International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 13, No.1 (Jan-2023)

http://dx.doi.org/10.12785/ijcds/130126

Voice Based Pathology Detection from Respiratory Sounds using Optimized Classifiers

Vipul Chudasama[1], **Krina Bhikadiya**[1], **Sapan H Mankad**[1], **Ajaykumar Patel**[1] **and Maunil P Mistry**[2]

[1]*Department of Computer Science and Engineering,Nirma University, Ahmedabad, India*
[2]*Department of Power Electronics,Vishwakarma Governent Engineering College , Ahmedabad, India*

Received 29 Apr. 2022, Revised 13 Dec. 2022, Accepted 12 Jan. 2023, Published 31 Jan. 2023

Abstract: Speech is an important tool for communication. When a person speaks, the vocal cords come closer and the glottis is partially closed. The airflow which passes through glottis is disturbed by vocal cords and speech waveform is produced. The person who suffers from the vocal cord paralysis or vocal cord blister, his lungs are filled with fluid and airway blockage cannot generate a similar waveform as a healthy person. In this work, we compare traditional approaches with deep learning based approaches for respiratory disease detection to distinguish between a healthy person and the victim of pathological voice disorder. Four conventional machine learning classifiers and a one-dimensional convolution neural network based classifier have been implemented on two benchmark datasets ICBHI 2017 and Coswara. Our experiments show that the CNN based approach and Random Forest algorithm exhibit superior performance over other approaches on ICBHI 2017 and Coswara datasets, respectively.

Keywords: respiratory disease detection, audio, Coswara, ICBHI 2017

1. INTRODUCTION

Voice is the easiest medium of communication for humans. The pathological vocal fold vibration generates some changes in speech generation. Using signal processing methods, we can differentiate between normal and pathological speech. The vocal folds vibration is directly related to breathing, so it can directly affect the generation of speech signal. Thus, speech is an important parameter for voice pathology detection. Some methods used to detect larynx condition are as follows. The physician can use a technique to identify the voice disorder through the examination of the voice quality with the help of some specialist doctor, or can use some equipment like telescopic oral and nasal fiber optic. They also may use endoscopy methods to detect the pathological voice [1].

Many diseases are related to pathological voice disability. These include Alzheimer, Parkinson, and Depression. Some respiratory diseases like Chronic Obstructive Pulmonary Disease (COPD), Upper Respiratory Tract Infection (URTI), Lower Respiratory Tract Infection (LRTI), Bronchiolitis, Asthma, and Pneumonia are also common. All these diseases directly affect the voice signal so we can apply voice processing techniques and detect the presence of the disease. In this paper, we mainly focus on respiratory disease or cough related disease.

Cough is the primary symptom in respiratory diseases. Most of the pulmonary diseases are auscultated through the anterior and posterior chest wall using the stethoscope. The physician examines different lung sounds which are crackle, wheeze sound and stridors. Lungs are important organs in our respiratory system. When the human breathes in and out, he inhales oxygen from the air into the blood and exhales the carbon dioxide from the blood to the air. Lungs have two main components: airways and alveoli [2]. The airway is used to inhale and exhale the air. The presence of any respiratory disease indicates lack in lungs efficiency. The respiratory diseases are classified into three types: (i) Obstructive, (ii) Restrictive and (iii) combination of obstructive and restrictive. In obstructive respiratory disease, person's air exhaling capacity from lungs may decrease [2]. In obstructive disease, the lungs airways become widening and narrowing so exhaling capacity reduces. In restrictive respiratory disease, the lungs are filled with fluid so it results into expansion of lungs [2]. In a combination of obstructive and restrictive disease, the patient starts suffering from short breathing, cough and pain in the chest [2]. Different diseases have different sound quality and can be classified using the sound features like frequency, energy, intensity, wavelength, pitch. For example, the wheeze and stridors both are the high-frequency sounds. High pitch wheeze

E-mail address: vipul.chudasama@nirmauni.ac.in,

sounds are produced due to the high breathing which indicate asthma and COPD. The stridors sounds are mostly due to the obstruction in the airflow. The airway is blocked because of the blockage of the larynx.

Auscultation is safest, non-invasive, cheapest and user-friendly method to analyze the lungs sounds. First, auscultation requires accurate and perfect training. If this auscultation is done by some expert then also it may be risky, hence the computer-aided system performs the automatic respiratory system analysis. Researchers have proposed many methods for respiratory disease identification. The sound characteristics are examined first and then extracted features are fed into the classification model for prediction of the disease. Such a system may be mainly divided into two phases: (i) feature extraction from the audio sample (ii) prediction of the input sound a pathological or a healthy voice.

A. Human Voice Generation System

The human voice generation system contains lungs, larynx and vocal tract (Fig 1). The lungs are the main organ in the human voice generation system. During the breathing, one can inhale air from the rib cage surrounded by the lungs and exhale air from the diaphragm which are the bottom of the lungs [3]. The airflow is controlled by rib cage depending on the length of the speech or sentence. The larynx consists of the cartilages, muscles and ligament [4]. The larynx is used to control the vocal folds. The gap between two vocal folds is referred to as glottis. In a breathing state, the vocal folds are in relaxed condition and glottis is open, so air is passed through the glottis. Depending on the scenario of glottis, two kinds of speech is generated: voiced speech and unvoiced speech. In voiced speech, the mass of the vocal folds come closer and glottis is partially closed. The airflow which passes through glottis is disturbed by vocal cords and periodic waveforms are generated. In unvoiced condition, the vocal folds come closer and generate disturbance.

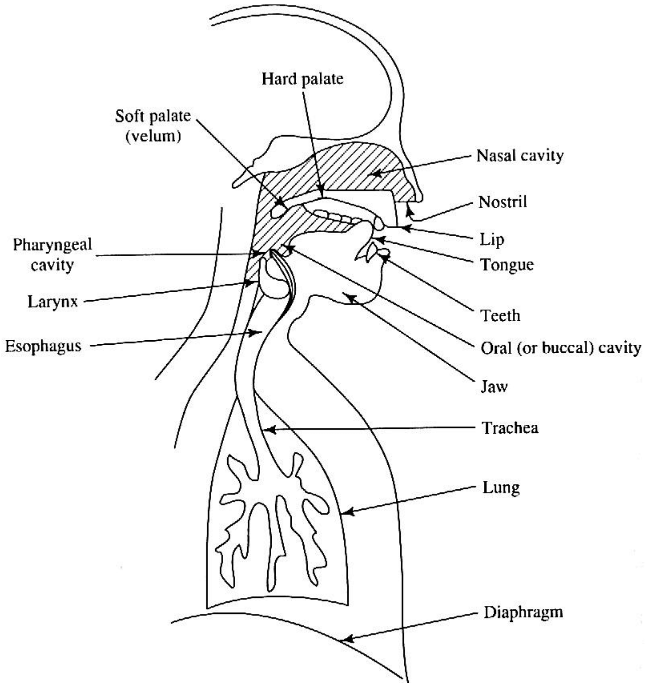

Figure 1. Human voice generation system [3]

B. Voice Disorder Generation

The vocal cords play an important role for voice generation. All causes of voice disorder are still not identified. The voice disorder is generated by vocal cord paralysis, vocal cord blister, vocal cord swelling etc. Some other diseases (such as brain injury, neurological disorder, or mental disability) can also affect voice disorder. The person can also suffer from a temporary disorder like tonsils, swelling in the throat, some kind of allergies or some smoking habit. The physician can use endoscopy to detect voice disorder in which a probe is entered into the throat for examination. However, this method is painful for the patient. Thus, voice disorder detection using audio processing may be effective in terms of painless diagnosis. It can detect the voice disorder in less time, less cost and mainly it is not painful to humans.

C. Why clinic based COVID-19 testing is not sufficient to stop spreading?

China, South Korea and Singapore and many more countries adopt the "Trace, Track and Treat" strategy [5]. Still the pandemic is spread in whole over the world that proves that this strategy is not enough. The 81% of the COVID-19 is spread out because symptoms are not detected at an early stage and people did not visit the clinic because of fear, and those became active spreaders [5]. After that countries adopted a new strategy of personal testing, and if detected as positive then isolate them in one place and this way reduces the pandemic effect. But still, this method was not enough to reduce the pandemic spreading due to these reasons:

- The limited capability of testing because of the geographical location.
- This testing requires that a person should visit a hospital, clinic and lab. This leads to crowd in one place, and breaches the rule of social distance. In this crowd, if one person is diagnosed as positive then there is a high chance of others getting infected.
- In this type of testing, the medical staff has a high risk if they don't have enough protection equipment and measures.

2. Related Work

Researchers have made several attempts to develop the methods to identify diseases using speech processing. In this section, we discuss existing literature for disease detection through voice.

Rudraraju et al. [2] explained the importance of lungs in detection of respiratory diseases. They classified COPD, URTI, Bronchiectasis, Bronchitis, Pneumonia, and Asthma. They used zoom handy recorder to record the audio sounds. Audio samples were represented using 40-dimensional MFCC, Zero Crossing Rate, Energy, Spectral Centroid, Spectral Roll-off, Spectral Bandwidth etc. as primary features. In addition, the cough type (Dry/Wet), cough duration, cough frequency, kurtosis were used as secondary features. On decision tree, Random Forest and XGBoost, they obtained overall 91.97% accuracy, 87.2% sensitivity and 93.69% specificity.

Rumana Islam et al.[4] considered that the generated signals are noisy, hence it is difficult to find the meaningful information. The digital pre-processing techniques like Mel Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficient (LPCC), and Perceptual Linear Prediction (PLP) were applied on the speech signal. The MFCC and PLP feature extraction techniques were applied on Massachusetts Eye and Ear Infirmary (MEEI) dataset. Various classifiers such as Support vector machine (SVM), Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Artificial Neural Network (ANN), Deep Neural network were appplied on MEEI dataset. The GMM and HMM classifiers reported the accuracy of 94.56% and 90.52% respectively with male and female subjects.

In [5], Imran et al. showed that the clinical COVID-19 test is not advisable due to possibility of infection spread and outbreak. They explained the need for AI4COVID–19 system which can diagnose COVID 19 base on sounds. They signified the use of cough as the unique feature to distinguish between a COVID affected person and normal person. However, for overlapping cases, it becomes difficult to clearly differentiate them. On ESC-50 dataset, they classified sounds into four categories: normal person, COVID-19 person, pneumonia person and bronchitis. On mel-spectrogram representation of audio samples, they got 92.64% accuracy on a CNN based model.

Haizhen et al.[6] worked on the depression detection from the speech. The mental health disorder can be considered as the classification or regression problem. The normal depression methods are Beck Depression Index, Hamilton Rating Scale for depression, and Quick inventory for depressive symptomatology. They applied basic interview process assessment between physician and patient. The speech signals represented as spectrograms were generated on AVEC dataset for feeding into a CNN based model and accuracy with 70.05% was reported.

Mingyu et al. [7] demonstrated in their work approach for cough detection from speech. The cough is a common symptom and has a serious effect on the patients' life. In their approach, the speech signal was down-sampled (from 44.1kHz to 16kHz) to reduce the cost of transmission followed by windowing. After that, MFCC and GFCC features were extracted and the final feature vector was passed through a classifier model. A comparison between cough and normal speech showed that the cough energies are scattered between the whole area. They used HMM, GMM and SVM as classification model. In cough detection system, the noise data reduced the accuracy. So in this paper, the authors worked on synthesized data, applied some ensemble methods like threshold voting and subband methods. The result showed 79.3% (MFCC), 77.6% (GFCC) and 80.9% (Subband) and 81.1% threshold voting accuracy.

Carlos et al. [8] presented disease detection approaches using nonlinear features. Voice impairment is directly related to nonlinear pressure flow in the glottis, nonlinear vocal fold and nonlinear vocal fold collisions. In [8], the nonlinear dynamic features were calculated using Correlation dimension (CD) and were used to identify the sum of the number of possible pair plot at given distance r. Largest Lyapunov Exponent (LLV) was used to find the average divergence of neighbor in space. Lample Ziv Complexity (LZC) was used to identify different patterns into the sentence. Two classification methods based on HMM were used. HMM worked on the transition probability of the two states. In DHMM, the decision was taken from the previous input vector. The result of the DHMM was given to the SVM kernel, and classification algorithm was evaluated. They used three databases, namely, Cleft Lip and palate (CLP), Parkinson Disease (PD) and Laryngeal pathologies (LP) and achieved performance of 85%, 70% iand 90%, respectively.

Gabor et al. [9] showed detection of mild cognitive impairment and Alzheimer disease. They mainly focused on vocal and linguistic features. Feature extraction was done manually based on articulation rate, speech tempo, and hesitation ratio. To analyze the linguistic features they used semantic features, morphological feature, and demographic features. They collected the data from the hospital using a recording of 75 subjects. On Hungarian MCI-mAD database, the performance of around 76 to 78% was obtained.

Muhammad et al [10] developed automatic voice pathology detection system wherein features were extracted from vocal tract area related to the glottis. They explained that vocal tract was connected through vocal folds and glottis. For feature extraction, they used the signal to noise ratio, harmonic to noise ratio, spectral flatness, and pitch amplitude. They used the MEEI database which contains 53 normal and 173 pathological people. The performance on SVM classifier was 99.22%.

3. Analysis of Speech on different diseases

The diseases which can be diagnosed using speech are Asthma, Parkinson, Alzheimer, Depression, COPD, Cough, LRTI, Pneumonia, and CoViD-19. In this section, we describe the representational differences among all these diseases.

A. Asthma

In Asthma, the vocal cords get swollen and hence they cannot properly vibrate. Due to this, the voice of the patient becomes thick and weak. To determine the difference between a normal and asthmatic person, both people need to speak five minutes. During this period, the total number of syllable per breath, word duration, number of pauses, pause duration, and total speaking time by both persons is measured [4]. The waveforms of a healthy and asthmatic person are shown in Fig 2 which indicate that the asthmatic person takes more and longer pause during the speech.

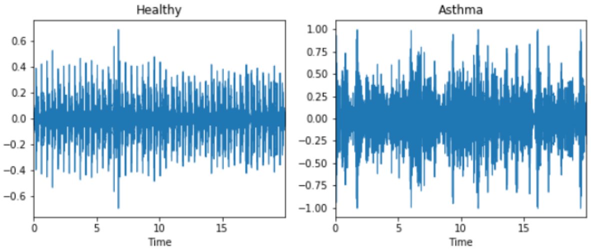

Figure 2. Waveforms representing the speech of a healthy and Asthma Person.

B. Alzheimer

In Alzheimer, the person suffers from a mental disorder, inability to retrieve information from memory, memory loss, and mood swing. Alzheimer is diagnosed when the symptoms start interfering into the daily activity. Patients who suffer from Alzheimer also have anemia (i.e. difficulty in word-finding) and it often creates problem in word list generation [11]. The waveforms of a healthy and Alzheimer person are shown in Fig 3. This difficulty directly affects on spontaneous speech generation. Based on acoustic features, the Alzheimer disease patients have variation in the rhythm of pronunciation, varying pitch level, word-finding pause, and slowness in their speech [11].

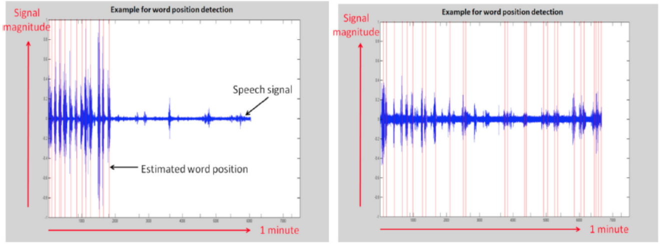

Figure 3. (Left) Healthy person audio sample and (Right) Alzheimer person audio sample [11].

C. Parkinson's Disease (PD)

In Parkinson's disease, the person loses the neurons in his brain so that the other body parts are also affected. These neurons produce the dopamine which carries a message from mid-brain to another part of the brain. It is used to control the whole body movements [12]. The waveforms of a healthy and Parkinson affected persons are shown in Fig 4. The PD patients also suffer from slowness in movement, rigidity, and imbalance posture. They speak slowly and roughly, and their voice tone becomes decreased so they experience difficulty in pronouncing a word correctly. They also feel depressed.

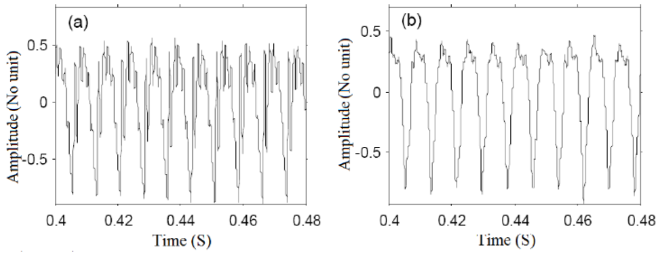

Figure 4. (Left) Waveform of a healthy person and (Right) waveform of a person with Parkinson's disease [12].

D. Depression

The depression directly affects the person's mood, thoughts, behavior and alignments. Each individual's voice quality is directly related to the mental state information and these vocal cords features are used as bio-markers to detect the depression [13]. The voice quality features are detected using jitter, shimmer, and change in vibration of vocal cords [13]. The waveforms of a normal and a depressed person are shown in Fig 5 and 6.

E. Cough

Cough is the most common and important symptom of respiratory disease. Cough is directly related to the respiratory tract and is responsible for internal infections in the vocal tract. Cough can be classified into two types: wet cough and dry cough [2]. The waveforms of a normal person and a person with cough are shown in Fig 7. When the cough sounds contain the saliva, it is known as wet

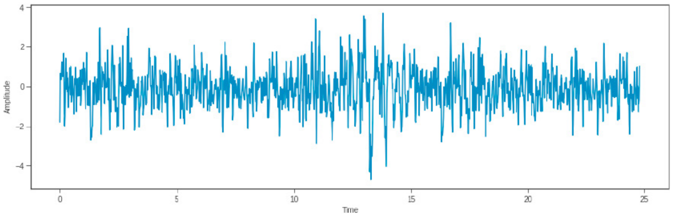

Figure 5. Speech of a normal person

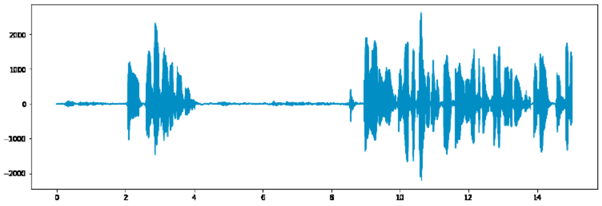

Figure 6. Speech of a depressed person

cough and absence of wetness is known as dry cough. Some changes into cough sound directly affect on lungs pathology condition [2].

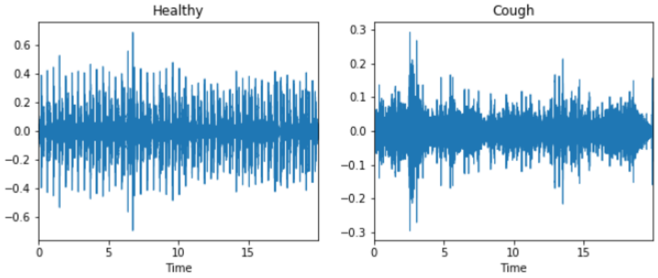

Figure 7. Waveforms of normal and coughing persons

F. *Pneumonia*

In pneumonia, people face difficulty in breathing, and the infection can be caused in one or both lungs. The lungs are filled with fluid and create difficulty in breathing. The main symptoms of pneumonia are greenish, yellow, and bloody cough. So it is also one type of respiratory disease. Another main symptom of pneumonia is crackled voice. The physicians also use the crackle sound (based on low and high frequency of the crackled sound) to see possibility of pneumonia [14]. The waveform of a healthy and pneumonic person are shown in Fig 8.

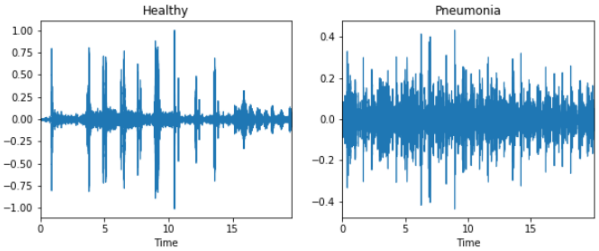

Figure 8. Healthy and pneumonic persons sound sample

G. *COVID-19*

The main symptoms of COVID-19 are dry cough, difficulty in breathing, loss of the speech movement, chest pain etc. COVID-19 and pneumonia are different. The waveform of the healthy person's breathing and COVID - 19 person's breathing are shown in Fig 9.

4. Feature Extraction from Speech

Audio signals are classified into three types[15] : human speech, music and environmental sounds. The speech is produced by different organs like lungs, mouth, nose etc. The speech production starts at frequency 100Hz and goes to up to 17kHz [15]. The music sound is generated using different musical instruments. The music contains different characteristics like genre, mood and sound property. The music frequency starts from 40Hz and ranges upto 190.5 kHz [15]. The environmental sound is generated by a different number of devices like car, doorbell, water, animals, factory noise etc.

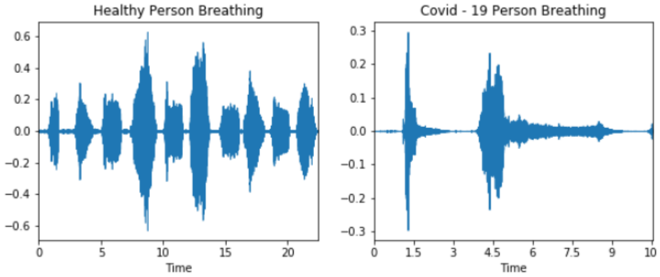

Figure 9. Healthy and COVID-19 person breathing sound sample

Features from audio can be derived in variety of ways such as time, frequency and cepstral domain features.

A. *Time Domain features*

This method analyzes the signal into its original form. The time-domain analysis is applied onto the short term energies of the signal assuming stationary characteristics through framing. Fig 10 describes feature extraction from time domain representation of the signal.

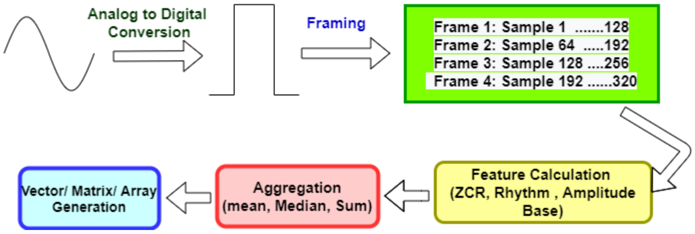
Analog to Digital Conversion
Framing
Frame 1: Sample 1 128
Frame 2: Sample 64 192
Frame 3: Sample 128 256
Frame 4: Sample 192 320
Feature Calculation (ZCR, Rhythm , Amplitude Base)
Aggregation (mean, Median, Sum)
Vector/ Matrix/ Array Generation

Figure 10. Time domain feature extraction

1) *Zero-Crossing Rate (ZCR)*

It is the rate of amplitude change in the signal from positive to negative and vice versa. The ZCR mainly is used to detect the voiced, unvoiced and silence portion. The ZCR value is higher in unvoiced condition compared to voice condition. For the silence portion, the ZCR is zero. ZCR is mainly useful in speech discrimination, music classification, voice detection, and vowel detection and analysis [15].

B. *Frequency Domain features*

The time-domain features are used to find out time-related information from the signal. However, they fail to reveal significant details of the signals, hence the time-domain signal is converted into frequency domain features using Fourier transform (Fig 11).

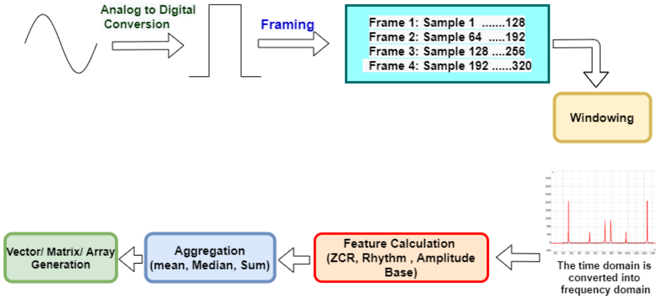
Analog to Digital Conversion
Framing
Frame 1: Sample 1 128
Frame 2: Sample 64 192
Frame 3: Sample 128 256
Frame 4: Sample 192 320
Windowing
The time domain is converted into frequency domain
Feature Calculation (ZCR, Rhythm , Amplitude Base)
Aggregation (mean, Median, Sum)
Vector/ Matrix/ Array Generation

Figure 11. Frequency domain feature extraction

1) *Time-Frequency based features*

In time-domain, features show how the amplitude of the signal is changed over time. The frequency signal gives the frequency-related information from the signal. So time-frequency both together give both information. Speech is a continuous signal so short-time Fourier transform (STFT) is used to catch up relevant information. Using STFT, the signal is converted into the time-frequency representation. Spectrograms are similar features belonging to time-frequency analysis.

2) *Chroma Related Features*

It is used to gain the tonal related information from the musical sound. Chroma spectrum is similar to spectrogram representation, it works on 12 different intensity musical octave [16]. Without knowing the frequency of the original audio signal, the chroma features give musical audio information. From audio signal to chroma representation, the signal suffers some information loss.

3) *Spectrum shape based features*

The spectrum means how signal changes their phase of frequency and magnitude.

- Spectrum Centroid: It shows the center of mass of the spectrum. It is used to find the brightness of the signal. It is computed based on the frequency and probabilities value. It used to measure the timbre information of sound, music classification. Timbre means character, texture and color of the sound. The spectrum centroid of music is higher compared to spectrum centroid of speech.

- Spectral Spread: It is related to the bandwidth of the signal. For noise-related sounds, the spread is more compared to the real tonal sound. It is mostly used for environmental and musical sound detection.

- Spectral Skewness: It measures the symmetry of the signal based on the mean value. Its value is zero for the silent part and high for the voice part. The skewness value describes the energy distribution. If skewness is zero, it denotes symmetric distribution; if skewness is negative, the energy distribution is on right side; otherwise it is on the left side. It is used for music genre classification, Parkinson disease detection.

- Spectrum kurtosis: It is used to measure the flatness of the signal based on their mean value. If kurtosis is less than zero than there is flat distribution and if kurtosis is greater than zero then high peaked spectrum in the signal. It is used in Parkinson disease detection, mood classification.

4) *Cepstral Domain features*

The cepstral analysis is used to differentiate between the vocal tract and source excitation form speech.

- Mel Frequency Cepstral Coefficients (MFCC): MFCC is derived from the audio clip. In MFCC, first windowing is applied, then discrete cosine transform is applied, then take the log of the magnitude and warping frequency in mel scale (pitch listen by the listener are equal from one part to another part). This mel frequency is near to human auditory system. MFCC is used in music classification, speech recognition, speech enhancement, vowel detection. The MFCC extraction flow is shown in Fig 12.

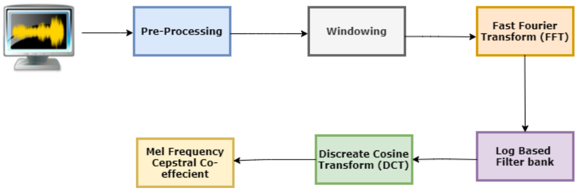
Pre-Processing
Windowing
Fast Fourier Transform (FFT)
Log Based Filter bank
Discreate Cosine Transform (DCT)
Mel Frequency Cepstral Co-effecient

Figure 12. MFCC Work Flow

- Linear Prediction Cepstral Coefficient (LPCC): The LPCC features are used to perform source-filter separation, orthogonality, compactness.

5. Dataset

In this work, we have experimented with two datasets : (i) ICBHI Challenge dataset (ii) a crowd-sourced COVID-19 dataset [17] hosted by Indian Institute of Science (IISc) Bangalore where people can record their cough and breathing sound.

A. *ICBHI dataset*

The ICBHI database is publicly available as part of the challenge announced by the International Conference on Biomedical and Health Informatics 2017. This dataset contains a total of 126 participants and 920 different lungs sound records [18]. The lung sound is collected anterior and posterior chest location based on the disease requirement. The recording is done with a stethoscope, it is put tight on chest location to reduce the human artifacts. The sound recording is done based on a disease like upper left and right of anterior/posterior, middle left or right and lower left or right. A total of 8 classes, namely, COPD, URTI, LRTI, Bronchitis, Bronchiectasis, Asthma, Pneumonia, and healthy are part of this database.

B. Coswara dataset

The COVID-19 dataset is downloaded from the GitHub repository [1]. This repository contains data from Coswara project[2]. The Coswara project is started by IISc Bangalore, which records the person's respiratory, breathing and cough sound using the web application. They record the person breathing sound (fast/slow), cough sound (deep/shallow), counting number (fast/normal), phonation vowel sound (/a/e/o). Thus, a total of 9 files are recorded by one person.

The Coswara COVID-19 dataset contains some empty file folders, so first, the file duration was obtained and the files with zero duration were removed from the dataset. Finally, a total of 369 people's sound recording were used for prediction.

6. Proposed Methodology

The architecture of the proposed approach is shown in Fig 13.

A. Machine Learning Model

Classification models are used to predict the disease from speech signal. The classifiers used in this work are k-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, and XGBoost.

B. Evaluation Measures

The performance of the proposed system is evaluated using various parameters including F1-Score, recall, precision and accuracy.

C. CNN Architecture

Mainly CNN architecture consists of three layers: convolution layer, pooling layer and fully connected layer. The filter size of the model gradually increases at each layer. To reduce the overfitting, L2 regularization and drop out layer is used. The learning is nonlinear so the ReLU activation function is used. The architecture of the system is shown in Fig 14. The parameter details of the network for respiratory disease detection and COVID-19 detection are listed in Table II and III.

7. Results and Discussion

In this section, we demonstrate the performance of proposed approach on both datasets and discuss the results.

A. Experiments on ICBHI dataset

1) Results on Machine Learning model

The result of all machine learning models on ICBHI dataset are shown Tables IV through VII.

To summarize these classwise results on multiple classifiers, we carried out Borda ranking approach [19] to determine which classifier performed the best and which class label was predicted correctly for the most of the times overall.

TABLE I. Performance measures used in this work

True Positive (TP)	The predicted positive value is the same as the actual value. It means the predicted and actual both values are "yes".
True Negative (TN)	The predicted negative value is the same as the actual value. It means the predicted and actual both values are "no".
False Positive (FP)	The predicted value is "yes" and the actual class is "no".
False Negative (FN)	The predicted value is "no" and the actual class is "yes".
F1-Score	It is calculated based on the average of the precision and recall. Accuracy is good estimator when the dataset is balanced. But when the dataset is unbalanced that time f1-score is good estimator because it is the average of precision and recall.
Recall	It is known as the ratio of correct prediction of labels to the sum of correct prediction labels and false prediction labels. In medical domain system recall is the good estimator because recall shows the false-negative value means the patients are suffering from the initial stage of COVID-19 but the system shows healthy.
Precision	It is the ratio of correct predicted value to the sum of correct prediction label and false-positive label.
Accuracy	It means the ratio of correctly predicted value to the total value in dataset.

TABLE II. Parameters for the respiratory disease detection model on ICBHI 2017

Network Type	Input	Filters	Stride	Kernel	Activation
MFCC Input	130 * 13	-	-	-	-
Conv. layer	130*13*32	32	2,2	3*3	ReLU
Conv. layer	64*4*32	32	2,2	3*3	ReLU
Conv. layer	30*1*32	32	2,2	2*2	ReLU
Flatten	480	-	-	-	-
Dense	64	-	-	-	-
Dropout	0.30	-	-	-	-
Output	8	-	-	-	Softmax

Figure 15,Figure 16,Figure 17 shows the performance comparison of these four classifiers in terms of precision, recall and F1 score.

A quantitative comparison of the proposed approach on ICBHI 2017 dataset is shown in Table VIII. This

[1] https://github.com/iiscleap/Coswara-Data

[2] https://coswara.iisc.ac.in/

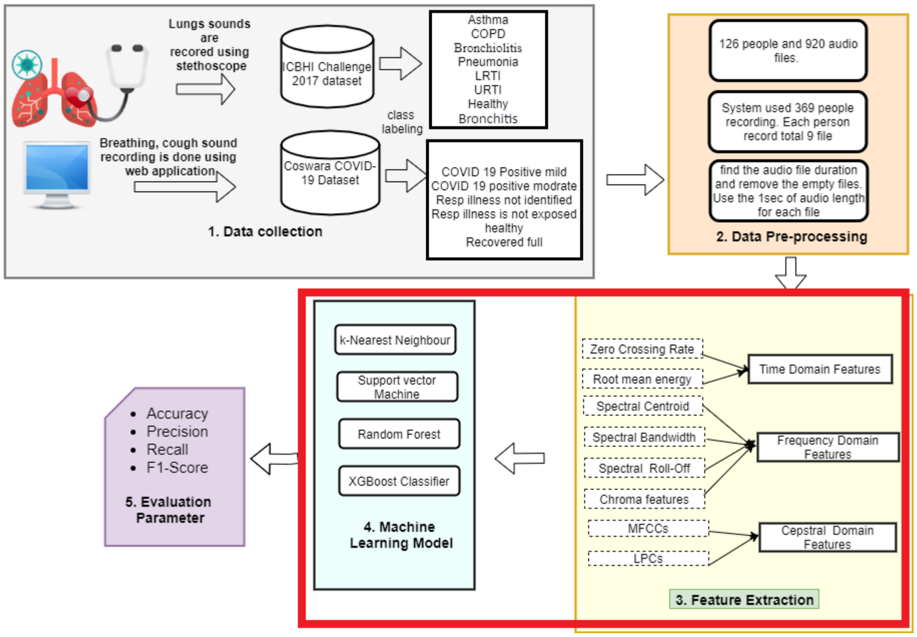

Figure 13. Architecture of the proposed approach

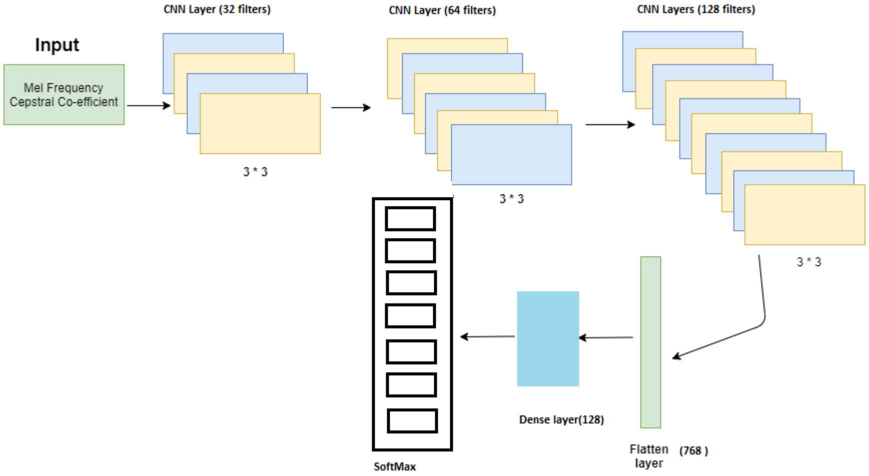

Figure 14. CNN Architecture

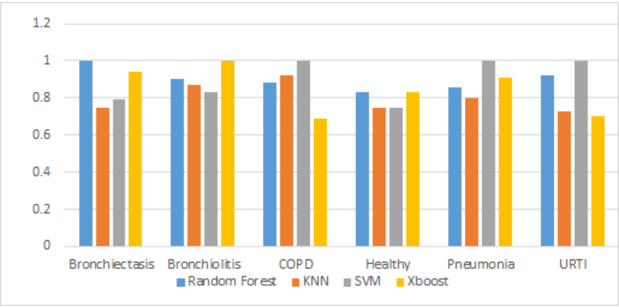

Figure 15. Precision

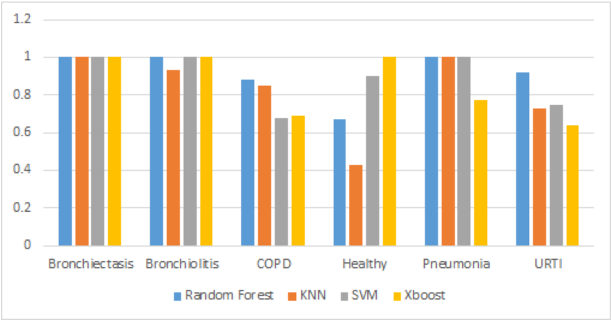

Figure 16. Recall

TABLE III. Parameters for the COVID-19 detection model on Coswara dataset

Network Type	Input	Filters	Stride	Kernel	Activation
MFCC Input	52 * 13	-	-	-	-
Conv. layer	52*13*32	32	2,2	3*3	ReLU
Conv. layer	25*6*32	64	2,2	3*3	ReLU
Conv. layer	12*2*64	128	2,2	3*3	ReLU
Flatten	768	-	-	-	-
Dense	128	-	-	-	-
Dropout	0.50	-	-	-	-
Output	7	-	-	-	Softmax

TABLE IV. Random Forest on ICBHI Challenge dataset

Disease Name	Precision	Recall	F1-Score
Bronchiectasis	1.00	1.00	1.00
Bronchiolitis	0.90	1.00	0.95
COPD	0.88	0.88	0.88
Healthy	0.83	0.67	0.74
Pneumonia	0.86	1.00	0.92
URTI	0.92	0.92	0.92

TABLE V. KNN on ICBHI Challenge dataset

Disease Name	Precision	Recall	F1-Score
Bronchiectasis	0.75	1.00	0.86
Bronchiolitis	0.87	0.93	0.90
COPD	0.92	0.85	0.88
Healthy	0.75	0.43	0.55
Pneumonia	0.80	1.00	0.89
URTI	0.73	0.73	0.73

TABLE VI. SVM on ICBHI Challenge dataset

Disease Name	Precision	Recall	F1-Score
Bronchiectasis	0.79	1.00	0.88
Bronchiolitis	0.83	1.00	0.91
COPD	1.00	0.68	0.81
Healthy	0.75	0.90	0.82
Pneumonia	1.00	1.00	1.00
URTI	1.00	0.75	0.86

TABLE VII. XGBoost on ICBHI Challenge dataset

Disease Name	Precision	Recall	F1-Score
Bronchiectasis	0.94	1.00	0.97
Bronchiolitis	1.00	1.00	1.00
COPD	0.69	0.69	0.69
Healthy	0.83	1.00	0.91
Pneumonia	0.91	0.77	0.83
URTI	0.70	0.64	0.67

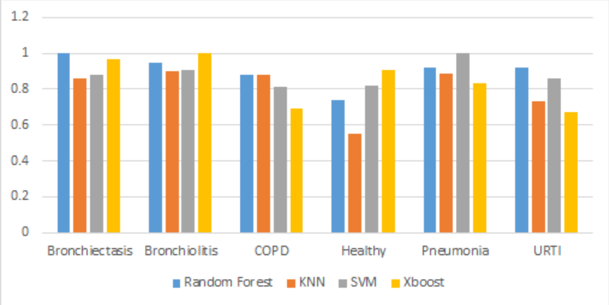

Figure 17. F1 score

indicates the strong impact of Random Forest in comparison to other classifiers. The ranking achieved from Borda method also places Random Forest first, followed by SVM, XGBoost and kNN. Another insight which we gain from Borda method is that all classifiers are good at classifying Bronchiectasis, Bronchiolitis and Pneumonia patients in comparison to other three classes of ICBHI dataset. Overall, random forest classifier gives the best performance, and most of the methods find it slightly difficult to classify healthy persons as healthy. This shows there is huge scope to design strategies to avoid such false alarms specifically applicable in medicine domain.

2) *Results on Deep Learning model*

The CNN based approach fails to outperform the conventional classifier model (Table IX). Since deep classifiers work well when training data are huge in size, the impact of such models may be enhanced either through augmentation or transfer learning.

B. *Experiments on Coswara dataset*

1) *Results on Machine Learning model*

In Coswara dataset for feature extraction ZCR, spectral roll-off, spectral centroid, spectral bandwidth, chroma related features, root mean energy, MFCCs (with 13 coefficients) were used as audio representation. Three classifiers were implemented: Random forest, KNN and XGBoost. The result of all machine learning based models on Coswara dataset are shown in Tables X through XII.

Table XIII shows a comparison of the proposed approach on Coswara dataset with existing approaches in literature. The random forest based approach outperforms all systems. Figures 18, 19, 20, depict the comparison of three classifiers on Coswara dataset for individual classes. Again, it can be seen that for healthy class, the performance is not upto the satisfactory level.

The top performer random forest was further exploited using 13-dimensional LPC features based audio representation. Table XIV shows the classwise results and Table XV shows the overall performance of the same experiment.

TABLE VIII. Comparison of proposed approach with existing results on ICBHI 2017 challenge dataset

Reference	Classifier	Accuracy	Recall	F1 score
[20]	MFCC-HMM	39.56	—	—
[21]	low level features, Decision Tree	49.62	20.81	—
[22]	STFT+Wavelet, SVM	57.88	—	—
[23]	BiResNet	52.79	31.12	—
[24]	transfer learning CNN + softmax	63.09	—	—
[24]	deep features + SVM	65.5	—	—
[25]	deep features with CNN + LDA	71.15	—	—
Our approach	Random Forest	**90.9**	91.07	90.15
Our approach	kNN	80.51	82.27	80.02
Our approach	SVM	87.01	88.9	88.01
Our approach	XGBoost	85.71	84.96	84.51

TABLE IX. CNN Model for ICBHI Challenge Dataset

Classifier	Accuracy	Precision	Recall	F1 Score
CNN	0.9406	0.7127	0.9870	0.6241

TABLE X. Random Forest on Coswara dataset

Disease Name	Precision	Recall	F1-Score
Healthy	0.94	0.89	0.92
No resp illness exposed	0.96	0.96	0.96
Positive asymp	1.00	1.00	1.00
Positive Mild	0.93	0.98	0.95
Positive Moderate	1.00	1.00	1.00
Recovered Full	1.00	1.00	1.00
Resp illness not identified	1.00	1.00	1.00

TABLE XI. KNN on Coswara dataset

Disease Name	Precision	Recall	F1-Score
Healthy	1.00	0.36	0.53
No resp illness exposed	0.86	0.98	0.92
Positive asymp	0.95	1.00	0.98
Positive Mild	0.90	0.98	0.94
Positive Moderate	0.90	1.00	0.95
Recovered Full	0.89	1.00	0.94
Resp illness not identified	0.84	1.00	0.92

TABLE XII. XGBoost on Coswara dataset

Disease Name	Precision	Recall	F1-Score
Healthy	0.90	0.69	0.78
No resp illness exposed	0.84	0.98	0.91
Positive asymp	0.98	1.00	0.99
Positive Mild	0.94	0.91	0.92
Positive Moderate	0.96	1.00	0.98
Recovered Full	0.98	1.00	0.99
Respillness not identified	0.89	0.95	0.92

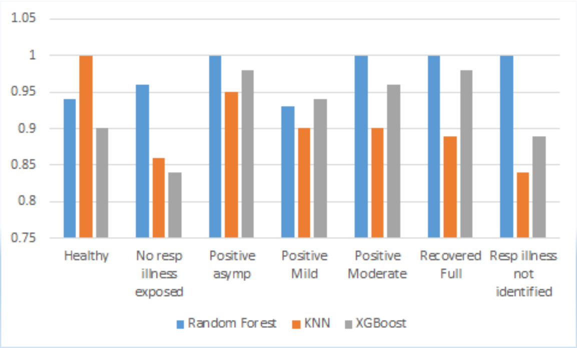

Figure 18. Precision

2) *Results on Deep Learning model*

Table XVI shows the performance of CNN model on Coswara datset. In this case also, it gives poor performance showing limitation of deep learning approaches in small-scale dataset classification.

8. Conclusion

In this work, we demonstrated conventional machine learning based systems for predicting respiratory diseases from recorded audio signals. Experiments on ICBHI 2017 and Coswara dataset suggest that the proposed systems are good enough to detect such diseases in most of the cases, however, deep learning based mechanism fails to deliver good results in case of Coswara dataset. This issue may be resolved by adopting a larger database, or augmenting into the existing dataset, or through transfer learning. Among all classifiers used in this study, the random forest

TABLE XIII. Comparison of proposed approach with existing results on Coswara dataset

Reference	Classifier	Accuracy	Recall	F1 score
[26]	Logistic Regression	75.7	94.0	—
[26]	kNN	74.7	83.0	—
[26]	SVM	73.91	74.0	—
[26]	MLP	87.51	88.0	—
[26]	CNN	94.57	90.0	—
[26]	LSTM	94.02	91.0	—
[26]	ResNet50	95.33	93.0	—
Our approach	Random Forest	97.42	97.62	97.60
Our approach	kNN	89.42	90.36	88.15
Our approach	XGBoost	92.85	93.24	92.72

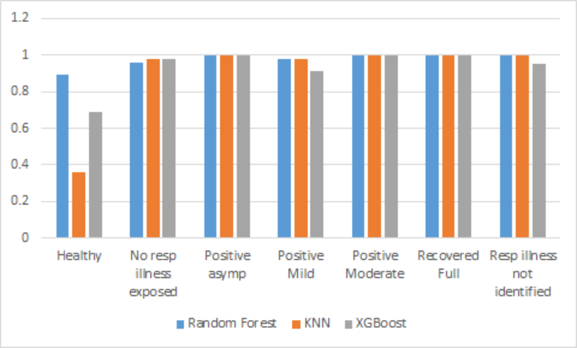

Figure 19. Recall

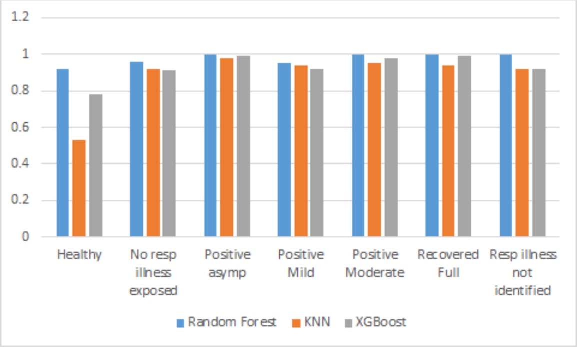

Figure 20. F1 score

algorithm achieved reasonable performance. In future, other approaches of deep learning may be attempted. Further, spectrogram based audio representations may also help in improving the overall system performance.

TABLE XIV. Classwise Performance of Random Forest on Coswara dataset using LPC Feature Extraction

Disease Name	Precision	Recall	F1-Score
Healthy	1.00	0.76	0.87
No resp illness exposed	0.96	1.00	0.98
Positive asymp	1.00	1.00	1.00
Positive Mild	0.90	1.00	0.95
Positive Moderate	0.96	1.00	0.98
Recovered Full	0.98	0.98	0.98
Resp illness not identified	0.93	0.98	0.95

TABLE XV. Random Forest on Coswara dataset using LPC Feature Extraction

Classifier Name	F1-Score	Recall	Accuracy
Random Forest	0.9580	0.9601	0.9571

TABLE XVI. Performance on CNN Model for Coswara dataset

Classifier Name	F1-Score	Recall	Accuracy	Precision
CNN	0.2925	0.2527	0.7031	0.4438

References

[1] G. Vaziri, F. Almasganj, and R. Behroozmand, “Pathological assessment of patientsâ€™ speech signals using nonlinear dynamical analysis,” *Computers in Biology and Medicine*, vol. 40, no. 1, pp. 54 – 63, 2010.

[2] G. Rudraraju, S. Palreddy, B. Mamidgi, N. R. Sripada, Y. P. Sai, N. K. Vodnala, and S. P. Haranath, “Cough sound analysis and objective correlation with spirometry and clinical diagnosis,” *Informatics in Medicine Unlocked*, vol. 19, p. 100319, 2020.

[3] X. Huang, A. Acero, H.-W. Hon, and R. Reddy, *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR Upper Saddle River, 2001, vol. 95.

[4] R. Islam, M. Tarique, and E. Abdel-Raheem, “A survey on signal processing based pathological voice detection techniques,” *IEEE Access*, vol. 8, pp. 66 749–66 776, 2020.

[5] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel, “Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via

an app," *Informatics in Medicine Unlocked*, vol. 20, p. 100378, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.imu.2020.100378

[6] H. An, X. Lu, D. Shi, J. Yuan, R. Li, and T. Pan, "Mental health detection from speech signal: A convolution neural networks approach," in *2019 International Joint Conference on Information, Media and Engineering (IJCIME)*, 2019, pp. 436–439.

[7] M. You, Z. Liu, C. Chen, J. Liu, X.-H. Xu, and Z.-M. Qiu, "Cough detection by ensembling multiple frequency subband features," *Biomedical Signal Processing and Control*, vol. 33, pp. 132 – 140, 2017.

[8] C. M. Travieso, J. B. Alonso, J. Orozco-Arroyave, J. Vargas-Bonilla, E. NÃ¶th, and A. G. Ravelo-GarcÃ¬a, "Detection of different voice diseases based on the nonlinear characterization of speech signals," *Expert Systems with Applications*, vol. 82, pp. 184 – 195, 2017.

[9] "Identifying mild cognitive impairment and mild alzheimerâ€™s disease based on spontaneous speech using asr and linguistic features," *Computer Speech & Language*, vol. 53, pp. 181 – 197, 2019.

[10] G. Muhammad, G. Altuwaijri, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, and A. Al-nasheri, "Automatic voice pathology detection and classification using vocal tract area irregularity," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 2, pp. 309 – 317, 2016.

[11] A. KÃ¶nig, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, and R. David, "Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112 – 124, 2015.

[12] R. A. Shirvan and E. Tahami, "Voice analysis for detecting parkinson's disease using genetic algorithm and knn classification method," in *2011 18th Iranian Conference of Biomedical Engineering (ICBME)*, 2011, pp. 278–283.

[13] K. Chlasta, K. WoÅk, and I. Krejtz, "Automated speech-based screening of depression using deep convolutional neural networks," *Procedia Computer Science*, vol. 164, pp. 618 – 628, 2019.

[14] K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, "Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1185–1194, 2015.

[15] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Applied Acoustics*, vol. 158, p. 107020, 2020.

[16] ""chroma features analysis and synthesis" https://labrosa.ee.columbia.edu/matlab/chroma-ansyn/.[accessed on: 08 – oct -2020]."

[17] N. Sharma, V. PrashantKrishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, R. Nirmala, P. Ghosh, and S. Ganapathy, "Coswara - a database of breathing, cough, and voice sounds for covid-19 diagnosis," in *INTERSPEECH*, 2020.

[18] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A. M. Ibnian, and M. Alkhodari, "Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 1 – 14, 2021.

[19] P. Emerson, "The original borda count and partial voting," *Social Choice and Welfare*, vol. 40, pp. 353–358, 2013.

[20] N. Jakovljevic and T. Loncar-Turukalo, "Hidden markov model based respiratory sound classification," in *BHI 2017*, 2017.

[21] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6, 2018.

[22] G. Serbes, S. Ulukaya, and Y. Kahya, "An automated lung sound preprocessing and classification system based on spectral analysis methods," 2018.

[23] Y. Ma, X. Xu, Q. Yu, Y. Zhang, Y. Li, J. Zhao, and G. Wang, "Lungbrn: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm," *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4, 2019.

[24] F. Demir, A. Şengür, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Information Science and Systems*, vol. 8, 2020.

[25] F. Demir, A. M. Ismael, and A. Sengur, "Classification of lung sounds with cnn model using parallel pooling structure," *IEEE Access*, vol. 8, pp. 105 376–105 383, 2020.

[26] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "Covid-19 cough classification using machine learning and global smartphone recordings," *Computers in Biology and Medicine*, vol. 135, pp. 104 572 – 104 572, 2021.

Dr Vipul Chudasama Dr Vipul Chudasama is working as an Assistant Professor in Computer Science and Engineering Department. He has academic experience of more than 14 years.His research interests include Distributed Computing, Cloud Computing, Parallel Processing, Machine Learning and Artificial Intelligence

Ms Krina Bhikadiya Krina Bhikadiya was student in M.Tech Computer Science and Engineering Department and now working as software developer. Her area of interest includes Machine learning and Deep learning.

Dr. Sapan Mankad Dr. Sapan Mankad is working as an Assistant Professor in Computer Science and Engineering Department. He has more than 17 years of teaching experience. His research interests include Audio and Speech Processing, Voice Biometrics, Machine Learning and Music Information Retrieval.

Prof Ajaykumar Patel Prof Ajaykumar Patel is working as an Assistant Professor in Computer Science and Engineering Department. He has more than 12 years of teaching experience. His research interests are Computer Networks and Vehicular Adhoc Networks.

Mr Maunil P Mistry Maunil P Mistry is a student at Department of Power Electronics,Vishwakarma Governent Engineering College. His area of interest are robotics and machine learning.