**COVID-19 detection using Cough Recordings**

Project-II (18ME10024) report submitted to

Indian Institute of Technology, Kharagpur

In partial fulfilment for the award of the degree of

Bachelor of Technology (Hons.)

In

Mechanical Engineering

**By**

**Himanshu**

**(**18ME10024**)**

**Under the supervision of**

**Prof. Ram Babu Roy**



**Department of Mechanical Engineering**

**Indian Institute of Technology, Kharagpur**

**Autumn Semester,**

**November 21, 2021**

**DECLARATION**

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources whenever necessary.

Date: Himanshu

Place: IIT Kharagpur (18ME10024)



**DEPARTMENT OF MECHANICAL ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR**

**KHARAGPUR-721302, INDIA**

***CERTIFICATE***

This is to certify that the project report entitled **“COVID-19 Detection using Cough Recordings”** submitted by Himanshu (Roll No. 18ME10024) to the Indian Institute of Technology, Kharagpur towards the fulfilment of requirements for the award of the degree of Bachelor of Technology (Hons.) in Mechanical Engineering is a record of bonafide work carried out by him/her under my/our supervision and guidance during Autumn Semester 2021-22.

**Prof. Ram Babu Roy**

**Rajendra Mishra School of Engineering Entrepreneurship**

**Indian Institute of Technology**

**Kharagpur, India**

**Date –**

**Place- IIT Kharagpur**

Table of Contents

[A. Creating an ML Model for Detection of COVID-19 using Cough Recordings **Error! Bookmark not defined.**](#_Toc87793299)

[1. Abstract 5](#_Toc87793300)

[2. Introduction 5](#_Toc87793301)

[3. Problem Definition 5](#_Toc87793302)

[4. Literature Review 6](#_Toc87793303)

[5. Methodology 7](#_Toc87793304)

[ Audio Preprocessing 7](#_Toc87793305)

[ Feature Extraction 7](#_Toc87793306)

[ Classifier Architectures 10](#_Toc87793307)

[ Results 12](#_Toc87793308)

[6. Conclusion and Discussion: 14](#_Toc87793309)

[B. Creation of a Web Application for implementing the Model **Error! Bookmark not defined.**](#_Toc87793310)

[1. Front - End App Workflow 14](#_Toc87793311)

[ App Overview 14](#_Toc87793312)

[ User Input 15](#_Toc87793313)

[ Classification of the Recording 16](#_Toc87793314)

[2. Back–End App Workflow 16](#_Toc87793315)

[C. Comparison of Different waves during COVID-19 pandemic 17](#_Toc87793316)

[Analysis Method 17](#_Toc87793317)

[Observation 18](#_Toc87793318)

[Results 19](#_Toc87793319)

[D. References 19](#_Toc87793320)

# COVID-19 Cough Pattern Analysis

## Abstract

COVID-19  was discovered on January 7, 2020, in Wuhan, China; since then, much technological advancement is already in place to identify COVID-19 patients. We present a cough pattern analysis method that would help detect patterns in COVID-19 cough sounds by comparison of Cough Patterns in COVID and Non-COVID coughs.

## Introduction

COVID-19 was declared as an international pandemic on March 11, 2020, by World Health Organization (WHO). COVID-19 is a result of SARS-CoV-2(Severe Acute Respiratory Syndrome Corona-Virus). It begins by contaminating the mucous films within the throat and developments down the breath tract driving to the lungs, with coughing being a common symptom [1]. Studies have reported that Cough sound contains underutilised respiratory health info, which can be used for analysis. Until November 16 2021, more than 254 million cases confirmed COVID-19 infections worldwide and close to 5.1 million people have perished after getting infected by the virus [2].

Methods like X-rays and Chest CT scans have been used to identify COVID-19 patients [4]. Furthermore, these methods suggest that the respiratory system is influenced distinctively by COVID-19 [5]. Hence, cough sounds carry essential information related to the respiratory system [6].

Artificial Intelligence has many audio processing and analysis applications that can help medical professionals screen and detect infected people, which can help reduce the spread of infection. Usage of cough processing techniques might provide us with vital diagnostic leads. These techniques can be built using the existing feature extraction techniques. Machine Learning algorithms have the potential to identify COVID-19 patients by observing the *coughing patterns*. This report develops one such feature extraction technique that can extract vital information from cough sounds.

The development of such feature extraction and classification models can help lessen the effort of COVID-19 testing laboratories and also help increase the accessibility of the tests to a broader community. Using these models, test Results can be quickly obtained, thus reducing the intermediate spread of the disease. This report then presents the classification techniques used to separate COVID-19 cough from a non-COVI-19 cough using Data Analysis.

## Literature Review

This review attempts to summarise the vital studies in COVID-19 detection and identify diseases based on cough audio samples’ features like frequency, duration, and intensity. A literature review is provided, keeping in mind the current technological progressions in this Area of research. In recent years, numerous studies have suggested acoustic features to identify respiratory diseases in cough signals. Furthermore, conditions associated with the respiratory system can be identified with the help of machine learning techniques. These algorithms can also process respiratory data and coughs to diagnose COVID-19 [8][11].

Abeyratne et al. [9] inspected the modifications in coughing patterns of asthma, bronchitis coughs and pneumonia. This paper extracted features from the cough sound such as Mel Frequency Cepstrum Coefficients, Formant frequencies and Zero Crossing Rate. The Logistic Regression Model trained on recordings from 91 subjects showed a specificity of 73% and sensitivity of 80% on the validation set. This study provided vital evidence supporting the hypothesis that the information related to the lower respiratory tract is carried by the cough sound.

In the system proposed by Belkacem et al. [10], the cough recording is initially passed through a cough detection system, a source separation system and then features such as Mel-frequency Cepstral Coefficients, Shannon Entropy (SH) and Zero Crossing Rate are extracted from the recording which is then passed through Deep Neural Network (DNN). A logarithmic scaled Mel-spectrogram has also been used for feature extraction, passed through a fully convolutional network (FCN). The authors also designed a pipeline in MATLAB to analyse the cough recordings, which consists of computing Fourier Transform (FFT).

Chatrzarrin et al. [12] have listed preprocessing and feature extraction techniques obtained from cough signals to classify them into dry and wet coughs. Vikrant Et al. [13] classified the cough sounds using a Support Vector Classifier(SVC). The classifier performed well with an accuracy of 98.9% and the True Positive Rate (TPR) close to 96%. Matos Et al. [14] used hidden Markov models (HMMs) to detect cough sounds with the extraction of features like linear predictive coding (LPC) coefficients and Mel frequency cepstral coefficients (MFCC). The algorithm was accurately able to recognise 82% of the events of coughing.

Imran et al. [16] have made an app to classify COVID-19 cough from an audio recording close to 90% precision. They used a model which was initially trained on everyday noises and then tweaked the model on the COVID-19 dataset. Their app uses a mediator to merge a Neural Network (CNN) that uses a Mel spectrogram as input and a Machine Learning model that uses Mel Frequency Cepstrum coefficients (MFCC) as input to generate predictions.

Sharma et al. [17] have compiled a database of respiratory sounds like coughing, breathing, and voice known as Coswara for COVID-19 Detection. The Random Forest (RF) model trained on Coswara data gave a mean accuracy of 70% for coughing. Wang et al. [18] compared the similarity of Cepstral Coefficients to perform Voice Activity Detection (VAD). The usage of MFCC has proven to be the most suitable method for Voice Activity Detection (VAD) in a noisy background compared to other features.

## Methodology

### Data Collection and EDA

A Dataset consisting of 2665 cough audio recordings has been used for the analysis, consisting of 681 COVID-19 positive patients’ and 1984 non-COVID-19 patients’ recordings. These recordings were obtained from the Indian Institute of Science, Bangalore [17], containing 2515 records from Indian citizens and 150 from other countries.

### Audio Preprocessing

Cough signals consist of vital information. The capturing of such time-varying characteristics would help us distinguish between a variety of coughs. For extracting these features, the cough recordings are in **.wav** format, and these waves are digitalised. These are converted into a one-dimensional array of digital values using a sampling technique. This conversion is done using Analog Digital Conversion, which consists of sampling, quantisation, and encoding. These digital values represent the amplitude, frequency at that given instance. Here the sampling rate is fixed to 22050 for all the recordings.

### Resolving Data Imbalance

As the dataset was imbalanced, the Synthetic Minority Oversampling Technique (SMOTE)[41] was used for balancing the data. This technique has been established successfully in the field of cough detection and classification [20].

### Feature Extraction

Cough signals consist of vital information. The capturing of such time-varying characteristics would help us distinguish between a variety of coughs. For extracting these features, the cough recordings are in **.wav** format, and these waves are digitalised. These are converted into a one-dimensional array of digital values using a sampling technique. This conversion is done using Analog Digital Conversion, which consists of sampling, quantisation, and encoding. These digital values represent the amplitude, frequency at that given instance. Here the sampling rate is fixed to 22050 for all the recordings.

As the dataset was imbalanced, the Synthetic Minority Oversampling Technique (SMOTE)[41] was used for balancing the data. This technique has been established successfully in the field of cough detection and classification [20].

### Feature Extraction

A total of 25 numerical-frequency features were extracted from each cough recording which was done to extract the frequencies at which the disturbances in the cough sound were amplified. These features were extracted and used after a careful literature review on cough identification, detection and classification [35][36][37].

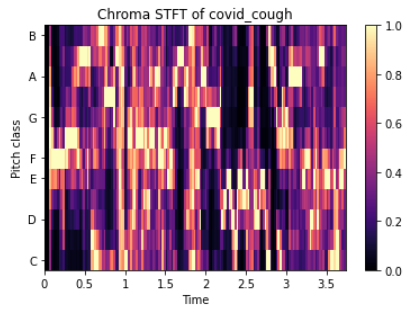
#### Extraction Technique

The Zero Crossing Rate (ZCR) [15] is the number of times the signal changes sign within a frame.

Here, if the sign of and differ then = 1 else = 0.

#### Chroma STFT

A 12-element illustration of spectral energy, here the bins represent the 12 distinctive pitch categories used to study music where each representation indicates how much energy each pitch categories has. The figure below uses short term Fourier transformation in order to compute Chroma features. These features carry harmonic and melodic characteristics of the audio while being robust to changes in timbre.



#### Mel Frequency Cepstral Coefficients (MFCCs)

MFCC is a compact representation of the spectrum, a primary feature in research areas that includes audio signals ranging from detecting cough sounds to automatic speech recognition [21].

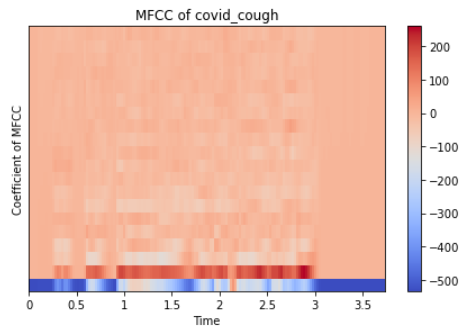
COVID-19 is said to cause inflammation of the upper airway and larynx [39]. This inflammation affects the flexibility of the vocal cords hence causing minor alterations in frequency. This alteration in frequency is captured using MFCC and is a vital component in detecting COVID-19 in cough sounds.

MFCC represents the sound spectrum by converting the audio signal via a sequence of steps to imitate the vocal tract.

The STFT uses a sliding window function centred at time to consecutively perform a Fourier transform of the signal , 𝑥(𝑛) is a finite length sequence, 𝑁 is the number of samples in the frequency domain. The result reveals that the alteration of the signal’s frequency content as time evolves.

Using MFCC, we can simulate the human auditory function and extract the physical characteristics of the spectrum. In filter-source theory, “the source is the vocal cords, and the filter represents the vocal tract.”

The audio output of an individual depends on the *shape* and *length* of the vocal tract. *Cepstrum coefficients* capture such characteristics of the vocal tract in an ordered manner [22].



#### Root Mean Square Energy

Root Mean Square Energy’s magnitude is the energy of the signal. It tells us about the loudness of the audio signal. RMSE is the square root of the mean square (the average of the squares of the magnitude of the audio frames) [23].

where is the amplitude of a particular audio frame. is the number of audio frames in one cough audio.

#### Spectral Roll off

Spectral Roll-off is the frequency below which 85% of the total spectral energy lies.

where is the percentile cutoff, is the spectral value at bin and and are the band edges.

#### Spectral Centroid

Spectral centroid is a metric to compute the “centre of mass” of a given signal. It tells us about sound “brightness”, which indicates the high-frequency content in a signal.

where  is the spectral magnitude at frequency bin ,  is the frequency at bin .

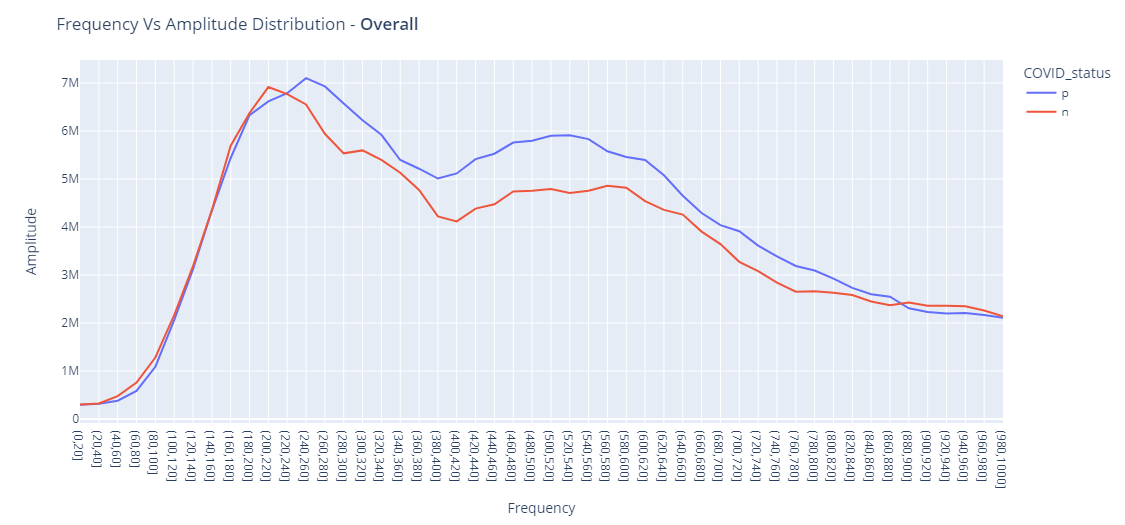
#### Spectral Bandwidth

The spectral bandwidth is defined as the extent of the power transfer function around the centre frequency [24].

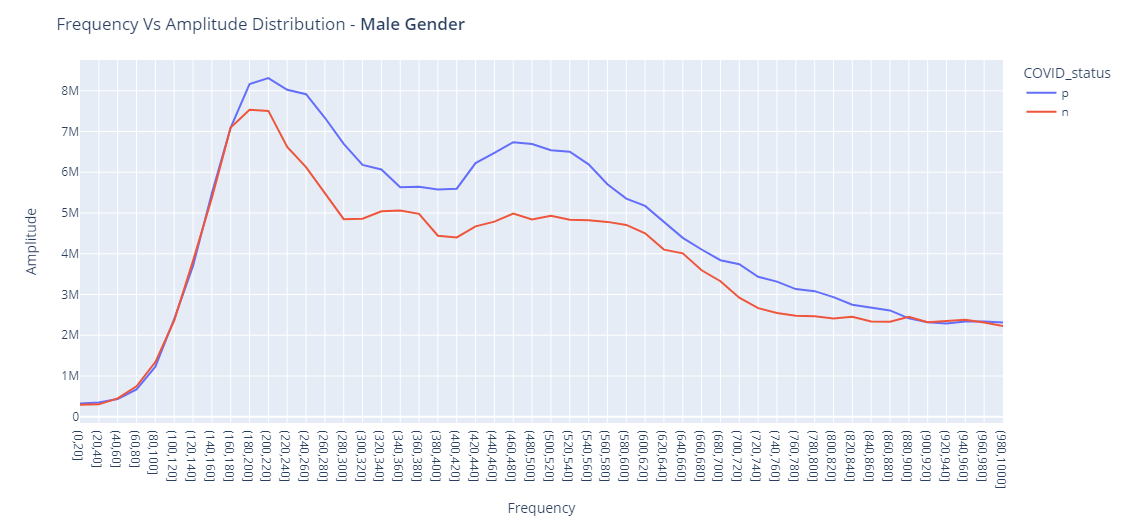
where  is the spectral magnitude at frequency bin ,  is the frequency at bin , and  is the spectral centroid.

### Graphical Results and Observations

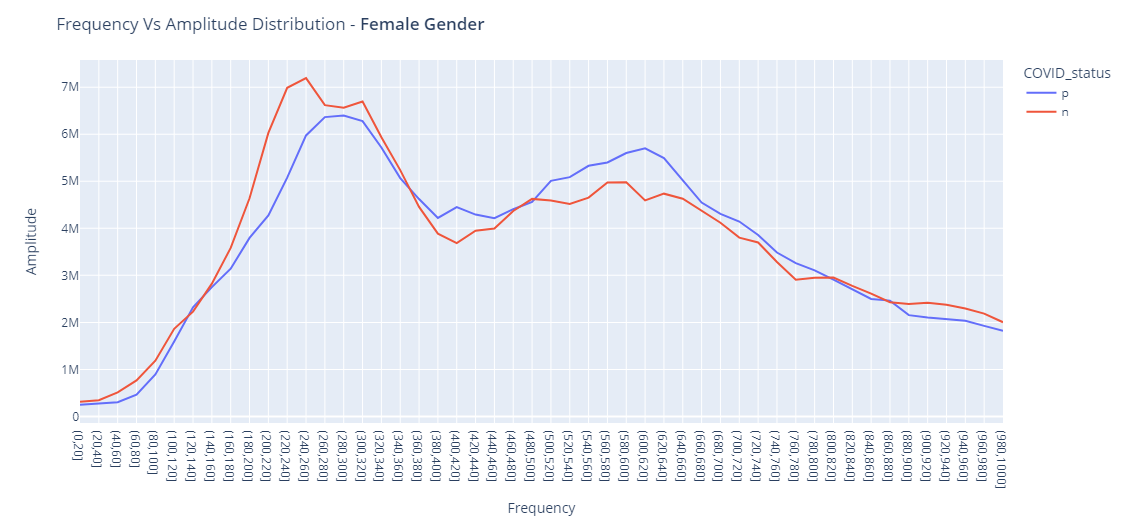
* These charts were produced in python using plotly library which is used to create interactive plots.



* Most of the characteristic frequency differences have been noticed in between **0-1kHz**
* Male:
  + A sudden change in the frequency amplitudes were noticed after 180 Hz
  + COVID positive cough showed consistently higher amplitudes of cough in range of 200 to 800 Hz



* Female:
  + Higher Amplitude for COVID Positive cough in certain frequency ranges: 140-360 Hz, 480-680 Hz —> These values are the Fundamental and 2nd harmonic frequency zones



### Physical Aspect of the Obtained Result

Another way of evaluating various models is the auROC. A ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. The Area Under the Curve (AUC) is the measure of the ability of a classifier to discriminate between classes.

## Conclusion and Discussion

In this summary, a pipeline for the Detection of COVID-19 in cough has been made. This pipeline includes processing cough sound, signal processing, feature extraction, training individual and ensemble classifiers, and classifying the cough sound. Feature extraction significantly impacted the model, hence much time was spent on ensuring the most negligible data loss. Small and unclean dataset

Future works include:

* Instalment of the App in devices like Google Home, Amazon Alexa
* Using the same idea for detecting other diseases
* Training the model on more extensive datasets for increased usability

Using other feature extraction techniques to make the model further better. This model will not exact the current testing techniques but will undoubtedly be a preliminary diagnostic tool.

# Cough Detection Model

## Methodology

### Data Collection

A Dataset consisting of 2665 cough audio recordings has been used for the analysis, consisting of 681 COVID-19 positive patients’ and 1984 non-COVID-19 patients’ recordings. These recordings were obtained from the Indian Institute of Science, Bangalore [17], containing 2515 records from Indian citizens and 150 from other countries.

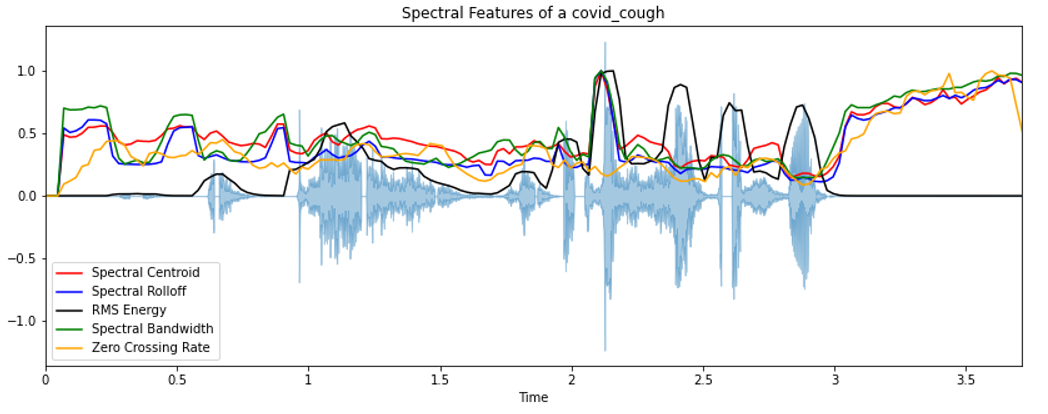
### Audio Preprocessing

Cough signals consist of vital information. The capturing of such time-varying characteristics would help us distinguish between a variety of coughs. For extracting these features, the cough recordings are in **.wav** format, and these waves are digitalised. These are converted into a one-dimensional array of digital values using a sampling technique. This conversion is done using Analog Digital Conversion, which consists of sampling, quantisation, and encoding. These digital values represent the amplitude, frequency at that given instance. Here the sampling rate is fixed to 22050 for all the recordings.

As the dataset was imbalanced, the Synthetic Minority Oversampling Technique (SMOTE)[41] was used for balancing the data. This technique has been established successfully in the field of cough detection and classification [20].

### Feature Extraction

A total of 25 time-frequency features were extracted from each cough recording, such as MFCC (first 20 coefficients), Chroma STFT, Root Mean Square Energy, Spectral Roll-off, Spectral Centroid and Zero Crossing Rate. These features were extracted and used after a careful literature review on cough identification, detection and classification [35][36][37].



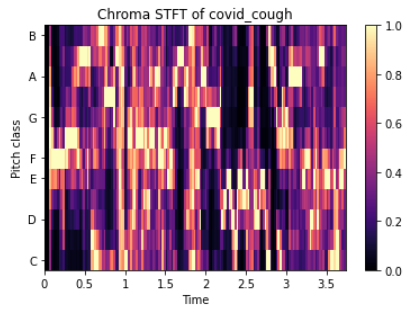
#### Zero-Crossing Rate

The Zero Crossing Rate (ZCR) [15] is the number of times the signal changes sign within a frame.

Here, if the sign of and differ then = 1 else = 0.

#### Chroma STFT

A 12-element illustration of spectral energy, here the bins represent the 12 distinctive pitch categories used to study music where each representation indicates how much energy each pitch categories has. The figure below uses short term Fourier transformation in order to compute Chroma features. These features carry harmonic and melodic characteristics of the audio while being robust to changes in timbre.



#### Mel Frequency Cepstral Coefficients (MFCCs)

MFCC is a compact representation of the spectrum, a primary feature in research areas that includes audio signals ranging from detecting cough sounds to automatic speech recognition [21].

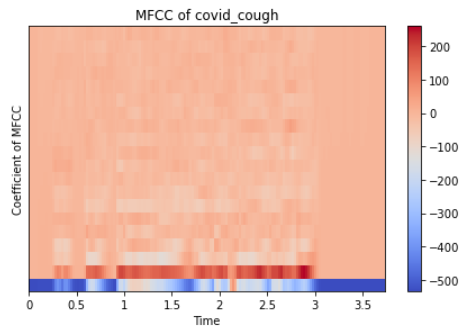
COVID-19 is said to cause inflammation of the upper airway and larynx [39]. This inflammation affects the flexibility of the vocal cords hence causing minor alterations in frequency. This alteration in frequency is captured using MFCC and is a vital component in detecting COVID-19 in cough sounds.

MFCC represents the sound spectrum by converting the audio signal via a sequence of steps to imitate the vocal tract.

The STFT uses a sliding window function centred at time to consecutively perform a Fourier transform of the signal , 𝑥(𝑛) is a finite length sequence, 𝑁 is the number of samples in the frequency domain. The result reveals that the alteration of the signal’s frequency content as time evolves.

Using MFCC, we can simulate the human auditory function and extract the physical characteristics of the spectrum. In filter-source theory, “the source is the vocal cords, and the filter represents the vocal tract.”

The audio output of an individual depends on the *shape* and *length* of the vocal tract. *Cepstrum coefficients* capture such characteristics of the vocal tract in an ordered manner [22].



#### Root Mean Square Energy

Root Mean Square Energy’s magnitude is the energy of the signal. It tells us about the loudness of the audio signal. RMSE is the square root of the mean square (the average of the squares of the magnitude of the audio frames) [23].

where is the amplitude of a particular audio frame. is the number of audio frames in one cough audio.

#### Spectral Roll off

Spectral Roll-off is the frequency below which 85% of the total spectral energy lies.

where is the percentile cutoff, is the spectral value at bin and and are the band edges.

#### Spectral Centroid

Spectral centroid is a metric to compute the “centre of mass” of a given signal. It tells us about sound “brightness”, which indicates the high-frequency content in a signal.

where  is the spectral magnitude at frequency bin ,  is the frequency at bin .

#### Spectral Bandwidth

The spectral bandwidth is defined as the extent of the power transfer function around the centre frequency [24].

where  is the spectral magnitude at frequency bin ,  is the frequency at bin , and  is the spectral centroid.

### Classifier Architectures

Five different classification models have been trained and tested on the dataset, keeping in mind the interpretability and explainability of the models:

#### Logistic Regression

We start with Logistic Regression. Logistic Regression is superior to other models such as SVM, ANN in the field of Clinical Diagnosis [25]. The output of a logistic regression model is given below:

where and are the model parameters. is interpreted as probability and hence is used for binary classification. Hyperparameter tuning and Regularisation has been done on the Logistic Regression model to reduce overfitting. Two types of penalties were considered used to minimise the loss function: lasso, ridge and used AUC scores to compare the models.

#### Explainable Boosting Machine

Explainable Boosting Machine [26] is a glass-box model created by Microsoft. It is a modification of a generalised additive model (GAM), also known as GA2M model [27] and is of the form:

Where  is the link function that adjusts the GAM to different configurations,   is a feature function that is learned by Explainable Boosting Classifier using machine learning techniques like Gradient Boosting and Bagging, represents the pairwise interaction function [29] of these features. Being an additive model, contributions of each feature to the prediction can be observed and hence these contributions can be understood, making the model completely interpretable.

#### Decision Tree Classifier

Decision Tree Classifier [30] is a tree-structured classification model that replicates the human thinking ability while making a decision. The tree structure of the Decision Tree Classifier is ideal for capturing the interactions between the model’s features, hence maintaining the interpretability of the model. The relation between outcome and features is given by [31]:

Each instance falls into precisely one leaf node (=subset ). Where is an identity function that returns one if is a subset of otherwise 0. If an instance falls into a leaf node , the predicted outcome is , where  is the average of all training instances in the leaf node . The selection of best attributes (also known as Attribute selection measure) is made using the Information Gain (or Entropy) technique, and the maximum depth of the tree is taken as 5.

#### Random Forest Classifier

Random Forest Classifier [31] is a supervised learning algorithm that creates, fits decision trees on randomly selected subsamples, and selects the best solution/prediction among these trees by voting and aggregation in classification and Regression. These individual trees are generated using indicators such as information gain and the Gini index. Random Forest Classifier is also a good indicator of feature importance [32]. RFC also overcomes the overfitting caused by individual decision trees through the randomness of subsample and features’ selection. Here Importance of node is given by

Where is the weighted number of samples reaching node , is the impurity value of node , is the child node from left split on node , and is the child node from right split on node .

#### XG Boost

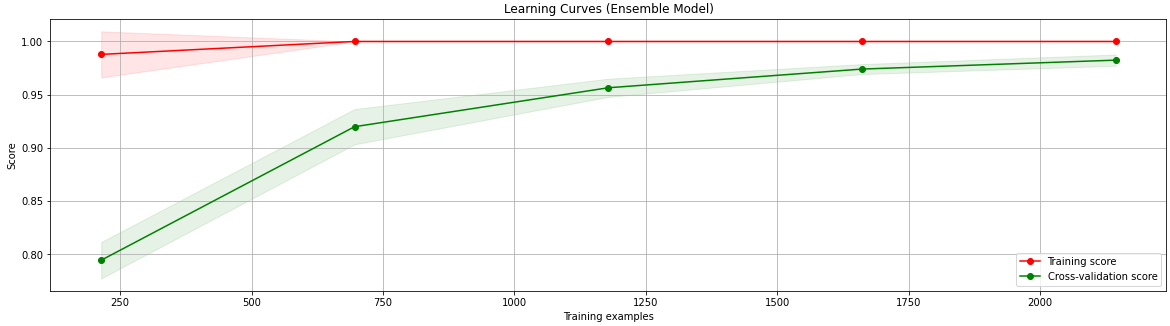
Extreme Gradient Boosting[33] is a supervised learning algorithm implements boosting, an ensemble learning technique that corrects for deficiencies in previous models.

#### Ensemble Model

Machine Learning algorithms have certain limitations, and developing a high accuracy and sensitivity model can be challenging. Hence building and combining multiple models can help boost these metrics.

Ensemble Learning is a machine learning approach to obtain better precision than any single model performance by combining results of multiple models.

Here Stacking Ensemble model has been used. Stacking [34] uses a “Level 1” classifier to combine multiple “Level 0” classification models. “Level 0 or Base” Model are those models which fit the training data and the predictions are made. “Level 1 or Meta” model is the model that learns how to merge the Base models’ predictions effectively.

Here, Logistic Regression, Explainable Boosting Machine, XG Boost, Decision Tree and Random Forest Classifier has been used as “Level 0” models. Random Forest Classifier has been used as the “Level 1” model.

### Results

The table below compares the different classification models’ results using Accuracy, Precision, Recall, and F1 Score metrics. These metrics are for the Detection of COVID positive labels. The results from the ensemble model show that cough sound can be used to detect COVID-19. The addition of more data could improve the models’ sensitivity.

The table shows that the Ensemble Model exhibits the superlative performance, with an auROC of 0.974 and a corresponding accuracy of 98% with Precision and Recall of 97% each.

The most straightforward metric to understand here is accuracy. Accuracy is the ratio of correct predictions to the total number of predictions. In this case, we need a high accuracy model, but it is important to note that we also need to ensure that the coughs of a COVID positive patient are not wrongly labelled. This brings us to the metrics: Recall and Precision, where Recall is the measure of the Total True positives (in this case COVID positive patients’ cough) detected by the Model and Precision is the ratio of true positives to all the Positives. In this case, we need a model which would:

* Detect all the COVID positive patients’ coughs correctly
* Differentiate between a COVID positive and a healthy patients’ cough

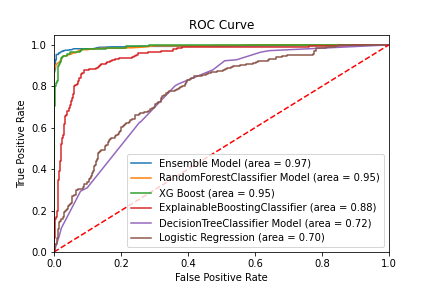
Now among Recall and Precision, we need to prioritise a metric to help us compare various models. If precision is taken as the metric, the model will differentiate between the different types of coughs, but there is a possibility that a COVID positive patient’s cough is wrongly labelled as healthy, which can cause further problems as that patient will now be considered as healthy and can further spread the virus.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Decision Tree | 72% | 68% | 80% | 74% |
| EBM | 88% | 84% | 92% | 88% |
| Logistic Regression | 70% | 69% | 72% | 70% |
| RFC | 95% | 97% | 96% | 95% |
| XGB | 95% | 93% | 97% | 95% |
| **Ensemble Model** | **98%** | **97%** | **98%** | **98%** |

Using Recall ensures that all the specific label (COVID positive cough here) data points are correctly identified. Using Recall would also mean that certain Healthy coughs could be wrongly labelled as COVID positive coughs, but this inaccurate labelling will not lead to the problems encountered while using precision as the metric because now there will not be a further spread of the virus as all the COVID patients are identified.

Hence, Recall is the correct metric for evaluating the models used for detecting COVID positive patients, and a Recall of 1 would be ideal for such a model.

Based on the above result, Ensemble Model performs better than the rest of the classifiers with a Recall of 0.98, while Random Forest and Ensemble both that the same precision of 0.97. The figure below compares the Area under Receiver Operating Characteristics of various models.



Another way of evaluating various models is the auROC. A ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. The Area Under the Curve (AUC) is the measure of the ability of a classifier to discriminate between classes.

## Methodology

### Data Collection

A Dataset consisting of 2665 cough audio recordings has been used for the analysis, consisting of 681 COVID-19 positive patients’ and 1984 non-COVID-19 patients’ recordings. These recordings were obtained from the Indian Institute of Science, Bangalore [17], containing 2515 records from Indian citizens and 150 from other countries.

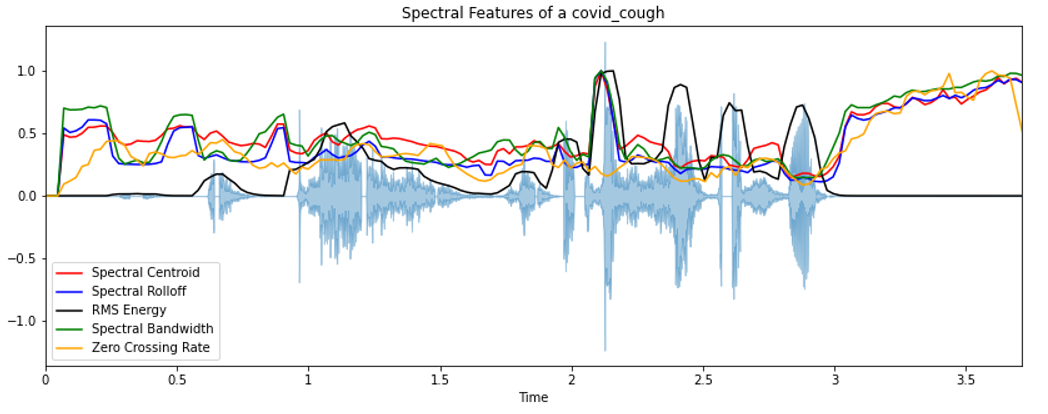
### Audio Preprocessing

Cough signals consist of vital information. The capturing of such time-varying characteristics would help us distinguish between a variety of coughs. For extracting these features, the cough recordings are in **.wav** format, and these waves are digitalised. These are converted into a one-dimensional array of digital values using a sampling technique. This conversion is done using Analog Digital Conversion, which consists of sampling, quantisation, and encoding. These digital values represent the amplitude, frequency at that given instance. Here the sampling rate is fixed to 22050 for all the recordings.

As the dataset was imbalanced, the Synthetic Minority Oversampling Technique (SMOTE)[41] was used for balancing the data. This technique has been established successfully in the field of cough detection and classification [20].

### Feature Extraction

A total of 25 time-frequency features were extracted from each cough recording, such as MFCC (first 20 coefficients), Chroma STFT, Root Mean Square Energy, Spectral Roll-off, Spectral Centroid and Zero Crossing Rate. These features were extracted and used after a careful literature review on cough identification, detection and classification [35][36][37].



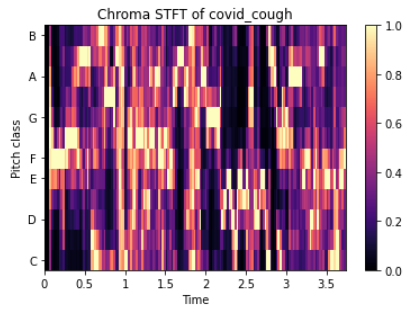
#### Zero-Crossing Rate

The Zero Crossing Rate (ZCR) [15] is the number of times the signal changes sign within a frame.

Here, if the sign of and differ then = 1 else = 0.

#### Chroma STFT

A 12-element illustration of spectral energy, here the bins represent the 12 distinctive pitch categories used to study music where each representation indicates how much energy each pitch categories has. The figure below uses short term Fourier transformation in order to compute Chroma features. These features carry harmonic and melodic characteristics of the audio while being robust to changes in timbre.



#### Mel Frequency Cepstral Coefficients (MFCCs)

MFCC is a compact representation of the spectrum, a primary feature in research areas that includes audio signals ranging from detecting cough sounds to automatic speech recognition [21].

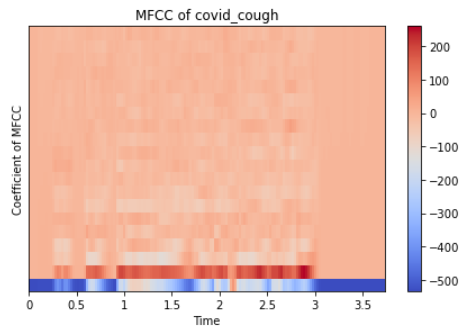
COVID-19 is said to cause inflammation of the upper airway and larynx [39]. This inflammation affects the flexibility of the vocal cords hence causing minor alterations in frequency. This alteration in frequency is captured using MFCC and is a vital component in detecting COVID-19 in cough sounds.

MFCC represents the sound spectrum by converting the audio signal via a sequence of steps to imitate the vocal tract.

The STFT uses a sliding window function centred at time to consecutively perform a Fourier transform of the signal , 𝑥(𝑛) is a finite length sequence, 𝑁 is the number of samples in the frequency domain. The result reveals that the alteration of the signal’s frequency content as time evolves.

Using MFCC, we can simulate the human auditory function and extract the physical characteristics of the spectrum. In filter-source theory, “the source is the vocal cords, and the filter represents the vocal tract.”

The audio output of an individual depends on the *shape* and *length* of the vocal tract. *Cepstrum coefficients* capture such characteristics of the vocal tract in an ordered manner [22].



#### Root Mean Square Energy

Root Mean Square Energy’s magnitude is the energy of the signal. It tells us about the loudness of the audio signal. RMSE is the square root of the mean square (the average of the squares of the magnitude of the audio frames) [23].

where is the amplitude of a particular audio frame. is the number of audio frames in one cough audio.

#### Spectral Roll off

Spectral Roll-off is the frequency below which 85% of the total spectral energy lies.

where is the percentile cutoff, is the spectral value at bin and and are the band edges.

#### Spectral Centroid

Spectral centroid is a metric to compute the “centre of mass” of a given signal. It tells us about sound “brightness”, which indicates the high-frequency content in a signal.

where  is the spectral magnitude at frequency bin ,  is the frequency at bin .

#### Spectral Bandwidth

The spectral bandwidth is defined as the extent of the power transfer function around the centre frequency [24].

where  is the spectral magnitude at frequency bin ,  is the frequency at bin , and  is the spectral centroid.

### Classifier Architectures

Five different classification models have been trained and tested on the dataset, keeping in mind the interpretability and explainability of the models:

#### Logistic Regression

We start with Logistic Regression. Logistic Regression is superior to other models such as SVM, ANN in the field of Clinical Diagnosis [25]. The output of a logistic regression model is given below:

where and are the model parameters. is interpreted as probability and hence is used for binary classification. Hyperparameter tuning and Regularisation has been done on the Logistic Regression model to reduce overfitting. Two types of penalties were considered used to minimise the loss function: lasso, ridge and used AUC scores to compare the models.

#### Explainable Boosting Machine

Explainable Boosting Machine [26] is a glass-box model created by Microsoft. It is a modification of a generalised additive model (GAM), also known as GA2M model [27] and is of the form:

Where  is the link function that adjusts the GAM to different configurations,   is a feature function that is learned by Explainable Boosting Classifier using machine learning techniques like Gradient Boosting and Bagging, represents the pairwise interaction function [29] of these features. Being an additive model, contributions of each feature to the prediction can be observed and hence these contributions can be understood, making the model completely interpretable.

#### Decision Tree Classifier

Decision Tree Classifier [30] is a tree-structured classification model that replicates the human thinking ability while making a decision. The tree structure of the Decision Tree Classifier is ideal for capturing the interactions between the model’s features, hence maintaining the interpretability of the model. The relation between outcome and features is given by [31]:

Each instance falls into precisely one leaf node (=subset ). Where is an identity function that returns one if is a subset of otherwise 0. If an instance falls into a leaf node , the predicted outcome is , where  is the average of all training instances in the leaf node . The selection of best attributes (also known as Attribute selection measure) is made using the Information Gain (or Entropy) technique, and the maximum depth of the tree is taken as 5.

#### Random Forest Classifier

Random Forest Classifier [31] is a supervised learning algorithm that creates, fits decision trees on randomly selected subsamples, and selects the best solution/prediction among these trees by voting and aggregation in classification and Regression. These individual trees are generated using indicators such as information gain and the Gini index. Random Forest Classifier is also a good indicator of feature importance [32]. RFC also overcomes the overfitting caused by individual decision trees through the randomness of subsample and features’ selection. Here Importance of node is given by

Where is the weighted number of samples reaching node , is the impurity value of node , is the child node from left split on node , and is the child node from right split on node .

#### XG Boost

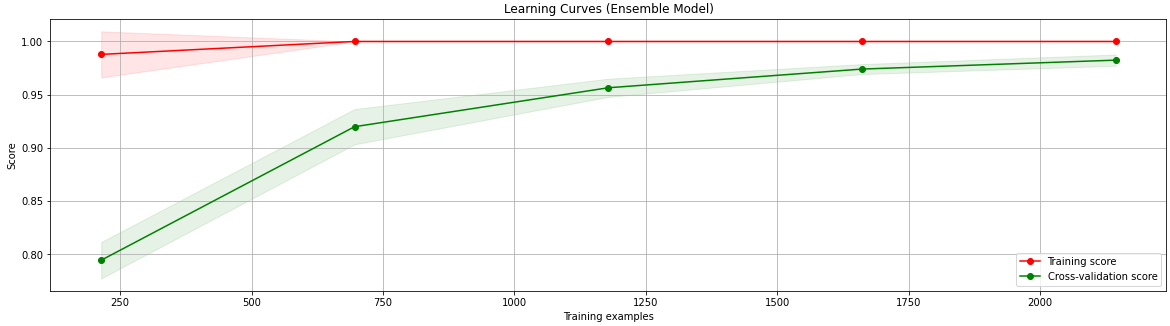
Extreme Gradient Boosting[33] is a supervised learning algorithm implements boosting, an ensemble learning technique that corrects for deficiencies in previous models.

#### Ensemble Model

Machine Learning algorithms have certain limitations, and developing a high accuracy and sensitivity model can be challenging. Hence building and combining multiple models can help boost these metrics.

Ensemble Learning is a machine learning approach to obtain better precision than any single model performance by combining results of multiple models.

Here Stacking Ensemble model has been used. Stacking [34] uses a “Level 1” classifier to combine multiple “Level 0” classification models. “Level 0 or Base” Model are those models which fit the training data and the predictions are made. “Level 1 or Meta” model is the model that learns how to merge the Base models’ predictions effectively.

Here, Logistic Regression, Explainable Boosting Machine, XG Boost, Decision Tree and Random Forest Classifier has been used as “Level 0” models. Random Forest Classifier has been used as the “Level 1” model.

### Results

The table below compares the different classification models’ results using Accuracy, Precision, Recall, and F1 Score metrics. These metrics are for the Detection of COVID positive labels. The results from the ensemble model show that cough sound can be used to detect COVID-19. The addition of more data could improve the models’ sensitivity.

The table shows that the Ensemble Model exhibits the superlative performance, with an auROC of 0.974 and a corresponding accuracy of 98% with Precision and Recall of 97% each.

The most straightforward metric to understand here is accuracy. Accuracy is the ratio of correct predictions to the total number of predictions. In this case, we need a high accuracy model, but it is important to note that we also need to ensure that the coughs of a COVID positive patient are not wrongly labelled. This brings us to the metrics: Recall and Precision, where Recall is the measure of the Total True positives (in this case COVID positive patients’ cough) detected by the Model and Precision is the ratio of true positives to all the Positives. In this case, we need a model which would:

* Detect all the COVID positive patients’ coughs correctly
* Differentiate between a COVID positive and a healthy patients’ cough

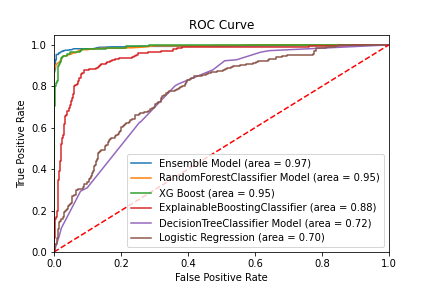
Now among Recall and Precision, we need to prioritise a metric to help us compare various models. If precision is taken as the metric, the model will differentiate between the different types of coughs, but there is a possibility that a COVID positive patient’s cough is wrongly labelled as healthy, which can cause further problems as that patient will now be considered as healthy and can further spread the virus.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Decision Tree | 72% | 68% | 80% | 74% |
| EBM | 88% | 84% | 92% | 88% |
| Logistic Regression | 70% | 69% | 72% | 70% |
| RFC | 95% | 97% | 96% | 95% |
| XGB | 95% | 93% | 97% | 95% |
| **Ensemble Model** | **98%** | **97%** | **98%** | **98%** |

Using Recall ensures that all the specific label (COVID positive cough here) data points are correctly identified. Using Recall would also mean that certain Healthy coughs could be wrongly labelled as COVID positive coughs, but this inaccurate labelling will not lead to the problems encountered while using precision as the metric because now there will not be a further spread of the virus as all the COVID patients are identified.

Hence, Recall is the correct metric for evaluating the models used for detecting COVID positive patients, and a Recall of 1 would be ideal for such a model.

Based on the above result, Ensemble Model performs better than the rest of the classifiers with a Recall of 0.98, while Random Forest and Ensemble both that the same precision of 0.97. The figure below compares the Area under Receiver Operating Characteristics of various models.



Another way of evaluating various models is the auROC. A ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. The Area Under the Curve (AUC) is the measure of the ability of a classifier to discriminate between classes.

## Conclusion and Discussion:

In this summary, a pipeline for the Detection of COVID-19 in cough has been made. This pipeline includes processing cough sound, signal processing, feature extraction, training individual and ensemble classifiers, and classifying the cough sound. Feature extraction significantly impacted the model, hence much time was spent on ensuring the most negligible data loss. Small and unclean dataset

Future works include:

* Instalment of the App in devices like Google Home, Amazon Alexa
* Using the same idea for detecting other diseases
* Training the model on more extensive datasets for increased usability

Using other feature extraction techniques to make the model further better. This model will not exact the current testing techniques but will undoubtedly be a preliminary diagnostic tool.

# References

[1] Mouawad, P., Dubnov, T. & Dubnov, S. Robust Detection of COVID-19 in Cough Sounds. *SN COMPUT. SCI.* **2,**34 (2021). <https://doi.org/10.1007/s42979-020-00422-6>

[2] [Online]. Available: <https://www.worldometers.info/coronavirus/> [Accessed 16 June 2021].

[3] Kucharski, A. J., Klepac, P., Conlan, A. J., Kissler, S. M., Tang, M. L., Fry, H., ... & Simons, D. (2020). Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*, *20*(10), 1151-1160.

[4] Alafif, T., Tehame, A. M., Bajaba, S., Barnawi, A., & Zia, S. (2021). Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *International Journal of Environmental Research and Public Health*, *18*(3), 1117.

[5] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, *395*(10223), 497-506.

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

[7] Vrindavanam, J., Srinath, R., Shankar, H. H., & Nagesh, G. (2021, April). Machine Learning based COVID-19 Cough Classification Models-A Comparative Analysis. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 420-426). IEEE.

[8] Deshpande, G., & Schuller, B. (2020). An overview on audio, signal, speech, & language processing for covid-19. *arXiv preprint arXiv:2005.08579*.

[9] Abeyratne, U. R., Swarnkar, V., Setyati, A., & Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia. *Annals of biomedical engineering*, *41*(11), 2448-2462.

[10] Belkacem, A. N., Ouhbi, S., Lakas, A., Benkhelifa, E., & Chen, C. (2021). End-to-End AI-Based Point-of-Care Diagnosis System for Classifying Respiratory Illnesses and Early Detection of COVID-19: A Theoretical Framework. *Frontiers in Medicine*, *8*, 372.

[11] Schuller, B. W., Schuller, D. M., Qian, K., Liu, J., Zheng, H., & Li, X. (2021). COVID-19 and Computer Audition: An Overview on What Speech & SoundAnalysis Could Contribute in theSARS-CoV-2 Corona Crisis. Frontiers in Digital Health, 3, 14.

[12] Chatrzarrin, H., Arcelus, A., Goubran, R., & Knoefel, F. (2011, May). Feature extraction for the differentiation of dry and wet cough sounds. In 2011 IEEE international symposium on medical measurements and applications (pp. 162-166). IEEE.

[13] Bhateja, V., Taquee, A., & Sharma, D. K. (2019, November). Pre-processing and classification of cough sounds in noisy environment using SVM. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 822-826). IEEE.

[14] Matos, S., Birring, S. S., Pavord, I. D., & Evans, H. (2006). Detection of cough signals in continuous audio recordings using hidden Markov models. IEEE Transactions on Biomedical Engineering, 53(6), 1078-1083.

[15] Bachu, R. G., Kopparthi, S., Adapa, B., & Barkana, B. D. (2010). Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In Advanced techniques in computing sciences and software engineering (pp. 279-282). Springer, Dordrecht.

[16] Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, M. S., Ali, K., ... & Nabeel, M. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Informatics in Medicine Unlocked, 20, 100378.

[17] Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S. R., Ghosh, P. K., & Ganapathy, S. (2020). Coswara--A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. arXiv preprint arXiv:2005.10548.

[18] Wang, H., Xu, Y., & Li, M. (2011, August). Study on the MFCC similarity-based voice activity detection algorithm. In 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC) (pp. 4391-4394). IEEE.

[19] Dunne, Robert (2020), “COVID-19 CNN MFCC classifier”, Mendeley Data, V1, doi: 10.17632/ww5dfy53cw.1

[20] Windmon, A., Minakshi, M., Bharti, P., Chellappan, S., Johansson, M., Jenkins, B. A., & Athilingam, P. R. (2018). Tussiswatch: A smart-phone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and congestive heart failure. IEEE journal of biomedical and health informatics, 23(4), 1566-1573.

[21] Han, W., Chan, C. F., Choy, C. S., & Pun, K. P. (2006, May). An efficient MFCC extraction method in speech recognition. In 2006 IEEE international symposium on circuits and systems (pp. 4-pp). IEEE.

[22] Abushakra, A., Faezipour, M., Abumunshar, A., & Patra, P. (2012). Breathing Movement Classification Using MFCCs. .

[23] [Online]. <https://rramnauth2220.github.io/blog/posts/code/200525-feature-extraction.html>

[24] Liu, Z., Huang, J., Wang, Y., & Chen, T. (1997, June). Audio feature extraction and analysis for scene classification. In Proceedings of First Signal Processing Society Workshop on Multimedia Signal Processing (pp. 343-348). IEEE.

[25] Botha, G. H. R., Theron, G., Warren, R. M., Klopper, M., Dheda, K., Van Helden, P. D., & Niesler, T. R. (2018). Detection of tuberculosis by automatic cough sound analysis. Physiological measurement, 39(4), 045005.

[26] Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). Interpretml: A unified framework for machine learning interpretability. arXiv preprint arXiv:1909.09223.

[27] Hastie, T., & Tibshirani, R. (1987). Generalised additive models: some applications. Journal of the American Statistical Association, 82(398), 371-386.

[28] Lou, Y., Caruana, R., Gehrke, J., & Hooker, G. (2013, August). Accurate intelligible models with pairwise interactions. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 623-631).

[29] Lou, Y., Caruana, R., Gehrke, J., & Hooker, G. (2013, August). Accurate intelligible models with pairwise interactions. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 623-631).

[30] Patel, H., & Prajapati, P. (2018). International Journal of Computer Sciences and Engineering Open Access. Int. J. Comput. Sci. Eng, 6(10).

[31] Molnar, Christoph. “Interpretable machine learning. A Guide for Making Black Box Models Explainable”, 2019. <https://christophm.github.io/interpretable-ml-book/>.

[32] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

[33] Chen, T. (2016). Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’16) (pp. 785-794).

[34] Ting, K. M., & Witten, I. H. (1997). Stacked Generalisation: when does it work?.

[35] Pramanik, A., & Raha, R. (2012, October). Automatic Speech Recognition using correlation analysis. In 2012 World congress on information and communication technologies (pp. 670-674). IEEE.

[36] Al-Khassaweneh, M., & Bani Abdelrahman, R. E. (2013). A signal processing approach for the diagnosis of asthma from cough sounds. Journal of medical engineering & technology, 37(3), 165-171.

[37] Boyanov, B., & Hadjitodorov, S. (1997). Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases. IEEE Engineering in Medicine and Biology Magazine, 16(4), 74-82.

[38] [Online] <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>

[39] [Online] <https://health.clevelandclinic.org/what-is-covid-19-voice-and-what-causes-it/>

[40] Lara Orlandic, Tomas Teijeiro, & David Atienza. (2021). The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms (2.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4498364>

[41] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.