistics (CENS) features. The number of features used is 39. The 39 MFCC features cover double delta features as well. a brief description of the 3 features is given in the following subsections. After the transformation of features they are stacked together so they behave as a single multidimensional feature which is used for the classification. Finally a newly proposed CNN architecture is used for the classification of 3 classes namely chronic, non-chronic and healthy audio. Detailed explanation is given in Section 4.

4.1 Mel-frequency cepstral coefficients (MFCCs)

Classification of pathologies is done by transforming different features from audio samples and one of them is Mel frequency cepstral coefficients which have produced good results in audio classification tasks [55] [28]. In audio processing, the short term spectrum of sound representation is called a Mel frequency cepstrum, which is based on the Mel scale of frequency. Various studies have shown the effectiveness of MFCCs in audio processing.
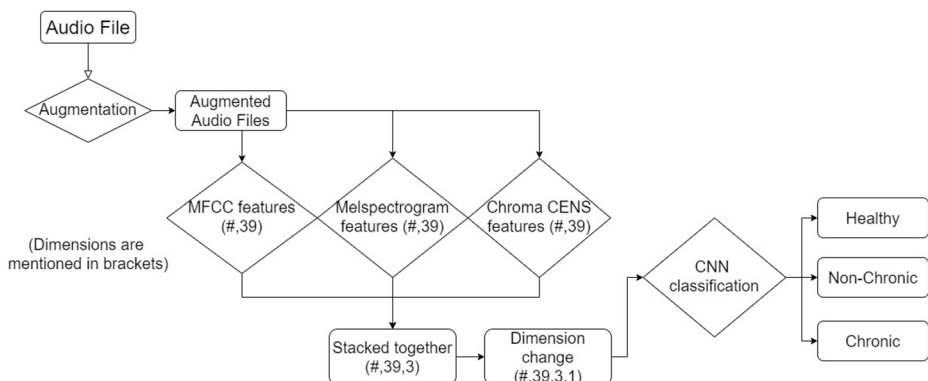

Fig. 1 Proposed execution flow with flowchart

MFCC features of audio are based on hearing perceptions of humans who can not perceive frequencies over 1 kHz. The MFCC features are based on the critical bandwidth variations of the ear [61]. MFCC consists of 2 filters that are logarithmically spaced above 10kHz and linearly spaced below 1kHz. MFCC has 7 computational steps [48].

1. Pre-emphasis: In this step, the audio sample is passed through a filter which increases the emphasis of higher frequencies.
2. Framing: This step consists of segmenting the audio sample from analog to digital, in frames with a duration of around 20 to 40 ms.
3. Windowing: Window named Hamming window is used for the next block in the process.
4. Fast Fourier Transform: The frames are converted into the frequency domain using the Fast Fourier transform (FFT).
5. Mel Filter Bank Processing: The filter bank according to mel scale is performed as shown in Figure 2. To calculate the weighted sum of filter spectral components the triangular filters are used. Also to make the output with respect to the mel scale.
6. Discrete Cosine Transform: Now the conversion of log mel spectrum in time domain takes place with the help of Discrete Cosine Transform. The results of the conversion are called the coefficients.
7. Delta Energy and Delta Spectrum: The cepstral features change over time so features are added, features like double delta or delta or velocity features.

Dealing with variable audio length, every 39 MFCC feature over all the frames are averaged and a final vector of 39 features is formed for a single audio file. The MFCC features are a bit more decorrelarated than melspectrogram features, due to which MFCC features have been proved beneficial with linear models such as gaussian mixture models.

The formula for converting from frequency to Mel scale is:

$$M(f) = 1125 \ln(1 + f/700) \tag{1}$$

4.2 Melspectrogram features

The Melspectrogram is calculated on a time series input using the librosa package. The spectrogram magnitude is calculated which is later mapped to mel scale. The melspectrogram is calculated with the help of following steps:

1. Separate windows: This step involves the sampling of input with a length of windows size 2048. After which the window is shifted 512 samples and the next window is considered.
2. Compute FFT: This step involves calculating the fast fourier transform of each window so that the conversion of the time domain to the frequency domain is done.
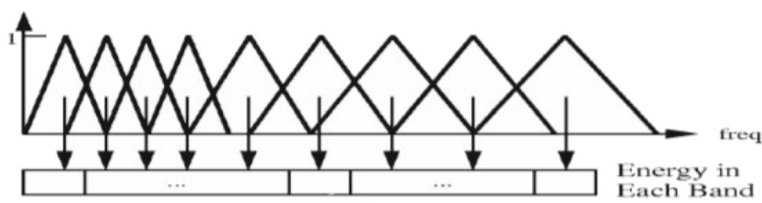

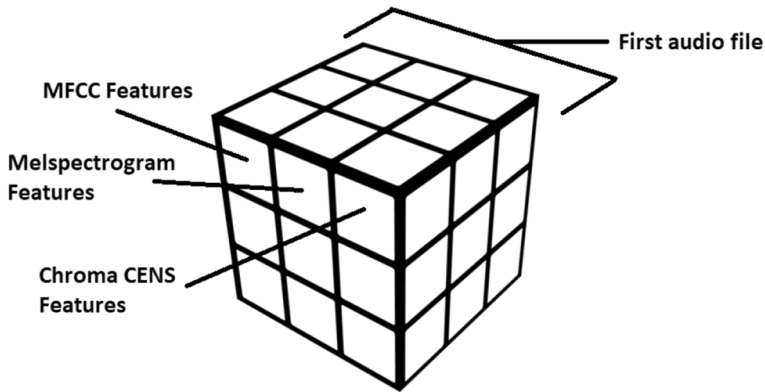**Fig. 2** Mel scale filter bank, from (young et al, 1997)

Fig. 3 Feature visualization

3. Generating a Mel scale: Next, the frequency spectrum is separated into 39 evenly spaced frequencies.

4.3 Chroma energy features

Chroma energy normalized statistics: The main idea of Chroma Energy Normalized Statistics (CENS) features is that by matching the similarity of audio sound with the selected features by taking statistics over large windows smooths local deviations in musical ornaments, tempo, and articulation such as chords and trills. The CENS features are calculated where each chroma is first normalized with the help of L1- norm which helps express the relative energy distribution. The next step is the application of quantization based on the chosen threshold.The last step is smoothing and downsampling.

5 Execution and implementation scenario

The implementation of the algorithm is carried out on a machine having 8GB DDR4 RAM, a 4Gb Nvidia GTX1050 Graphics Processing Unit (GPU), and an Intel Core i7-7700HQ Central Processing Unit (CPU) 2.80GHz which is a 64-bit processor. The implementation of this multi-class classification starts with the following order

1. Data Augmentation: A delay function is introduced which introduces a delay of 250ms, 500ms, 750ms and 1000ms in the audio files. The delayed audio and normal audio are added and saved in the local machine to help augment the data.
2. Feature Transformation: The 3 features namely MFCC, Melspectrogram, Chroma CENS are used from the audio samples. This happens in a way that first the audio is sliced in equal parts of 30ms and is passed from a pre-emphasis filter. After which the 3 features are transformed from the audio with the librosa[1] library and stacked together. The audio is loaded as a floating-point time series and keeping the sampling rate the same, with the help of the resample type as kaiser_best. The feature vector can be visualized by Fig. 3.

---

[1]https://github.com/librosa/librosa

Table 2 Original Data Size and Augmented data size

ID	Name of Disease	#Audio Sample	#Augmented samples
1	Bronchiectasis	16	80
2	Bronchiolitis	13	65
3	COPD	793	3965
4	Healthy	35	175
5	Pneumonia	37	185
6	LRTI	2	10
7	URTI	23	115
8	Asthma	1	5

3. Dividing Data: The selected dataset of audio file is split in three partitions that can be used for testing, training, and validation. The feature vectors are created based on split size, the split size is chosen as 70%,20% and 10% for the training, testing, and validation respectively. Keeping in mind that the samples used in validation are the original samples and not the augmented samples.
4. Reshaping Data: For the input of CNN we need to reshape the data, this is done using the NumPy[2] package in python.
5. Classifier: Then the feature vectors are passed to a CNN classifier which classifies them in different categories based on the probability given to them by the softmax layer.

As the dataset contains just 920 files there is a need to augment the data for the CNN layers to properly train on the dataset. The data augmentation step helps us augment the data so now the newly created dataset has 5 augmented audio files with every file at a delay of 250 ms. The number of augmented files is given in Table 2. So the final augmented dataset has a total of 4600 audio files. Next, the features transformation is done and a feature vector of 4600 x 39 x 3 is created. We can visualize this with the example of a regular image where the red layer is the MFCC layer, the green layer is the Melspectrogram features and the blue layer is the Chroma CENS features. So this tells us that every audio file has 3 features corresponding to a class. The next step is the reshaping of the data, this is done with the NumPy package so the final dimension of the feature vector becomes 4600 x 39 x 3 x 1. This helps us feed the feature vector in the CNN architecture. Figure 4 shows us healthy audio sample amplitude with respect to time (seconds). Next, the architecture is defined in Section 4.1. The model is trained for 100 epochs. The significance of our preprocessing step is that it helps include all features of the audio file and convolution layers have shown to perform greatly in the task of image classification so the CNN layers will help us correctly predict the classes of the dataset. The final softmax layer helps give probabilities to different classes. So the output of the CNN is one of the 3 cases, which helps us carry out our diagnosis of chronic disease, non-chronic disease, or healthy audio sample.

For a robust classification, we have applied K-fold cross validation. We have applied 5 folds to the model. So in every fold, the model is trained on different splits of the training and testing dataset and the validation dataset is kept separate on which the model is validated. This helps understand how our model performs on unseen data. So the model is trained on training and testing feature vectors and validated on validation feature vector. The mean
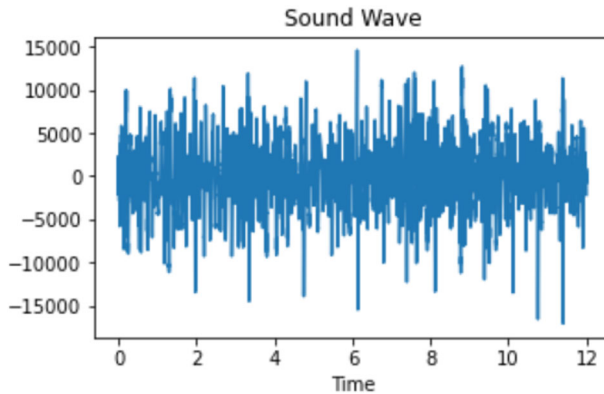
---

[2]https://github.com/numpy/numpy

Fig. 4 Plot of a healthy audio

recall obtained is 0.991, F1 score of 0.993 and precision of 0.994. The entire algorithm is shown as a pseudo code in Algoithm 1.

5.1 CNN architecture for our proposed approach

The CNN architecture contains 11 layers the first being a Conv2D layer with 64 filters, kernel size 3, stride as 1, activation as 'relu' and padding argument as 'same' which lets the output have the same length as the original input. The second layer being a max-pooling layer with padding kept as same. Third layer is a Conv2D layer with 128 filters, with the same parameters as the first layer. The fourth layer is a max-pooling layer with the padding kept as 'same'. Fifth layer is a dropout layer with dropout size kept at 0.3 which specifies the rate and so the probability of each input will be dropped by 30%. Sixth layer is a flatten layer which helps flatten the nodes. The seventh layer is a dense layer with 256 nodes and the activation kept as 'relu'. The eighth layer is a dropout layer with the same specification as the fifth layer. Ninth layer being a dense layer again with 512 nodes with 'relu' activation. Tenth layer being a dropout layer with the same specification as the fifth layer. The last layer being a dense layer with 3 nodes and the activation kept as 'softmax'. The optimizer used for this model is an Adam optimizer with its default parameters. The model can be visualized in Fig. 5.
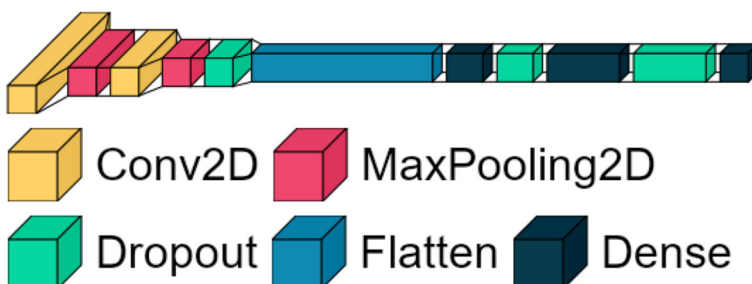

Fig. 5 CNN architecture

Kongtao Chen et al. [14] proposed that high computational and storage requirements sometimes stymie the implementation of convolutional neural networks. Structured model pruning is a possible way to get around these constraints. We show that systematic model pruning on TPUs may considerably reduce model memory use and performance without sacrificing accuracy, particularly for small datasets (e.g., CIFAR-10).

Algorithm 1 Model creation.

Require: Audio file (.wav)

1: **procedure** Augment(audio files)　▷ Add a delay of 250ms,500ms,750ms,1000ms to audio file and save
2: **end procedure**
3: **procedure** Matrix(audio files) ▷ Calculates MFCC, Melspectrogram, Chroma CENS features from audio files
4:　　**while** audio file in directory **do**　　　　　　　▷ Loops over every file in directory
5:　　　　$a \leftarrow MFCC, Melspectrogram, ChromaCENS$　　　▷ Transform MFCC, Melspectrogram, Chroma CENS feature using librosa
6:　　　　$b \leftarrow Name$　　　　　　　　　　　　　▷ Get the label of the file
7:　　**end while**
8:　　$a, b \leftarrow a.reshape, b.reshape$　　　　　　　▷ Converting into proper format
9:　　**return** a,b
10: **end procedure**
11: **procedure** Make model　　　　　　　　　　　　　▷ Creation of model
12:　　**while** layers **do**
13:　　　　$model \leftarrow model.layers$　　　　　　▷ Defining model architecture
14:　　**end while**
15:　　$model \leftarrow model.compile$　　　　　　　▷ Compiling the model
16:　　**return** model
17: **end procedure**
18: **procedure** Divide data　　　　　　　　　　　▷ Train, test and validation data
19:　　$[training + testing], validation feature \leftarrow [a, b], c$　　　▷ Splitting data 7:2:1
20:　　$x\_train, x\_test, y\_train, y\_test \leftarrow a, b$　　　▷ Diving the data using scikit-learn
21:　　**return** x_train,x_test,y_train,y_test
22: **end procedure**
23: **for** <number of epochs> **do**
24:　　<Pass train and test data into model>
25: **end for**
26: **save** $model.h5$　　　　　　▷ This model file can be used to make predictions

---

6 Experiment results

As a baseline each feature was first independently used for the task of classification. These independent features were passed through the same CNN network to correctly identify the 3 classes. The experimental results are shown in Table 3. As observed from the table, these independent features were not able to perform well. After which the concatenation of these features were done to improve the performance of the model.

Table 3 Analysis of independent features

Approach	Dataset	Classes	Precision(%)	F1-Score	Recall
MFCC	ICBHI	3(chronic, non-chronic, normal)	90.5	0.85	0.89
Melspectrogram	ICHBHI	3 (normal, chronic, non-chronic)	84.5	0.84	0.82
Chroma CENS	ICHBHI	3 (normal, chronic, non-chronic)	94.5	0.900	0.986

Three metrics have been selected for testing the performance of the model. Precision, Recall and F1-score. Where Precision is

$$\text{precision} = \frac{TP}{TP + FP} \tag{2}$$

where TP stands for true positive, FP stands for false positive. Recall is

$$\text{recall} = \frac{TP}{TP + FN} \tag{3}$$

where FN is false negative and TP is true positive. And F1-score is defined as

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

The evaluation of our proposed approach is done on the ICBHI corpus and plots of accuracy and loss have been plotted in in Figs. 6 and 7 respectively. We can see from the plots that there is no overfitting of the data and our proposed approach is better than previous approaches for the classification of diseases. The comparison from [59] also shows that our proposed method outperforms the previous approach. The comparison between the approaches is shown in Table 3. The newly proposed approach at preprocessing helps transform 3 kinds of features and in turn, helps us get the most of the information from the audio sample.

The method performs better than previous techniques as the transformation of different features helps understand enough information from the audio sample for proper classification of classes. The preprocessing proposed in this paper is unique than previous methods
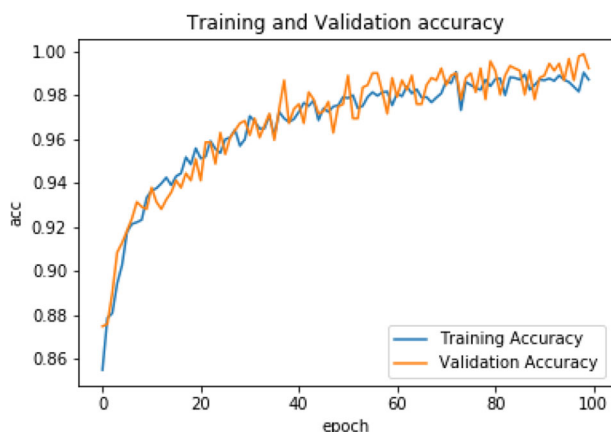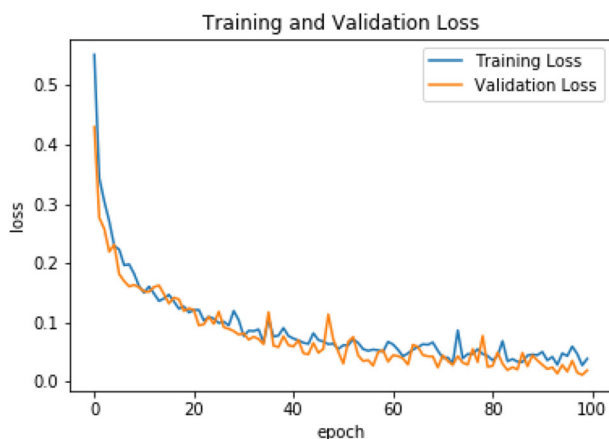

Fig. 6 Training and Validation Accuracy

Fig. 7 Training and Validation Loss

and as Table 4 suggests that the classification of a feature vector produced by preprocessing helps outperform previous approaches .

Figure 8 shows the comparison of features and highlights the first 10 coefficients (for viewing the difference in amplitude between different features) and their amplitude. From the figure we can see the variation of features in the audio sample. As we can see that the plot of a healthy audio sample in Fig. 4, the audio after preprocessing can be visualized as Fig. 8 (but with 39 features).

The combining of the features helps the model learn a better pattern of lung sounds to help it classify audio samples better. MFCC and Melspectrogram features provide the log scaled features and Chroma CENS provide us with matching the similarity of audio sound with the selected features by taking statistics over large windows smooths. These features together depict a far better understanding of the audio sample compared to a single feature alone. The results also point us to a similar direction.

Apart from this a few noise suppression techniques were looked at to help improve the quality of audio samples. But recent noise supression techniques degraded the quality of the audio samples as these techniques have a hard time differentiating lung sound with background noise. The noise supression techniques are better suited with speech and background noise problems.

Table 4 Experimental results

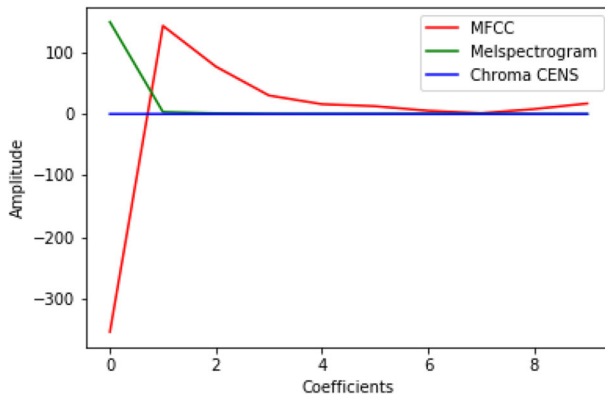| Approach	Dataset	Classes	Precision	F1-Score	Recall
2D CNN [59]	ICBHI	3(chronic, non-chronic, normal)	0.96	0.84	0.82
RNN [60]	ICHBHI	3 (normal, chronic, non-chronic)	0.93	0.91	0.90
CNN [21]	ICHBHI	3 (normal, chronic, non-chronic)	0.994	0.900	0.986
Proposed Approach	ICBHI	3(chronic, non-chronic, normal)	**0.9940**	**0.9930**	**0.9910**

Fig. 8 Comparison of features

7 Conclusion and future work

This paper aims at a new and effective approach towards the classification of respiratory diseases which can help in the biomedical field for early detection of the diseases. Our approach dwells upon a novel feature transformation method along with data augmentation which help us produce state-of-the-art performing model. The dataset used to evaluate the model is the ICBHI corpus [66] which is a publicly available dataset for research. The proposed approach combines pre-processing steps MFCC, Melspectrogram, and Chroma CENS with CNN, it helps for making an accurate diagnosis of lung sounds. Moreover, proposed approach outperforms previous approaches and presents a new way of preprocessing of audio files. The proposed approach was able to reach an recall of 0.991, F1 score of 0.993 and precision of 0.994. This works helps us extract maximum information from the audio sample and leverage it for a better, robust and accurate diagnosis. For future work, we plan to include other features like spectrogram, wavelet transform, FFT, or STFT and evaluate their performance for the task of respiratory pathology diagnosis. Along with improving performance this service can be deployed on edge devices to deliver, for example it can be deployed on a jetson nano and made available to public for usage.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. World health organization cardiovascular disease. https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), Online; Accessed on 24.04.2020
2. World health organization density of pyhsicians. https://www.who.int/gho/health_workforce/physicians_density/en/, [Online; Accessed on 24.04.2020]
3. Alsmadi S, Kahya YP (2008) Design of a dsp-based instrument for real-time classification of pulmonary sounds. Comput Biol Med 38(1):53–61
4. Aly M, Rahouma KH, Ramzy SM (2022) Pay attention to the speech: Covid-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings. Alexandria Engineering Journal 61(5):3487–3500

5. Aras S, Gangal A, Bülbül Y. (2015) Classification of healthy and pathologic lung sounds recorded with electronic auscultation. In: 2015 23Nd signal processing and communications applications conference (SIU). IEEE, pp 252–255

6. Aykanat M, Kılıç Ö, Kurt B, Saryal S (2017) Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing 2017(1):1–9

7. Bardou D, Zhang K, Ahmad SM (2018) Lung sounds classification using convolutional neural networks. Artif Intell Med 88:58–69

8. Chambres G, Hanna P, Desainte-Catherine M (2018) Automatic detection of patient with respiratory diseases using lung sound analysis. In: 2018 International conference on content-based multimedia indexing (CBMI). IEEE, pp 1–6

9. Charleston-Villalobos S, Martinez-Hernandez G, Gonzalez-Camarena R, Chi-Lem G, Carrillo JG, Aljama-Corrales T (2011) Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. Comput Biol Med 41(7):473–482

10. Chen CH, Huang WT, Tan TH, Chang CC, Chang YJ (2015) Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors 15(6):13132–13158

11. Chen CH, Huang WT, Tan TH, Chang CC, Chang YJ (2015) Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors 15(6):13132–13158

12. Chen CH, Huang WT, Tan TH, Chang CC, Chang YJ (2015) Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors 15(6):13132–13158

13. Chen H, Yuan X, Pei Z, Li M, Li J (2019) Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. IEEE Access 7:32845–32852

14. Chen K, Franko K, Sang R (2021)

15. Dokur Z (2009) Respiratory sound classification by using an incremental supervised neural network. Pattern Anal Applic 12(4):309–319

16. Du X, Cai Y, Wang S, Zhang L (2016) Overview of deep learning. In: 2016 31St youth academic annual conference of chinese association of automation (YAC). IEEE, pp 159–164

17. Emmanuel BS (2012) A review of signal processing techniques for heart sound analysis in clinical diagnosis. Journal of Medical Engineering & Technology 36(6):303–307

18. Fiz JA, Jané R., Lozano M, Gomez R, Ruiz J (2014) Detecting unilateral phrenic paralysis by acoustic respiratory analysis. PLoS One 9(4):e93595

19. Flietstra B, Markuzon N, Vyshedskiy A, Murphy R (2011) Automated analysis of crackles in patients with interstitial pulmonary fibrosis. Pulm Med 2011

20. Forgacs P (1978) The functional basis of pulmonary sounds. Chest 73(3):399–405

21. García-Ordás MT, Benítez-Andrades JA, García-Rodríguez I, Benavides C, Alaiz-Moretón H (2020) Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. Sensors 20(4):1214

22. Gibson GJ, Loddenkemper R, Lundbäck B, Sibille Y (2013) Respiratory health and disease in europe: the new european lung white book

23. Goudarzi S, Moradi MH (2015) Dynamical modeling of respiratory sound an aproach for pulmunary patients classification. In: 2015 22Nd iranian conference on biomedical engineering (ICBME). IEEE, pp 70–75

24. Gross V, Dittmar A, Penzel T, Schuttler F, Von Wichert P (2000) The relationship between normal lung sounds, age, and gender. Am J Respir Crit Care Med 162(3):905–909

25. Güler İ, Polat H, Ergün U. (2005) Combining neural network and genetic algorithm for prediction of lung sounds. J Med Syst 29(3):217–231

26. Güler İ, Polat H, Ergün U. (2005) Combining neural network and genetic algorithm for prediction of lung sounds. J Med Syst 29(3):217–231

27. Haider NS, Singh BK, Periyasamy R, Behera AK (2019) Respiratory sound based classification of chronic obstructive pulmonary disease: a risk stratification approach in machine learning paradigm. J Med Syst 43(8):1–13

28. Hasan MR, Jamil M, Rahman M et al (2004) Speaker identification using mel frequency cepstral coefficients. Variations 1(4):565–568

29. Hashemi A, Arabalibiek H, Agin K (2011) Classification of wheeze sounds using wavelets and neural networks. In: International conference on biomedical engineering and technology, vol 11. IACSIT press, pp 127–131

30. Islam MA, Bandyopadhyaya I, Bhattacharyya P, Saha G (2018) Multichannel lung sound analysis for asthma detection. Comput Methods Prog Biomed 159:111–123

31. Jakovljević N., Lončar-turukalo T (2018) Hidden markov model based respiratory sound classification. In: Precision medicine powered by phealth and connected health. Springer, pp 39–43

32. Jin F, Krishnan S, Sattar F (2011) Adventitious sounds identification and extraction using temporal–spectral dominance-based features. IEEE Trans Biomed Eng 58(11):3078–3087

33. Kahya YP, Yeginer M, Bilgic B (2006) Classifying respiratory sounds with different feature sets. In: 2006 International conference of the IEEE engineering in medicine and biology society. IEEE, pp 2856–2859

34. Kandaswamy A, Kumar CS, Ramanathan RP, Jayaraman S, Malmurugan N (2004) Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 34(6):523–537

35. Kraman S (1986) Lung sounds for the clinician. Arch Intern Med 146(7):1411–1412

36. Ku Y, Kwon SB, Yoon JH, Mun SK, Chang M (2022) Machine learning models for predicting the occurrence of respiratory diseases using climatic and air-pollution factors. Clinical and Experimental Otorhinolaryngology

37. Lee ES, Kim JY, Yoon YH, Kim SB, Kahng H, Park J, Kim J, Lee M, Hwang H, Park SJ (2022) A machine learning-based study of the effects of air pollution and weather in respiratory disease patients visiting emergency departments. Emergency Medicine International 2022

38. Lema GF, Berhe YW, Gebrezgi AH, Getu AA (2018) Evidence-based perioperative management of a child with upper respiratory tract infections (urtis) undergoing elective surgery; a systematic review. International Journal of Surgery Open 12:17–24

39. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M (2014) Medical image classification with convolutional neural network. In: 2014 13Th international conference on control automation robotics & vision (ICARCV). IEEE, pp 844–848

40. Liao S, Gao Y, Oto A, Shen D (2013) Representation learning: a unified deep learning framework for automatic prostate mr segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 254–261

41. Lin BS, Wu HD, Chen SJ (2015) Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. Journal of Healthcare Engineering 6(4):649–672

42. Lin BS, Wu HD, Chen SJ, Jan GE, Lin BS (2015) Using back-propagation neural network for automatic wheezing detection. In: 2015 International conference on intelligent information hiding and multimedia signal processing (IIH-MSP). IEEE, pp 49–52

43. Loudon R (1987) The lung exam. Clin Chest Med 8(2):265–272

44. Mahapoonyanont N, Mahapoonyanont T, Pengkaew N, Kamhangkit R (2010) Power of the test of one-way anova after transforming with large sample size data. Procedia-Social and Behavioral Sciences 9:933–937

45. Maruf SO, Azhar MU, Khawaja SG, Akram MU (2015) Crackle separation and classification from normal respiratory sounds using gaussian mixture model. In: 2015 IEEE 10Th international conference on industrial and information systems (ICIIS). IEEE, pp 267–271

46. Matsunaga S, Yamauchi K, Yamashita M, Miyahara S (2009) Classification between normal and abnormal respiratory sounds based on maximum likelihood approach. In: 2009 IEEE International conference on acoustics, speech and signal processing. IEEE, pp 517–520

47. Mhetre R, Bagal U (2014) Respiratory sound analysis for diagnostic information. IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) 9(5):42–46

48. Muda L, Begam M, Elamvazuthi I (2010)

49. Naves R, Barbosa BH, Ferreira DD (2016) Classification of lung sounds using higher-order statistics: a divide-and-conquer approach. Comput Methods Prog Biomed 129:12–20

50. Palaniappan R, Sundaraj K, Lam C (2016) Reliable system for respiratory pathology classification from breath sound signals. In: 2016 International conference on system reliability and science (ICSRS). IEEE, pp 152–156

51. Palaniappan R, Sundaraj K, Lam C (2016) Reliable system for respiratory pathology classification from breath sound signals. In: 2016 International conference on system reliability and science (ICSRS). IEEE, pp 152–156

52. Palaniappan R, Sundaraj K, Ahamed NU (2013) Machine learning in lung sound analysis: a systematic review. Biocybernetics and Biomedical Engineering 33(3):129–135

53. Palaniappan R, Sundaraj K, Ahamed NU, Arjunan A, Sundaraj S (2013) Computer-based respiratory sound analysis: a systematic review. IETE Tech Rev 30(3):248–256

54. Palaniappan R, Sundaraj K, Sundaraj S (2014) A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. BMC bioinformatics 15(1):1–8

55. Park JS, Kim JH, Oh YH (2009) Feature vector classification based speech emotion recognition for service robots. IEEE Trans Consum Electron 55(3):1590–1596

56. Pasterkamp H, Kraman SS, Wodicka GR (1997) Respiratory sounds: advances beyond the stethoscope. Am J Respir Crit Care Med 156(3):974–987

57. Pasterkamp H, Kraman SS, Wodicka GR (1997) Respiratory sounds: advances beyond the stethoscope. Am J Respir Crit Care Med 156(3):974–987

58. Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in mri images. IEEE Trans Med Imaging 35(5):1240–1251

59. Perna D (2018) Convolutional neural networks learning from respiratory data. In: 2018 IEEE International conference on bioinformatics and biomedicine (BIBM). IEEE, pp 2109–2113

60. Perna D, Tagarelli A (2019) Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. In: 2019 IEEE 32Nd international symposium on computer-based medical systems (CBMS). IEEE, pp 50–55

61. Price J, Student S (2005) Design of an automatic speech recognition system using matlab. Dept. of Engineering and Aviation Sciences. University of Maryland Eastern Shore Princess Ann

62. Reichert S, Gass R, Andrès E (2007) Analyse des sons auscultatoires pulmonaires. IRBM 28(3-4):169–180

63. Reichert S, Gass R, Brandt C, Andrès E (2008) Analysis of respiratory sounds: state of the art. Clinical medicine. Circulatory, respiratory and pulmonary medicine 2, CCRPM–S530

64. Reichert S, Gass R, Brandt C, Andrès E. (2008) Analysis of respiratory sounds: state of the art. Clinical medicine. Circulatory, respiratory and pulmonary medicine 2, CCRPM–S530

65. Riella R, Nohama P, Maia J (2009) Method for automatic detection of wheezing in lung sounds. Braz J Med Biol Res 42(7):674–684

66. Rocha B, Filos D, Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, Natsiavas P, Oliveira A, Jácome C, Marques A et al (2018) A respiratory sound database for the development of automated classification. In: Precision medicine powered by phealth and connected health. Springer, pp 33–37

67. Sankar AB, Kumar D, Seethalakshmi K (2011) Neural network based respiratory signal classification using various feed-forward back propagation training algorithms. Eur J Sci Res 49(3):468–483

68. Schoeffmann K, Münzer B., Riegler M, Halvorsen P (2017) Medical multimedia information systems (mmis). In: Proceedings of the 25th ACM international conference on Multimedia, pp 1957–1958

69. Serbes G, Sakar CO, Kahya YP, Aydin N (2011) Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 3314–3317

70. Society AT et al (1995) Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 152:77S–120S

71. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312

72. Thun MJ, Carter BD, Feskanich D, Freedman ND, Prentice R, Lopez AD, Hartge P, Gapstur SM (2013) 50-year trends in smoking-related mortality in the united states. N engl J med 368:351–364

73. Umeki S, Yamashita M, Matsunaga S (2015) Classification between normal and abnormal lung sounds using unsupervised subject-adaptation. In: 2015 Asia-pacific signal and information processing association annual summit and conference (APSIPA). IEEE, pp 213–216

74. Uwaoma C, Mansingh G (2017) On smartphone-based discrimination of pathological respiratory sounds with similar acoustic properties using machine learning algorithms. In: ICINCO (1), pp 422–430

75. Yadav A, Dutta MK, Prinosil J (2020) Machine learning based automatic classification of respiratory signals using wavelet transform. In: 2020 43Rd international conference on telecommunications and signal processing (TSP). IEEE, pp 545–549

76. Yamashita M, Matsunaga S, Miyahara S (2011) Discrimination between healthy subjects and patients with pulmonary emphysema by detection of abnormal respiration. In: 2011 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 693–696

77. Zorc JJ, Hall CB (2010) Bronchiolitis: recent evidence on diagnosis and management. Pediatrics 125(2):342–349

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.