iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56640

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Speech Processing on Detection of Respiratory Disease

Saptarshi Pramanik[1], SambitChanda[2], Prof. Dr. Soumik Poddar[3]
[1, 2]*Electronics and Communication Engg., Gurunanak Institute of Technology, Sodepur, India*
[3]*School of Computer Science and Artificial Intelligence, S R University, Warangal, India*

Abstract: ***Speech processing applied to the detection of respiratory diseases has emerged as an innovative and groundbreaking field in the medical domain.***
By analyzing audio signals, specifically cough and respiratory sounds, this cutting-edge approach offers a non-invasive and cost-efficient method for early diagnosis and continuous monitoring of various respiratory conditions. Herein we report a review on the process entailing extraction of crucial features like Mel-Frequency Cepstral Coefficients (MFCCs) and cochlegram image characteristics, inputting the subsequent ones advanced machine learning techniques like Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) for accurate classification.
Despite encouraging outcomes, certain challenges, including limited data availability and variability in cough sounds, pose obstacles to further advancements. Nevertheless, dedicated researchers are actively working on expanding datasets, bolstering the robustness of algorithms, and integrating multimodal data to surmount these hurdles.
The potential benefits of speech processing in respiratory disease detection are vast, encompassing prompt identification, remote assessment capabilities, and personalized medical interventions. Continued collaboration between engineers, healthcare professionals, and patients is essential to fully harness the potential of this technology in revolutionizing respiratory disease diagnosis and improving patient care.
Keywords: speech processing, respiratory disease detection, cough sounds, Mel-Frequency Cepstral Coefficients, machine learning, Support Vector Machines, Convolutional Neural Networks, early diagnosis, remote assessment.

I. INTRODUCTION

Speech processing applied to respiratory disease detection involves the use of sophisticated technologies and algorithms to analyse and interpret audio signals produced during speech, particularly cough and other respiratory sounds. This innovative field has garnered considerable attention in the medical community due to its potential to revolutionize diagnosing early and monitoring of various respiratory ailments.

The purpose of this overview aiming to offer insights into the current state of research concerning speech processing for respiratory disease detection.

Respiratory diseases, including conditions like pertussis, croup, and bronchitis,can pose significant public healthimplications if not promptly detected and treated. Conventional diagnostic methods often rely on subjective clinical assessments and laboratory tests, potentially leading to delays in diagnosis and treatment initiation.

Conversely, speech processing offers a non-invasive and cost-effective approach to analyse cough and speech patterns, presenting an opportunity for faster and precise disease detection.

The process involves converting audio signals into an analysable format, wherein crucial features like *Mel-Frequency Cepstral Coefficients (MFCCs)* and cochlegram image features are extracted, utilizing these features as inputs for ML algorithms, such as *Support Vector Machines (SVM)* and *Convolutional Neural Networks (CNN)*, to classify and differentiate among normal and abnormal respiratory sounds..

II. LITERATURE SURVEY

A. Cough based Algorithm for Automatic Diagonasis of Pertusis

Renard Xaviero Adhi Pramono, Sayed Anas Imtiaz, Esther Rodriguez – Villages modified a method which could diagnose respiratory disease. It is a typical method which includes time-based features, frequency-based features and usage of *MFCCs (Mel-Frequency Cepstral Coefficient).*This approach isused for classifying cough, detection and whooping sound detection. This method achieved some statistics which includes sensitivity - 92.38%, specificity - 90%, PPV – 96.5% and NPV – 79.84%. [1]

©IJRASET: All Rights are Reserved

B. Automatic Croup Diagonisis by using of Cough Sound Recognition.

Roneel V Sharan. Udanta R. Abeyratne, Vinayak R and Paul Porter developed a technique which could diagoni- se respiratory disease. It is a classical approach which uses amalgamation of linear MFCC and CIF produces good results for cough sounding classification which uses a LRM and SVM with stats sensitivity-72.84% and specificity-92.16%.[2]

By studying the literature survey based on these points we have certain spectral features such as-

1) *Cochlegram image Feature*: Cochlegram image features are derived from the cochlear filter bank, providing a visual representation of sound with non- linear frequency scaling based on the human auditory system. These featuresstimulate the frequency analysis of cochlea. This process includes cochlear filtering, logarithmic compression, temporal integration and image-like visualization. Extracted features, such as statistical measures or MFCCs, employed in speech recognition sound classification, and audio-related tasks.
2) *Mel-Frequency Cepstral Coefficient*: It prevalent feature extraction method in speech and audio processing .MFCCs are commonly used in *Automatic Speech Recognition (ASR),* speaker recognition, and other related tasks. [3]

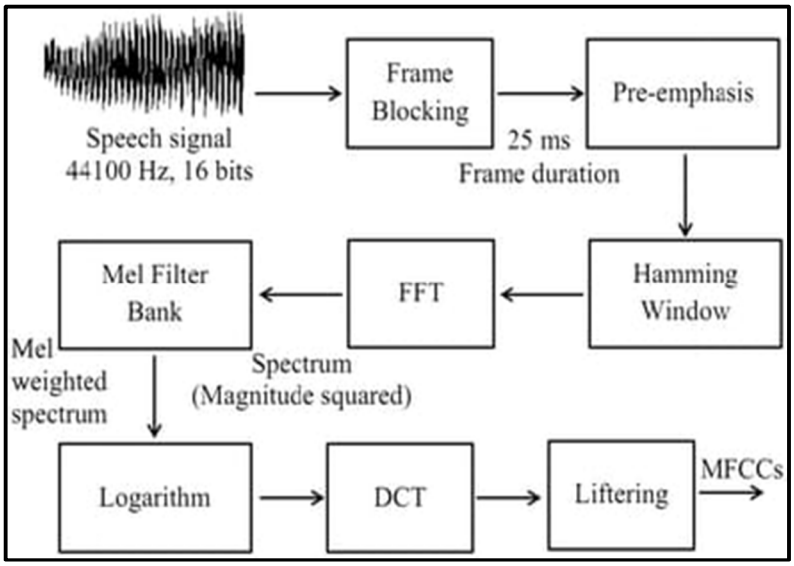


FIGURE 1 | Process for obtaining MFCC using audio/ speech signal. [4]

3) *Machine Leaning Tools:* Such as *SVM (Support Vector Machine), NN (Neural Network), LRM (Logic Regression Model).*

III. METHODOLOGY AND TECHNIQUES USED

Few methodologies used are discussed here below -

A. Speech Modality.

Machine learning algorithms can be used to process data from sound. This applicable forspeech analysis. Speech analysisapplicable for detecting different respiratory and lung diseases. The audio signal is first measured and represented in different temporal and harmonic domains. The harmonic domain is a good basis for deriving features that can be used to detect a lung infection. Basic features include pitch, variation, and irregularity of sound waves. Deriving elevated-level featureas statistical function, over the basic features. One standardized set of basic features is the extended *Geneva Minimalistic Acoustic Parameter Set (eGeMAPS).* It comprises 88 selected features by experts in audio fields [5]. Another standardized set of features is *Computational Paralinguistic ChallengE(ComParE)* features set [6]. It comprises 6,373 Characteristics that have been createdthrough a brute-force combination of numerous basic features and statistical functionals. The other recently developed that are not explicitly reliant onexpert knowledge. This includes deep spectral features, which are based on spectrograms and utilize hidden layers of CNNs, pre-trained on the ImageNet or other image corpora, for feature extraction.[7]

After the initial feature extraction stage, some approach exists that further process these features. A popular approach in this regard is the use of *bag-of-audio words (BOAW)* representations. BOAW representation summarize signal characteristics over time by means of their frequency.[8]

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Automatic speech-based disease detection

Lately, there has been an increasing fascination with using *artificial intelligence (AI)* to detect diseases from speech data. A literature search in PubMed, a leading medical research database, revealed that over 85,000 articles have been published on AI-based speech-based disease detection since 2017.Other investigations have similarly demonstrated that AI Various applications exist for detecting a broad spectrum of diseases, including respiratory diseases, psychiatric disorders, developmental disorders, and neurodegenerative diseases. For example, it has been demonstrated to be able to detect colds and flu with 90% accuracy, COVID-19 with 80% accuracy, anxiety disorder with 70% accuracy, autism spectrum disorder with 85% accuracy, Alzheimer's disease with 75% accuracy, and Parkinson's disease with 80% accuracy.[9,10,11]

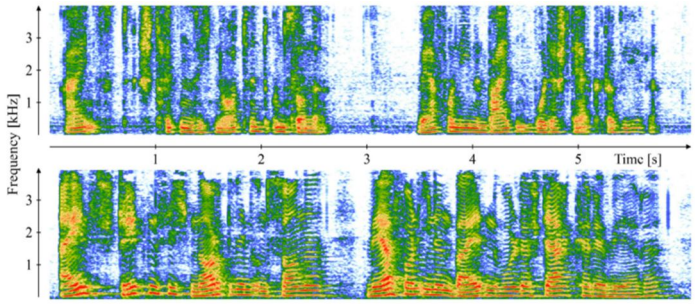


FIGURE 2 | Comparison of a 19-year-old symptomatic male with COVID-19 (top) and a 37-year-old asymptomatic COVID-19 negative male (bottom) by means of speech spectrograms of the recorded first clause of the German standard text “The Northwind and the Sun” [“Einststrittensich Nordwind und Sonne, wer von ihnenbeidenwohl der Stärkerewäre."]. The recordings are part of “Your Voice Counts” dataset [12, 13].

The findings from these studies indicate the same conclusionAI-based speech analysis have the potential to serve as a beneficial aid for early illness detection and diagnosis. However, more research is needed to validate the accuracy of AI-based speech analysis in clinical settings. Additionally, there are challenges to be addressed, like need for large and diverse datasets of speech data from healthy and diseased individuals.Despite these challenges, the progression of AI-based speech analysis is a promising area of research with the capability to improve early disease detection and diagnosis. Continued research and collaboration between engineers, healthcare professionals, and patient stakeholders will be essential to the successful development and implementation of this technology. [9,10,11]

C. Artificial Intelligence.

The most significant breakthroughs in recent AI research have primarily originated from the field of *machine learning (ML)*. ML encompasses various techniques where the algorithm designer only provides a learning framework, enabling the algorithm to learn from training data and make informed decisions. Among the ML subfields, supervised learning plays a crucial role in current automatic disease prediction systems. Here each datum is accompanied by a label that indicates the target of the ML algorithm. Successful algorithms in this domain predominantly fall in the field of parametric ML algorithms, which rely on fixed-sized sets of continuous-valued parameters for decision making. These algorithms undergo an optimization process during training to find a well-suited parameter set that yields high performance on specific evaluation metrics.

For regression tasks, such as predicting continuous values, evaluation metrics like *Root Mean Square Error (RMSE)* or *concordance correlation coefficient (CCC)* are used. For classification tasks, which involve assigning data points to predefined classes, evaluation metrics like *unweighted average recall (UAR)* or accuracy are employed, utilizing the confusion matrix to assess performance.

The general processing framework for most supervised ML tasks follows a similar pattern: preprocessed data, often in the form of features, is fed into an ML algorithm, which is then optimized to achieve high performance based on predefined regression or classification metrics. While details regarding data, preprocessing, ML algorithms, and evaluation metrics may differ from case to case, this basic idea has resulted inimmense success across a wide array of applications. Presently, the most successful technique for numerous ML tasks, such as self-driving cars or text generation, is *deep learning (DL)*. DL relies on *Artificial Neural Networks (ANNs)*, which construct hierarchical structures of neurons and propagate information through matrix multiplications and non-linear functions. These ANNs are divided into different classes based on their architectures [14]. *Feed-forward Neural Networks (FFNNs)* consist of fully-interwoven layers, while *Convolutional Neural Networks (CNNs)* connect consecutive layers through convolution operations with weight filters, akin to traditional image processing.

Furthermore, ANNs aredivided according to the tasks they perform. For example, *Generative Adversarial Networks (GANs)* involve two neural networks compete against each other—one creating authentic artificial data samples from noise, and the other trying to discriminate between fake samples and real samples from the database. [15]. Despite the advancement of deep learning in many ML tasks, more traditional ML algorithms remain popular, especially in the field of small datasets where ANNs may struggle to generalize effectively from training to test data. For speech-based disease detection, datasets often comprise only a few hours of speech, contrasting with larger corpora, such as those used in *Automatic Speech Recognition (ASR)* with several thousand hours of speech [16].

This apparent discrepancy in deep learning dominance in health-related speech tasks has been previously acknowledged. In such cases, traditional ML algorithms, like *Support Vector Machines (SVM*) for classification and *Support Vector Regression (SVR)* for regression, are commonly favored. These approaches involve transforming input features into a higher-dimensional space to separate data points using hyperplanes. In contrast, purely statistic-based analyses, such as those relying on mean and standard deviations of features, are not typically classified as AI methods.

IV. FLOW GRAPH FOR DISEASE DIAGONISIS

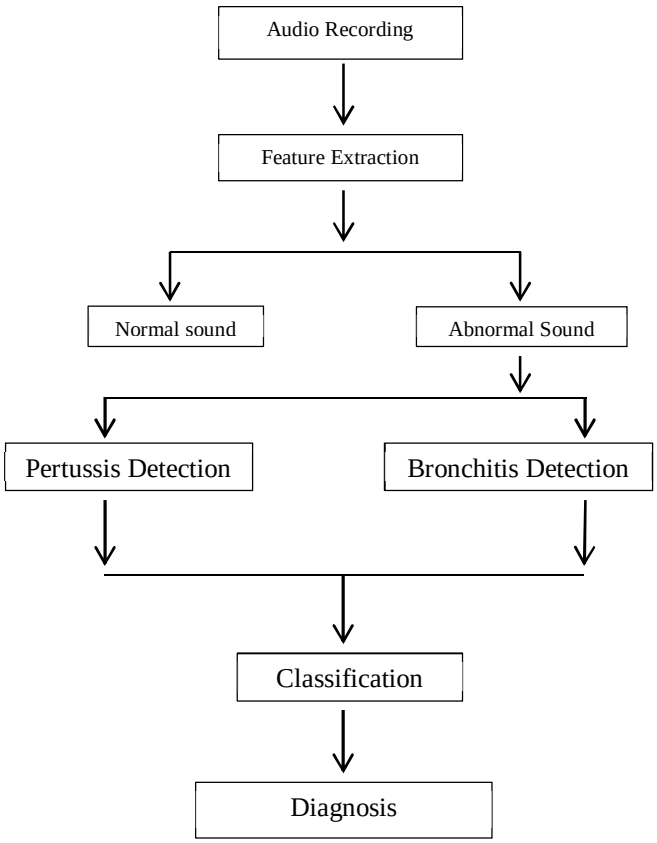
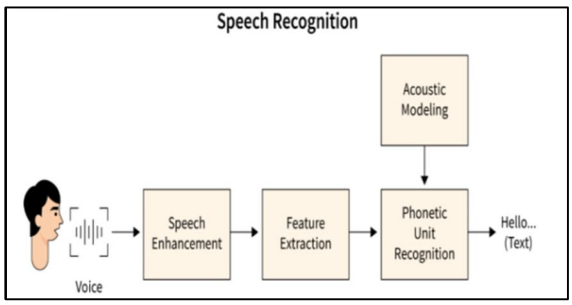


FIGURE 3 | Steps for disease diagnosis [11]



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

V. WORKING OF AI BASED SPEECHRECOGNITION.

1) *Recording:* The built-in sound capturer within the device is utilized to initiate the initial phase. The user's vocal tones are preserved as an auditory transmission subsequent to the recording process.[17]
2) *Sampling*: As per fundamental principles of physics, it is understood that a sound wave exhibits continuity. Hence, for the system to comprehend and handle it effectively, it undergoes transformation into discrete magnitudes. This conversion from continuous to discreet transpires at a specific frequency.[17]
3) *Transforming to Frequency Domain*: During this phase, the temporal characteristics of the audio signal undergo a transformation, transitioning into its frequency domain. This particular stage holds great significance, as the frequency domain offers ample opportunities for in-depth analysis of various audio details. The time domain involves the scrutiny of mathematical functions, physical signals, or chronological sequences of economic or environmental data, in relation to the element of time. Likewise, the frequency domain pertains to the examination of mathematical functions or signals concerning frequency, as opposed to the temporal aspect.[17]
4) *Information Extraction from Audio*: The cornerstone of every voice recognition system lies in this particular stage. During this phase, the audio is converted into a vectorized format that becomes readily employable. Various extraction techniques, such as *PLP (Perceptual Linear Prediction)* and *MFCC (Mel-Frequency Cepstral Coefficients)*, are implemented to facilitate this transformation.[17]
5) *Recognition of Extracted Information*: In this phase, the concept of pattern matching is put into practice. The recognition process involves taking the extracted information and juxtaposing it with predetermined data. Pattern matching techniques are utilized to execute this comparison and alignment. Among the widely acclaimed software for this purpose is the Google Speech API, which serves as one of the utmost favored tools in this domain.[17]

VI. CODE

Here is one algorithm and code which ismeant for disease detection with the help of speech-based diagnostic system using machine learning.

A. Code to detet COVID-19.[25]

Python code for a speech-based diagnostic system using machine learning

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
dataset = pd.read_csv('speech_dataset.csv')
X = dataset.drop('COVID-19', axis=1)
y = dataset['COVID-19']
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
classifier = RandomForestClassifier(n_estimators=100,
random_state=42)
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy: {:.2f}%".format(accuracy * 100))
```

Explanation and output of the above given code

Sure, here's the revised paragraph using synonyms while keeping the meaning intact:

1) The necessary modules are imported: numpy, pandas, train_test_splitfromsklearn.model_selection, Random Forest Classifier fromsklearn. ensemble, and accuracy_score from sklearn.metrics.
2) The dataset is loaded using pandas' read_csv function. It comprises speech samples and a binary label column named 'COVID-19'.

3) The dataset is split into features (X) and the target label (y).
4) The dataset is further divided into training and testing sets using train_test_split from sklearn.model_selection. The testing set size is set to 20% of the total data, and a random state of 42 is employed to ensure reproducibility.
5) A *Random Forest (RF*) classifier is instantiated with 100 estimators (decision trees) and a random state of 42.
6) The classifier is meant for the training set using the fit method.
7) The classifier then used to make predictions on the test set (X_test).
8) The precision of the classifier is computed by comparing the predicted labels (y_pred) with the actual labels from the test set (y_test).
9) Finally, the accuracy score is displayed on the screen.

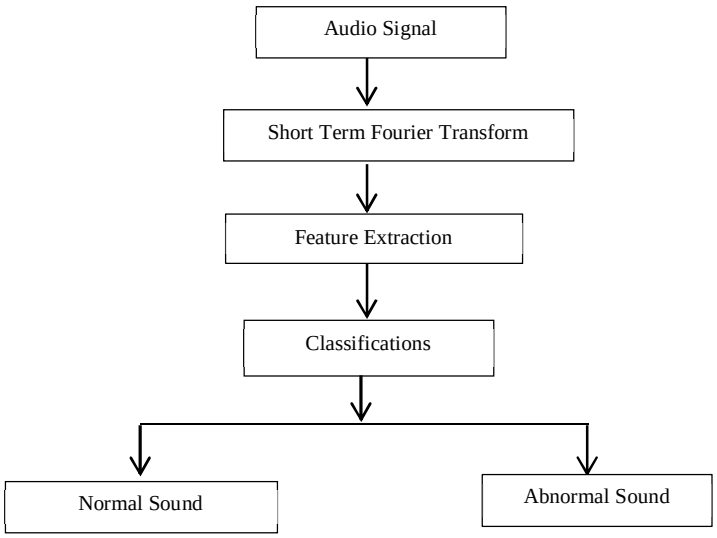
The result of this code will be a single line exhibiting the classifier's accuracy on the test set in percentage format, rounded to two decimal places. The value will vary each time the code is executed due to the randomness introduced during data splitting and classifier training. The result might resemble the following:

Accuracy: 86.25%

This signifies that the Random Forest classifier attained an accuracy of about 86.25% on the test set, indicating its capability to predict COVID-19 based on speech samples.

Similarly, we can utilise the same code for detection of the other respiratory diseases.

VII. PROPOSED ALGORITHM



A. Audio Signal

We are graphing the time-domain waveforms of typical sound, atypical sound, whooping cough, and bronchitis. The time-domain signals for all four state are given below.

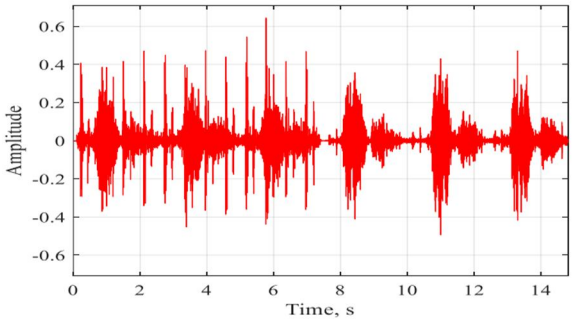


Figure 5 | Normal Sound time domain signal

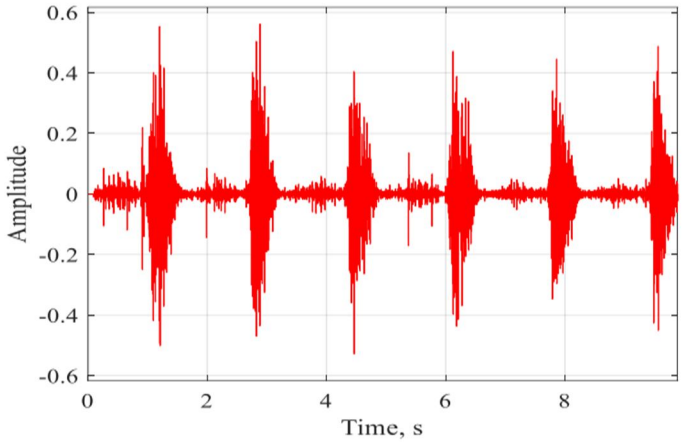


FIGURE 6 | Abnormal Sound time domain signal

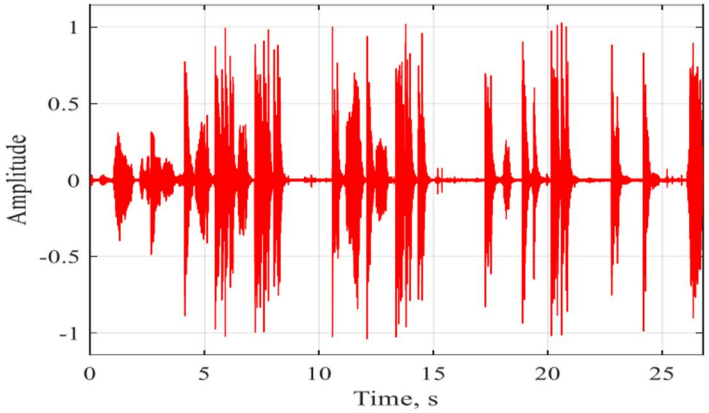


Figure 7 | Pertussis temporal signal

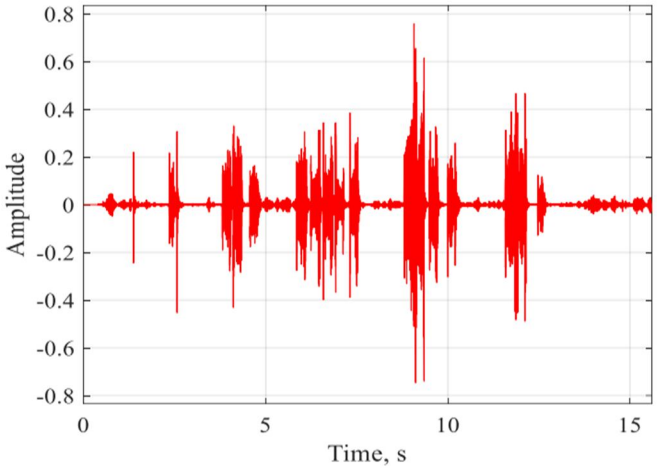


Figure8 | Bronchitis temporal signal

B. Short time Fourier Transform

Following the depiction of the signals of time domain, we are now visualizing their Short-time Fourier Transforms. The spectrograms for all four state are presented below:

Spectrogram of the signal

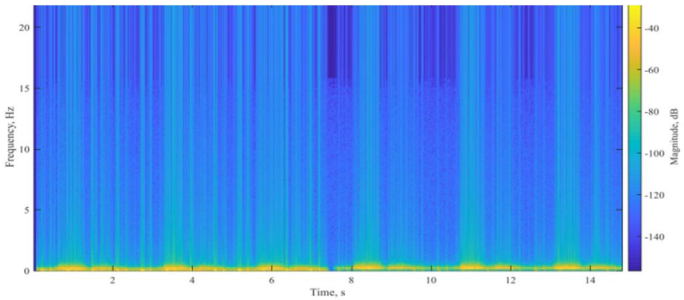


FIGURE 9 | Spectrogram of Normal sound

Spectrogram of the signal

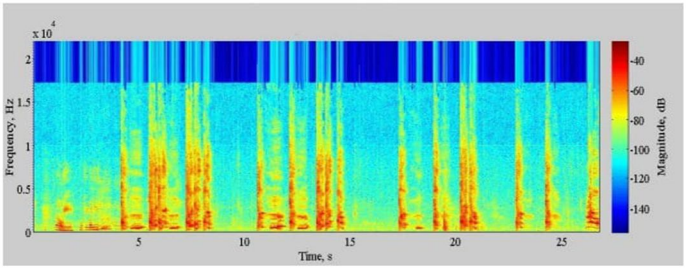


FIGURE 10 | Spectrogram of Abnormal sound

Spectogram of the signal

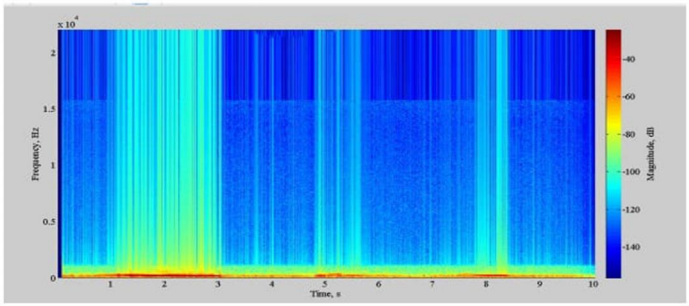


Figure 11 | Spectrogram of Pertussis

Spectogram of the signal

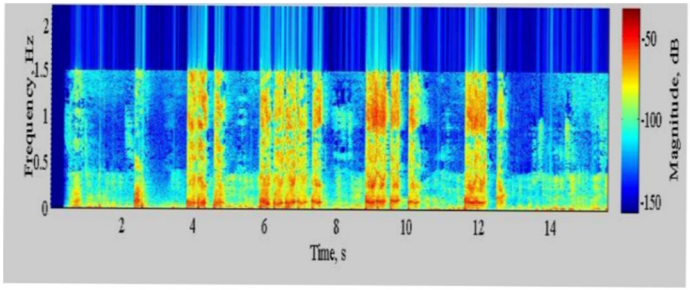


Figure 12 | Spectrogram ofBronchitis

C. Feature Extraction

After plotting its spectrogram, we are extracting two features mainly: Concentration Measure and Dominant Frequency. All the four dominant frequency plots are listed below:

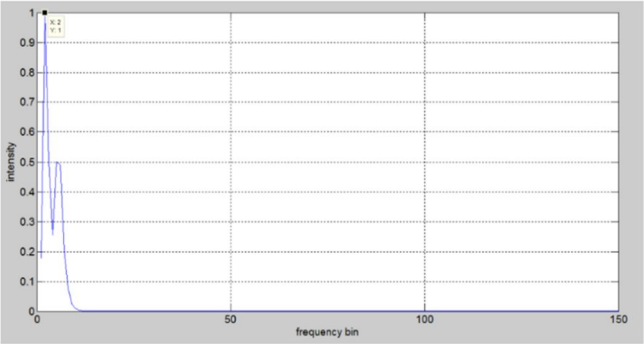


Figure 13 | Dominant Frequency plot for Normal sound

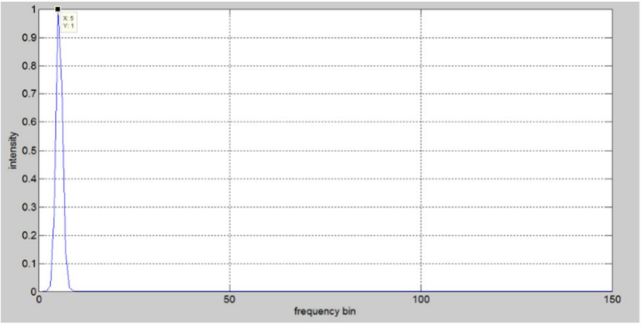


Figure 14 | Dominant Frequency plot for abnormal sound

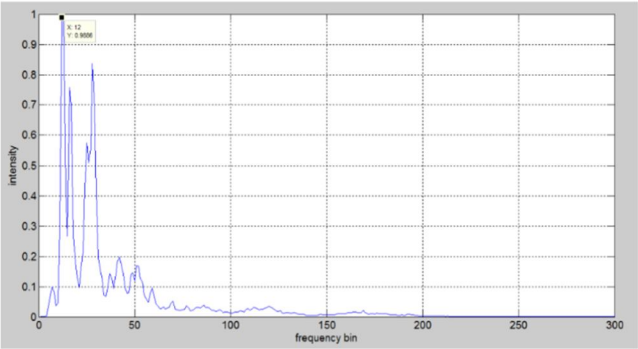


Figure 15 | Dominant Frequency plot for Pertussis

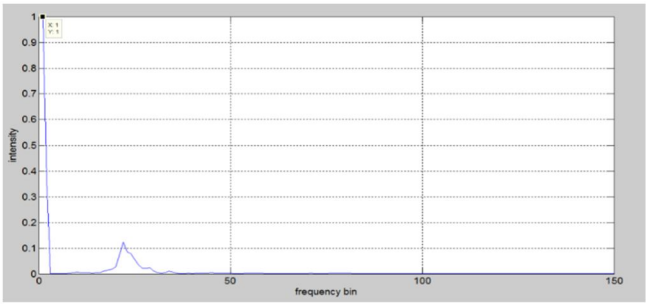


Figure 16 | Dominant Frequency plot for Bronchitis

TABLE I. CLASSIFICATION BETWEEN NORMAL AND ABNORMAL SOUND

	Classification between Normal and Abnormal Sound		
	Concentration measure	*Dominant Frequency (Hz)*	*Class*
Sample 1	0.0121	86.1238	One
Sample 2	0.011	81.3	One
Sample 3	0.012	89.78	One
Sample 4	0.0107	94.78	One
Sample 5	0.0117	80.245	One
Sample 6	0.0113	91.67	One
Sample 7	0.013	100.34	One
Sample 8	0.012	80.9	One
Sample 9	0.0129	103.678	One
Sample 10	0.0115	78.1238	One
Sample 11	0.0119	110.4	One
Sample 12	0.0128	88.657	One
Sample 13	0.0138	93.65	One
Sample 14	0.0109	115.67	One
Sample 15	0.0112	119.345	One
Sample 16	0.0114	99.56	One
Sample 17	0.0116	79.56	One
Sample 18	0.0138	85.123	One
Sample 19	0.0155	11.123	One
Sample 20	0.0133	93.56	One
Sample 21	0.011	87.345	One
Sample 1	0.0107	394.5315	Two
Sample 2	0.011	233.45	Two
Sample 3	0.0107	86.1328	Two
Sample 4	0.0109	245.788	Two
Sample 5	0.017	215.332	Two
Sample 6	0.02	199.78	Two
Sample 7	0.0137	129.112	Two
Sample 8	0.028	301.34	Two
Sample 9	0.0204	215.332	Two
Sample 10	0.03	99.345	Two
Sample 11	0.054	119.5	Two
Sample 12	0.09	186.56	Two
Sample 13	0.07	213.65	Two
Sample 14	0.012	254.765	Two

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

	Classification between Normal and Abnormal Sound		
	Concentration measure	*Dominant Frequency (Hz)*	*Class*
Sample 15	0.043	184.65	Two
Sample 16	0.0543	158.21	Two
Sample 17	0.0204	209.45	Two
Sample 18	0.0175	313.654	Two
Sample 19	0.045	123.55	Two
Sample 20	0.034	300.98	Two
Sample 21	0.0105	287.54	Two
		Class 1 --- Normal Sound	
		Class 2 --- Abnormal Sound	

TABLE II. CLASSIFICATION BETWEEN PERTUSSIS AND BRONCHITIS

	Classification between Normal and Abnormal Sound		
	Concentration measure	***Dominant Frequency (Hz)***	***Class***
Sample 1	0.0035	559.868	One
Sample 2	0.0039	560.954	One
Sample 3	0.0047	775.1953	One
Sample 4	0.004	498.54	One
Sample 5	0.0099	775.195	One
Sample 6	0.01	654.67	One
Sample 7	0.0041	645.9961	One
Sample 8	0.005	543.234	One
Sample 9	0.0053	538.33	One
Sample 10	0.0034	654.32	One
Sample 11	0.0065	698.634	One
Sample 12	0.0078	733.54	One
Sample 13	0.0023	754.343	One
Sample 14	0.0078	799.459	One
Sample 15	0.0036	589.55	One
Sample 16	0.0033	800.455	One
Sample 17	0.0087	899.54	One
Sample 18	0.0022	703.54	One
Sample 19	0.0033	1003.2	One
Sample 20	0.0041	698.432	One
Sample 21	0.0056	87.345	One
Sample 1	0.0055	775.1953	Two

©IJRASET: All Rights are Reserved

	Classification between Normal and Abnormal Sound		
	Concentration measure	***Dominant Frequency (Hz)***	***Class***
Sample 2	0.0039	789.432	Two
Sample 3	0.0064	1076.78	Two
Sample 4	0.0043	957.43	Two
Sample 5	0.0039	904.3945	Two
Sample 6	0.0055	987.432	Two
Sample 7	0.0047	947.46	Two
Sample 8	0.0087	1044.43	Two
Sample 9	0.0041	990.5273	Two
Sample 10	0.0068	967.54	Two
Sample 11	0.0076	1100.543	Two
Sample 12	0.0033	899.65	Two
Sample 13	0.0087	940.524	Two
Sample 14	0.0099	1205.43	Two
Sample 15	0.0043	1299.543	Two
Sample 16	0.0065	999.43	Two
Sample 17	0.0078	1009.496	Two
Sample 18	0.0089	785.74	Two
Sample 19	0.0021	846.654	Two
Sample 20	0.0029	879.43	Two
Sample 21	0.0012	1309.54	Two
		Class 1 --- Pertussis	
		Class 2 --- Bronchitis	

D. Classification Results

We have divided our information using an Artificial Neural Network, and we are providing the confusion matrix of the classified data. A confusion matrix is a tabular representation used to calculate the potency of a classification model applied to a test dataset with known true values, i.e., in the field of Supervised Learning.



Figure 17 | Confussion matrix for normal and abnormal sound.

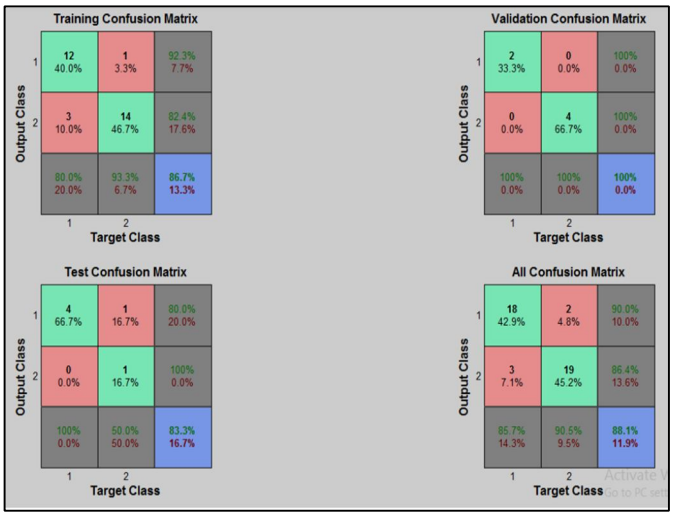


Figure 18 | Confussion matrix for Pertussis and Bronchitis.

VIII. PERFORMANCE ANALYSIS

1) *Sensitivity*, also known as recall or true positive rate, denotes the proportion of accurately classified instances. It is calculated as TP/(TP+FN), where TP represents true positives and FN stands for false negatives.
2) *Specificity*, also referred to as the true negative rate, signifies the percentage of false instances that are accurately rejected. It is computed as TN/(TN+FP), where TN represents true negatives and FP stands for false positives.
3) *Accuracy* reveals the frequency where the outcomes are correctly classified. It is proposed as (TP + TN) / (TP + TN + FP + FN), where TP represents true positives, TN stands for true negatives, FP denotes false positives, and FN represents false negatives.
4) *Positive Predictive Value (PPV)*, also known as precision, indicates the ratio of positive outcomes that are accurately identified. It is computed as TP / (TP + FP), where TP represents true positives, and FP stands for false positives.
5) *Negative Predictive Value (NPV)* illustrates the proportion of negative outcomes that are correctly declined. NPV is calculated as TN / (TN + FN), where TN represents true negatives, and FN stands for false negatives. In this context, TP refers to true positives, FP denotes false positives, and FN represents false negatives.

TABLE III. RESULT FOR NORMAL AND ABNORMAL SOUND

Normal & Abnormal	Results for Normal and Abnormal sound				
	Sensitivity	*Specificity*	*Accuracy*	*PPV*	*NPV*
	91.3%	8.7%	95.2 %	100%	90.5 %

.

TABLE IV. RESULT FOR PERTUSSIS AND BRONCHITIS SOUND

Pertussis & Bronchitis	Results for Normal and Abnormal sound				
	Sensitivity	*Specificity*	*Accuracy*	*PPV*	*NPV*
	90%	10%	88.1 %	85.7 %	90.5 %

IX. DRAWBACKS OF SPEECH PROCESSING USING AI AND ML

1) *Insufficient Data*: The availability of a limited dataset for training AI models to recognize respiratory illnesses from speech poses a challenge. This is primarily due to the elusive nature of diagnosing respiratory diseases, as patients often delay seeking medical attention until their symptoms become severe. Consequently, there is an inadequate amount of data to train AI models effectively, leading to difficulties in accurately detecting all types of respiratory diseases. [18,19,20]
2) *Variability in Cough Sounds*: The characteristics of a cough can changebased on the specific respiratory condition, the severity of the ailment, and the individual's unique vocal cords. These variations create complexities for AI models in precisely identifying the type of respiratory disease from the sound of a cough. [18,19,20]
3) *Interference from Backgroundnoise*: The presence of background noise can also hinder the accuracy of AI models designed to detect respiratory diseases from speech. Such noise can mask the distinct cough sounds, making it challenging for the AI model to differentiate and accurately identify the cough amidst the interference.[18,19,20]
4) *Accuracy Challenges*: Despite advancements, AI models still struggle with achieving high accuracy in detecting respiratory diseases from speech. This limitation arises from the AI's current inability to comprehensively grasp all the subtle nuances in cough sounds necessary for precise respiratory disease diagnosis.[18,19,20]

X. IDEAS TO OVERCOME THE DRAWBACKS

1) *Expand the Data Pool:* To label the scarcity of data, a potential solution is to broaden the available dataset utilized for schooling AI models. This can be achieved by gathering additional cough sounds from diverse patients afflicted with various respiratory illnesses. The data collection process could employ diverse methods, including crowdsourcing or the utilization of wearable devices capable of capturing cough sounds. [18,21,22,23]
2) *Enhance Algorithm Robustness:* An alternative strategy to tackle the challenges in speech processing is to advance the robustness of algorithms. These improved algorithms would be better equipped to handle the variations present in cough sounds and mitigate the impact of background noise. A powerful advancement to achieve this involves leveraging deep learning techniques, which have demonstrated proficiency in accurately classifying cough sounds. [18,21,22,23]
3) *Utilize Multimodal Data*: Another avenue to explore is the integration of multimodal data, wherein different types of information, such as cough sounds, images, and clinical data, are combined. This synergistic approach can bolster the precision of AI models by providing comprehensive insights into the patient's condition. [18,21,22,23]

XI. BENEFITS OF SPEECH BASED DISEASE DETECTION

1) *Non-intrusive and Economical Screening*: Utilizing AI and ML for speech processing offers a non-intrusive and cost-effective means of screening respiratory diseases. It obviates the need for physical contact with patients or the usage of expensive medical equipment, thus ensuring accessibility to a wider population.[21,24,25]
2) *Timelier identification:* The application of speech processing could also enable earlier identification of respiratory diseases. This is because alterations in respiratory sounds indicative of an ailment can be discerned before other symptoms, such as breathlessness, manifest. The significance of early diagnosis lies in its potential to facilitate prompt treatment, thereby enhancing the likelihood of positive outcomes.[21,24,25]
3) *Objective Evaluation:* Traditional methods of respiratory disease assessment often rely on subjective measures and interpretation. In contrast, AI and ML offer an objective and standardized approach to analysing speech data, minimizing inter-observer variability and enhancing diagnostic accuracy.[21,24,25]
4) *Early Alarm Systems:* By incorporating speech processing into wearable devices or smartphone applications, early warning systems can be developed. These systems promptly alert users to potential respiratory issues, prompting them to seek medical attention as needed by the help of AI. [21,24,25]

XII. RECENT WORKS ON SPEECH-BASED DISEASE DETECTION STUDIES

TABLE V. RECENT WORK ON SPEECH DISEASE DETECTION

Disease	RESULT WORKS ON SPEECH BASED DISEASE DETECTION STUDIES						
	Reference	*Unit*	*Data*	*Characteristics*	*ML Work*	*Method*	*Perform*
Bronchial Asthma	[26]	71 (N/A) , w/ and 135 (NA) w/o bronchial asthma, 8 ± N/A y	Sustained vowel /a:/	MFCCs, CQCCs	w/ vs w/o (2-class)	GMM-UBM	Acc: 0.72
COVID-19	[27]	52 (32/20), 63.4 ± 9.9 y Hospitalized w/, 3 Categories	Speech, 5 sentences	eGeMAPS, ComParE	severity prediction (3-class)	SVM	UAR ≤ 0.68
	[28]	20 (12/8) w/ 60 (40/20) w/o and healthy	Speech	(1/12)-MFCCs, LLDs	w/ vs w/o (2-class)	LSTM-RNN	Acc: 0.88
Pathological Speech	[29]	126 (N/A)	Speech	Cochleogram, Hilbert Spectrum	w/ vs w/o (2-class)	VGG-16 CNN	Acc: 0.92
Upper Respiratory Tract Infection	[30]	630 (382/248), 12–84 y (29.5 ± 12.1 y), w/ and w/o, WURSS-24 (German version)	Spontaneous speech, text reading	ComParE	w/ vs w/o (2-class)	DNNs	UAR: 0.67

Keywords – *N/A* (Not Available), *w/o* (without [corresponding disease], *Acc.* (accuracy), *MFCC* (Mel Frequency Cepstral Coefficient), *CQCC* (constant-Q cepstral coefficients), *GMM-UBM* (Gaussian mixture model-universal background mode), *eGeMAPS* (extended Geneva minimalistic acoustic parameter set), *ComParE* (computational paralinguistic challenge [representations]), *SVM* (support vector machine), *w/* (with [corresponding disease]), *CNN* (Convolutional neural network),*y* (years), *DNN*(Deep neural network), *UAR*(Unweight average recall), *WURSS*(Wisconsin upper respiratory symptom survey).

XIII. FUTURE SCOPES IN SPEECH BASED DISEASE DETECTION.

1) *Enhanced Precision and Responsiveness:* On-going advancements in speech processing algorithms have the potential to result in superior precision and sensitivity when detecting respiratory illnesses. Notably, the utilization of deep learning algorithms has already demonstrated their effectiveness in categorizing respiratory sounds, and as these algorithms continue to progress, their accuracy could further improve.[31,32,33]
2) *Tailored Medical approaches*: Additionally, speech processing can play a role in tailoring medicine for respiratory ailments. By analyzing the unique characteristics of a patient's respiratory sounds, healthcare providers can identify the most suitable treatment for the individual's specific requirements. [31,32,33]

3) *Remote Monitoring and Telemedicine*: The integration of speech-based respiratory disease detection into telemedicine platforms facilitates remote monitoring of patients. This proves especially advantageous for individuals residing in remote areas or with limited access to healthcare facilities.[31,32,33]

XIV. CONCLUSION

Research in AI-based disease recognition using speech has shown significant promise, with the potential to revolutionize healthcare. Advanced machine learning algorithms and natural language processing has demonstrated its effectiveness in early disease detection and diagnosis through speech patterns analysis. The findings demonstrate impressive accuracy rates in identifying various diseases, providing a non-invasive and cost-effective screening method. AI's ability to analyze vast amounts of data quickly helps healthcare professionals make informed decisions, leading to improved patient outcomes and potentially saving lives.

However, it's essential to acknowledge some limitations and challenges. Access to diverse and large-scale datasets is critical, especially for recognizing rarer diseases. Additionally, ethical considerations, privacy concerns, and potential biases in data and algorithms require attention to ensure responsible and equitable AI deployment in healthcare. To optimize and validate AI speech-based disease recognition, collaboration between AI experts, medical professionals, and ethicists is necessary. Conquering these obstacleswill make AI an indispensable device for early disease diagnosing and management, enhancing healthcare delivery worldwide. As technology evolves, a proactive and human-centric approach is vital, ensuring AI complements healthcare professionals' expertise while prioritizing patient well-being and ethics.

XV. ACKNOWLEDGEMENT

The completion of thispaper would not have been achievable without the contributions and support of many individuals and organizations. We are deeply grateful to all those who assumed a part in the success of this paper.

First and foremost, we would like to express our gratitude to our mentor, teacher and advisor, (Prof.) Dr. Soumik Poddar for his, idea, guidance, support, and encouragement throughout the entire process. His mentorship and expertise were invaluable in helping us to shape the direction of our research and to bring our ideas to fruition.

In addition, we desire to extend our sincere thanks to all of the members in our study, who liberally shared their time, experiences, and insights with us. Firstly, Saptarshi Pramanik for helping us with his immense knowledge in the field of AI based Speech Recognition and the methodologies, techniques and working of AI based Speech Recognition used in. Secondly, Sambit Chanda for helping us with the coding and algorithm of this paper. Thirdly, Arijit Chakraborty, for editing this paper.

Overall, this paper would not have succeeded without the support and contributions of so many people. Their willingness to engage with our work was essential for the success of this paper, and we are deeply obliged for their participation.

REFFERENCES

For completing this paper, we took references from the websites and research papers listed below :

[1] R. Pramono, S. Imtiaz and E. Rodriguez-Villegas, "A Cough-Based Algorithm for Automatic Diagnosis of Pertussis", PLOS ONE, vol. 11, no. 9, p. e0162128, 2016.

[2] R. Sharan, U. Abeyratne, V. Swarnkar and P. Porter, "Automatic Croup Diagnosis Using Cough Sound Recognition", IEEE Transactions on Biomedical Engineering, vol. 66, no. 2, pp. 485-495,2019.
Available: 10.1109/tbme.2018.2849502.

[3] BS. Lin, BS. Lin, "Automatic wheezing detection using speech recognition technique", Journal of Medical and Biological Engineering, vol. 36, no. 4, pp. 545-554, Aug. 2016.

[4] https://www.researchgate.net/figure/Block-diagram-of-Mel-Frequency-Cepstral-Coefficients-MFCCs-extraction_fig1_280027126

[5] Eyben F, Scherer KR, Schuller BW, Sundberg J, André E, Busso C, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Trans Affect Comput. (2015) 7:190–202. doi: 10.1109/TAFFC.2015.2457417

[6] Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, et al. The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Proceedings INTERSPEECH. Lyon: ISCA (2013). p. 148–52

[7] Amiriparian S, Gerczuk M, Ottl S, Cummins N, Freitag M, Pugachevskiy S, et al. Snore sound classification using image-based deep spectrum features. In: Proceedings INTERSPEECH. Stockholm: ISCA (2017). p. 3512–6.

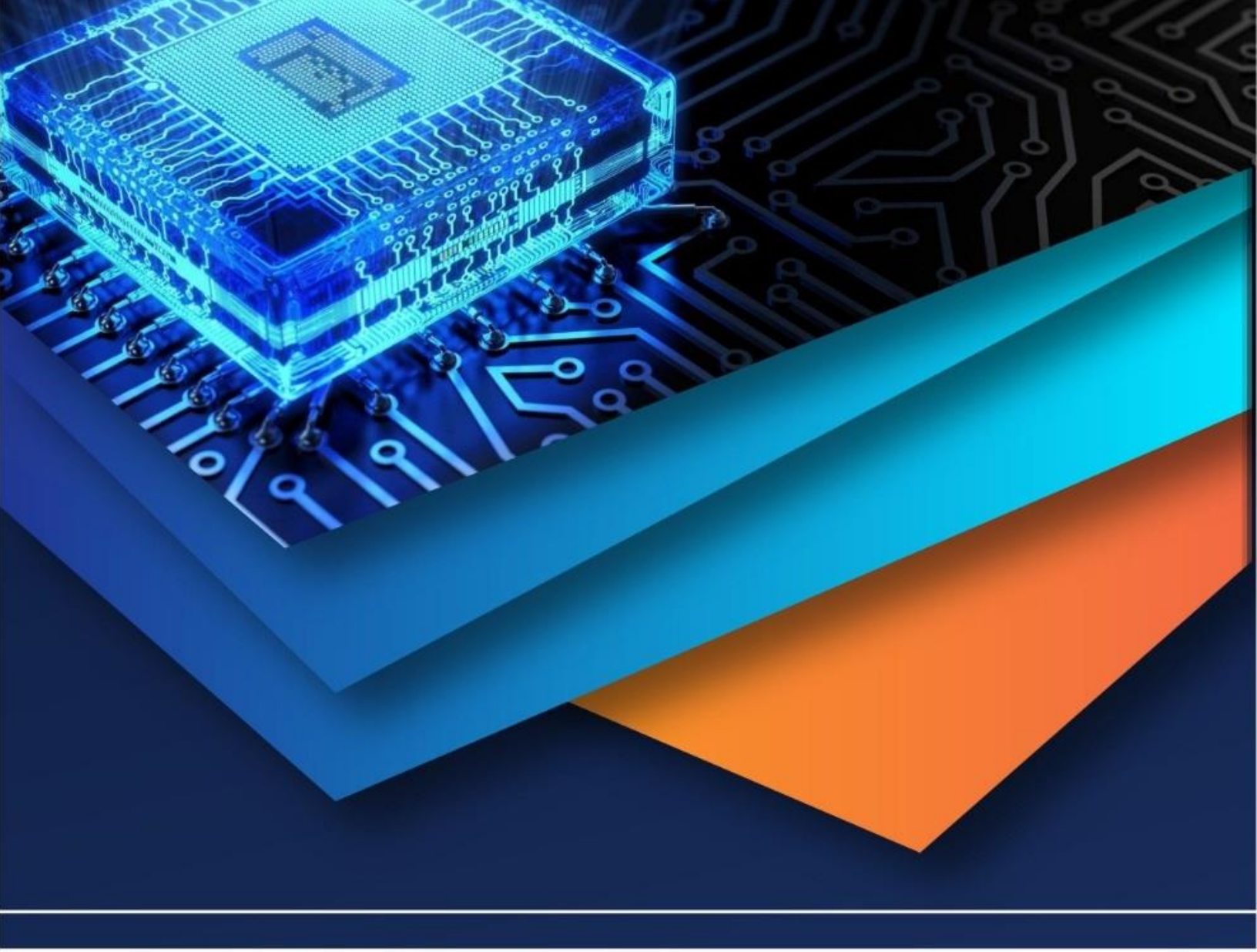
[8] Schmitt M, Janott C, Pandit V, Qian K, Heiser C, Hemmert W, et al. A bagof-audio-words approach for snore sounds' excitation localisation. In: ITGSymposium on Speech Communication. (2016).

[9] Is Speech the New Blood? Recent Progress in AI-Based Disease Detection from Audio in a Nutshell. Frontiers in Digital Health, 2022.https://www.frontiersin.org/articles/10.3389/fdgth.2022.886615/full

[10] Automatic Speech-Based Disease Detection: A Review of Recent Advances. Journal of Medical Speech-Language Pathology, 2022. https://pubmed.ncbi.nlm.nih.gov/35640027/

[11] Artificial Intelligence-Based Speech Analysis System for Medical Support. PLOS ONE, 2021. https://pubmed.ncbi.nlm.nih.gov/34649585/

[12] Bartl-Pokorny KD, Pokorny FB, Batliner A, Amiriparian S, Semertzidou A, Eyben F, et al. The voice of COVID-19: acoustic correlates of infection in sustained vowels. J Acoust Soc Am. (2021) 149:4377–83.doi: 10.1121/10.0005194
[13] Hecker P, Pokorny FB, Bartl-Pokorny KD, Reichel U, Ren Z, HantkeS,et al. Speaking Corona? Human and machine recognition of COVID-19 from voice. In: Proceedings INTERSPEECH. Brno, Czech Republic: ISCA (2021). p.701–5.
[14] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press. (2016). Available online at: http://www.deeplearningbook.org.
[15] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, editors. Advances in Neural Information Processing Systems. Vol. 27. Curran Associates, Inc. (2014). https://proceedings.neurips.cc/paper/2014/file/ 5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.
[16] Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: An ASR corpus based on public domain audio books. In: Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing. South Brisbane, QLD: IEEE (2015). p. 5206–10.
[17] https://www.scaler.com/topics/speech-recognition-in-ai/
[18] Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques: Challenges and Opportunities: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8545201/
[19] Speech Processing Challenges and How to Overcome: https://www.speechmatics.com/company/articles-and-news/speech-recognition-challenges-overcome
[20] Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis: https://www.sciencedirect.com/science/article/pii/S221321982100194X
[21] Towards using cough for respiratory disease diagnosis by leveraging Artificial Intelligence: A survey: https://www.sciencedirect.com/science/article/pii/S235291482100294X
[22] Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7959628/
[23] A Machine Learning Model for Detecting Respiratory Problems using Voice Recognition: https://ieeexplore.ieee.org/document/9033920
[24] Artificial Intelligence and Machine Learning for Speech Processing in Respiratory Diseases: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8241767/
[25] Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9791302/
[26] Balamurali BT, Hee HI, Teoh OH, Lee KP, Kapoor S, Herremans D, et al. Asthmatic versus healthy child classification based on cough and vocalised /a:/ sounds. J Acoust Soc Am. (2020) 148:EL253–9. doi: 10.1121/10.0001933
[27] Han J, Qian K, Song M, Yang Z, Ren Z, Liu S, et al. An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety. arXiv. (2020). doi: 10.48550/arXiv.2005.00096
[28] Hassan A, Shahin I, Alsabek MB. COVID-19 detection system using recurrent neural networks. In: Proceedings IEEE International Conference on Communications, Computing, Cybersecurity, and Informatics. virtual: IEEE (2020).
[29] Gumelar AB, Yuniarno EM, Anggraeni W, Sugiarto I, Mahindara VR, Purnomo MH. Enhancing detection of pathological voice disorder based on deep VGG-16 CNN. In: Proceedings International Conference on Biomedical Engineering. virtual: IEEE (2020). p. 28–33.
[30] Albes M, Ren Z, Schuller BW, Cummins N. Squeeze for sneeze: compact neural networks for cold and flu recognition. In: Proceedings INTERSPEECH. Shanghai: ISCA (2020). p. 4546–50.
[31] A review on lung disease recognition by acoustic signal analysis with deep learning networks: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00762-z
[32] Research Progress of Respiratory Disease and Idiopathic Pulmonary Fibrosis Based on Artificial Intelligence:https://www.mdpi.com/2075-4418/13/3/357
[33] Recent Developments, Challenges, and Future Scope of Voice Activity Detection Schemes—A Review: https://www.researchgate.net/publication/353005634_Recent_Developments_Challenges_and_Future_Scope_of_Voice_Activity_Detection_Schemes-A_Review



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)