International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

Detection of Covid 19 from Cough Audio using Convolutional Neural Networks

Apoorva Gondane, Bhushan Birwadkar, Prajwal Aher, Megha Gupta

Abstract— COVID-19, also known as the coronavirus which has sparked a global pandemic that has infected a lot of people and has also caused a large number of deaths globally. Trouble breathing, fever, and exhaustion are common early signs of the virus, and over 60% of patients had a dry cough. We propose a COVID-19 cough audio classifier based on machine learning technology that can distinguish COVID-19 positive cough audio from COVID-19 negative cough audio and healthy cough sample recorded on a smartphone. This type of screening is non-contact, easy to use, and can reduce the testing center effort while also limiting transmission by advising early self-isolation to patients with a cough that suggests COVID-19. The datasets which are used in this study are the COUGHVID and Virufy dataset. Over 25,000 crowdsourced cough audio recordings are available in the COUGHVID collection, covering different participant ages, genders, country, and COVID-19 statuses. The Virufy dataset is also used which contains segmented cough audio samples from patients. Our findings reveal that, while all classifiers were able to detect COVID-19 coughs, the CNN classifier with 10 layers had the greatest performance, with area under the ROC curve (AUC) of 0.80 in distinguishing between COVID-19 positive cough audio, symptomatic and healthy cough audio. This type of cough audio classification is possibly a beneficial and viable means of non-contact screening because it is cost-effective and simple to implement.

Index Terms— Audio Classification, CNN Classifier, Convolutional Neural Networks, Coronavirus, Cough Audio, Covid-19, Machine Learning

---

1 Introduction

THE Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. It starts by infecting the mucous membranes in the throat and moves down the respiratory tract leading to the lungs. It was declared a global pandemic on February 11, 2020, by the World Health Organization (WHO). The most common symptoms of the COVID-19 disease are fatigue, fever, and dry coughs. Other symptoms include shortness of breath, joint pain, muscle pain, gastrointestinal symptoms, and loss of smell or taste.

The use of human-generated sounds as a diagnostic tool for different diseases presents tremendous potential for early detection as well as an inexpensive solution that is embedded in commodity devices that could be rolled out to the people. It is even truer if these solutions unobtrusively cloud track individuals during their everyday lives.

A recent study has started to examine how human respiratory sounds such as cough audio, recorded voice and breathing audio from hospital-tested COVID-19 positive vary from healthy people's sounds. A cough-based diagnosis of COVID-19 will also have to be taken into account with the respiratory and non-respiratory sounds data associated with all the conditions are mentioned. Data review of a large crowd sourced respiratory speech/sounds dataset that has been obtained to accurately diagnose COVID positive cases was mentioned in. It explains cough and respiration to understand how noticeable the sounds of COVID-19 are from those in asthma or protection of individuals. Speech recordings from hospital patients with COVID-19 are analyzed in to categorize patient's health status automatically. The digital stethoscope, lung sounds data is being used in as a diagnostic signal for COVID-19; in, the COVID-19 related cough detection analysis obtained by telephone is presented using a cohort of nearly 50 patients with COVID-19 versus other pathological coughs trained in a series of models. Thus, our main objective is to build a machine learning model for the detection of COVID 19 from cough audio.

2 Traditional Covid-19 Detection Approach

The "reverse transcription polymerase chain reaction (RT-PCR)" test, known as the best test to detect COVID-19, is performed with the recommendation of WHO. This test is time-consuming and costly, and results are obtained late. Therefore, medical imaging methods are used to detect COVID-19 patients in addition to this test. Among these, the most used are Computed tomography (CT), and chest radiography (X-ray) images. Another helpful method is approaches based on the analysis of cough acoustic sounds produced by the respiratory system. There are many studies in the literature using CT and X-ray images in the detection of COVID-19 [1].

The Key steps in the COVID-19 PCR test are given as follows: 1. Sample collection: In this step a Swab collector uses a swab from the patient to collect respiratory particles in their nose. A swab stick a long and flexible stick that goes into the nose. Different types of nose swabs are there, such as nasal swabs which are collected from inside of nostrils and nasopharyngeal swabs which are collected from deep inside from the nasal cavity for collection. These types of swabs are collected as testing material for the COVID-19 PCR test. After collection, safely sealed swab in a tube is then sent to a laboratory. 2. Extraction: After receiving the sample, the genetic materials are extracted by the laboratory scientists.

International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

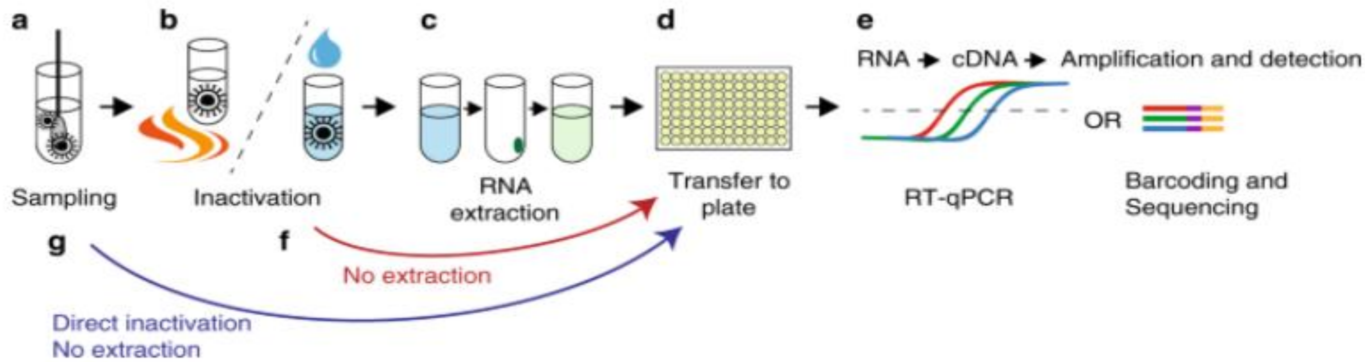

Fig. 1. Schematic overview of SARS-CoV-2 RT-PCR testing procedure [6]

3. PCR: In this step a PCR machine also called as thermal cycler uses some chemicals and enzymes. In this machine, the amount of the required genetic material in the test tube increases after each heating and cooling cycle. After number of cycles, millions of copies of a tiny portion of the SARS-CoV-2 virus's genetic material generates in the test tube. One of the chemicals present in the tube is responsible to produce a fluorescent light when SARS-CoV-2 is detected in the given sample. Once increased enough, the machine can detect this signal. A software is used by the scientists to interpret that the received signal is Covid Positive.

3 Literature Survey

Sound's utility as a potential predictor of actions and health has long been acknowledged by researchers and experts. To detect noises from the heart or lungs, for example, a digital stethoscope have utilized purpose-built external microphone recorders. These tooks require skilled physicians to listen and interpret, and they are being rapidly replaced by various technologies such as a range of imaging techniques (e.g., MRI, sonography), which are easier to inspect and interpret. Recent advances in automated audio interpretation and modelling, on the other hand, have the potential to revers this trend, making sound a low-cost and widely available alternative. Microphones have recently been used in consumer electronics such as smartphones and wearables for sound processing.

Orlandic Lara et al. uses the "COUGHVID" crowd sourced dataset for cough audio analysis in COVID-19 symptoms; More than 20,000 crowd sourced cough recordings reflecting a broad range of topic gender, age, locations, and COVID-19 status are there which are given in the COUGHVID dataset. Collected dataset has a series of 121 cough sounds and 94 sounds of no cough first-hand to train the classifier including various kinds of background noises such as voices, laughter, silence, etc. They have taken self-reported status variables where 25% of the recording audio has healthy values, 25% of the recording audio has COVID values, 35% of the recording audio has Symptomatic values and remaining 15% recording audio was non-reported status. The percentage of COVID positive, Symptoms of COVID, and healthy subjects were 7.5%, 15.5, and 77% from the subject of 65.5% males and 34.5% females respectively. It generated as 93.0% of audible dyspnea, 90% of wheezing, 98.7% of stridor, 99.1% of choking, or 99.2% of nasal congestion from 632 labelled COVID-19 cough records [3].

Chaudhari Gunavant et al. the research done by them show that to train machine learning models to detect COVID-19 they collected crowdsourced cough audio samples worldwide on smartphones; various teams have collected several COVID-19 cough recording datasets. However, each of these models has been trained on data from a variety of formats and recording settings; collected additional counting and vocal recordings, authors exclusively collect cough recordings. The datasets came from different sources, such as from clinical environments data, crowd sourcing, still it combined with COVID-19 status labels that can be used to create an Artificial Intelligence algorithm that correctly predicts COVID-19 infection with a 77.1% of ROC and AUC of 75.2–78.3%. In addition, without more training using the relevant samples, this AI algorithm can generalize to crowd-sourced samples from Latin America and clinical samples from South Asia [4].

Mohamed Bader et al. proposed a model with the combination of MFCCs (Mel-Frequency Cepstral Coefficients) and SSP (Speech Signal Processing) to the extraction of samples from non-COVID and COVID, and it finds the person correlation from their relationship coefficients. The result indicates high similarity between various breathing respiratory sound patterns and COVID cough sound patterns in MFCCs, although MFCC speech is more robust between non-COVID-19 samples and COVID-19 samples. Even though the findings are provisional; it is possible to remove the various patients cough audio with COVID-19 for future testing analysis. They have collected 3 female and 4 male voices from 7 healthy patients, and 2 female, 5 male voices from 7 COVID-19 patients were obtained from their dataset. From Zulekha hospital in Sharjah, they were obtained COVID-19 infected patients' data. The data contains four cough samples from each speaker, the voice of numbers counting from 1 to 10 of each speaker, and 4 to 5 times deep breath of each speaker. In addition, when record-

International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

ing their audio signals, the patients were their heads straight in a comfortable way; 3 recordings for each speaker are acquired from smartphone devices in data collection, which can affect the quality of sound [5].

Imran Ali et al. implemented an Artificial Intelligence based screening solution to detect COVID, transferable through a smart mobile phone application called AI4COVID-19, which was suggested, developed, and finally tested. The mobile application records and sends the sample to the AI-based clouds running in the cloud triple 3-second cough sounds and comeback reaction within two minutes. Generally, over 30 medical conditions associated with non COVID-19, cough is a basic indication. This makes it an incredibly difficult issue is cough alone to diagnose COVID disease. By investigating morphological direction changes with dissimilarities from cough respiratory achieves an accuracy of 88.76% [6].

Laguarta Jord et al. proposed an Artificial Intelligence model that detects the COVID symptoms from cough sound recordings. This model allows a solution with no cost to pre-screen COVID19 sound samples country-wide. It achieves 97.1% of accuracy to detet the COVID positive symptom from cough audio and 100% of accuracy to detect asymptomatic patients based on the cough audio samples selected from the dataset [7].

4 Proposed System

The system that we propose in this paper for identifying COVID 19 illustrated in Figure 2 is a faster and safer alternative to the traditional COVID 19 testing. This approach is cost effective and thus it can be easily implemented and distributed and does not require any extra accessories which increases the ease with which it can be used. Additionally, this approach will also ensure that social distancing is maintained and will decrease overcrowding at testing facilities making it easier for people to know whether they need medical attention or not.

The system involves the use of a smartphone which is available to most of the people. The user who suspects that he might have COVID-19 can use our app and record an audio recording while coughing. This recording is converted into a spectrogram which is a visual representation of the recorded audio. The spectrogram is given as an input to the machine learning model which classifies it as COVID-19 positive, healthy or COVID symptomatic.

5 Dataset

A dry cough was one of the most common symptoms in COVID-19, which is present approximately in 67.7% of cases. Cough sound classification is an emerging field of medical research that has successfully lifted up the signal processing and artificial intelligence tools to rapidly and inconspicuosly diagnose respiratory conditions like asthama, pertussis and pneumonia using nothing more than a smartphone and its built-in microphone devices. We have used two datasets to train our model first one being the COUGHVID dataset and the second one is the Virufy dataset.

COUGHVID is a large crowd sourced dataset, which is publicly available collection of cough audio recordings. It is one of the largest known public COVID-19 cough audio dataset, having approximately 25,000 cough audio recordings – 1,155 of which claim to have COVID-19 – originating from all over the world. Four professional physicians annotated a portion of the dataset to evaluate whether crowdsourced samples are really from COVID-19 patients, adding another layer of validation to the dataset. In addition to COVID-19 diagnosis, the expert labels and given metadata file, give a plethora of information not found in existing public cough datasets. Using the expert labels and participant metadata COUGHVID dataset can be used to train the classification models that can recognize a number of participant's covid status based on their cough audio samples. Overall, the dataset contains samples from participants having different gender, age, covid status, existing respiratory illnesses, and geographical locations, which should help machine learning models generalize them successfully.

Virufy is the crowd sourced dataset of cough audio which was collected in a hospital with patient's prior consent. While getting audio the physicians were there under highly supervision following Standard Operating Procedures. After collection the cough audio samples preprocessed and labeled with covid status which is taken from PCR testing, along with patient's basic information such as age, gender, medical history, etc. Hense it is the most accurate data. Here, we have provided total 121 prelabelled cough audio from 16 patients.

6 Data Preprocessing

The audio used in this approach is the combination of two datasets. Samples have been taken from the Virufy dataset and the Coughvid dataset. The preprocessing of the data was done manually using the Audacity software.

Audacity is an easy-to-use, multi-track audio editor which is open-source and compatible with the operating systems like Windows, macOS, Linux, etc. It may be used for recording live audio, edit wav and other audio formats, pre-processing of audio, including effects like normalizing, cropping, and fading in or fading out, as addition to recording audio from numerous sources.

The audio is loaded in the Audacity app and converted to the wav format. After this conversion is done, the noise from the audio recording is removed manually. This is followed up by removing the periods of silence from the audio. This step completes the preprocessing of the audio file and the cleaned processed audio recording can then be exported.

International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

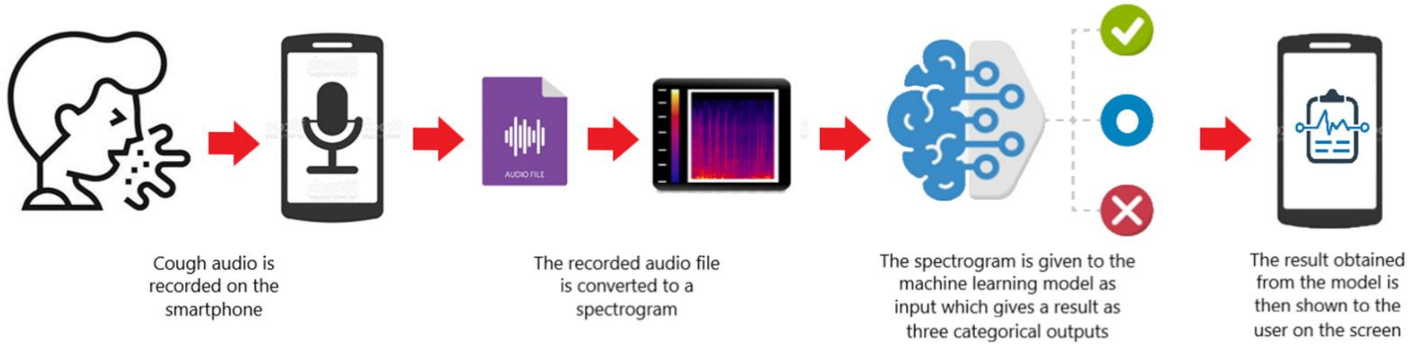

Fig. 2. Proposed Approach for COVID-19 testing

7 Spectrogram Generation

Generally, audio is represented in the form of a waveform which is basically a graph of amplitude vs time. In computer-memory, audio is represented as a time series of numbers, representing the amplitude at each timestep. For instance, if the sample rate was 25000, a one-second clip of audio would have 25000 numbers. Since the measurements are taken at fixed intervals of time, the data contains only the amplitude numbers and not the time values.A spectrogram, on the other hand, depicts the progression of a signal's frequencies over time. The amplitude is then shown in a third dimension with varying brightness or color.

Fourier Transforms are used to create spectrograms from sound signals. The amplitude of each frequency included in the signal is displayed using a Fourier Transform, which decomposes the signal into its constituent frequencies. A Spectrogram divides the duration of a sound stream into smaller time segments and uses the Fourier Transform to detect the frequencies contained within each segment. The Fourier Transforms for all of those segments are then combined into a single graphic. It shows Frequency (y-axis) vs. Time (x-axis), with different colors indicating the frequency's amplitude. The higher the signal's energy, the brighter the hue.

Unfortunately, there isn't much information to see when we display this spectrogram. This occurs as a result of how humans interpret sound. The majority of what we can hear is condensed into a small range of frequencies and amplitudes. Pitch refers to how we perceive sound frequencies. It's a personal assessment of the frequency. As a result, the frequency of a high-pitched sound is higher than that of a low-pitched sound. Frequencies are not perceived in a linear fashion by humans. Changes in lower frequencies are more noticeable to us than differences in higher frequencies. Rather than a linear scale, we hear them on a logarithmic scale.By performing experiments with a large number of listeners, the Mel Scale was established to take this into consideration. Listeners assess each unit to be equal in pitch distance from the next on this scale pf pitches. The loudness of a sound is a human impression of its amplitude. We hear loudness in a logarithmic rather than linear way, similar to how we hear frequency.

A Mel Spectrogram differs from a conventional spectrogram in that it plots Frequency vs. Time in two ways. Firstly, on the Y-axis, the Mel Scale is used instead of Frequency. Secondly, colors are indicated using the Decibel Scale rather tham the amplitude scale. We commonly use this instead of standard spectrogram for deep learning models. There are various libraries available for converting recorded audio into spectrogram. Librosa is one of the most commonly used library for this. Thus, we have also used Librosa to generate Mel Spectrograms.

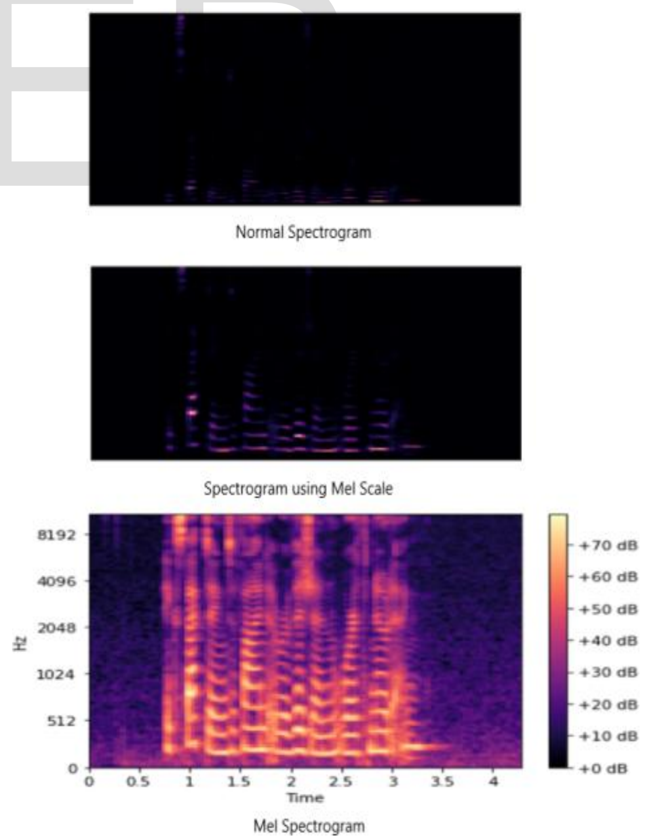

Fig. 3. Difference between Normal Spectrogram and Mel Spectrogram

International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

8 Model Architecture

In order to implement our proposed system, we have tried various machine learning models. After this trial-and-error approach we narrowed down to using the CNN model illustrated in Figure 4 which has 10 layers as it resulted in giving us the highest accuracy in comparison to the other models that we tried to integrate.

CNNs are a class of Deep Neural Networks that can recognize and classify particular features from images and are widely used for analyzing visual images. Their applications range from image and video recognition, image classification, medical image analysis, computer vision andnatural language processing.

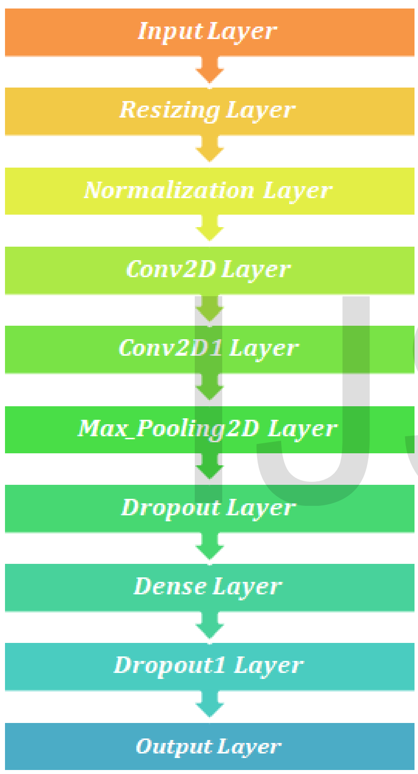

Fig. 4. CNN Architecture

8.1 Convolutional Layer

The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. Convolutional Networks need not be limited to only one Convolutional Layer. The first ConvLayer is responsible for capturing the Low-Level features such as edges, gradient orientation and color, etc. Further with added layers, the architecture adapts to the High-Level features as well, which gives us a network, having the wholesome understanding of images in the dataset, similar to how we would. We have used 2 convolutional layers in our implementation.

8.2 Pooling Layer

Our implementation has one max pooling layersimilar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model. Max Pooling and Average Pooling are the two types of pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel where the Average Pooling returns the average of all the values from the portion of the image covered by the Kernel.Max Pooling also performs as a Noise Suppressant. It discards the noise in activation and also performs denoising along with dimensionality reduction.

8.3 Dropout Layer

The Dropout layer is a mask that nullifies some neurons' contributions to the following layer while leaving all others unchanged. A Dropout layer can be applied to the input vector, nullifying some of its properties; however, it can also be applied to a hidden layer, nullifying some hidden neurons.Dropout layers are critical in CNN training because they prevent the training data from being overfit. If they aren't there, the first batch of training data has a disproportionately large impact on learning. As a result, learning of traits that occur only in later samples or batches would be prevented. We have used two dropout layers in this implementation.

8.4 Dense Layer

A dense layer is one that is deeply connected to the layer before it, meaning that each of the layer's neurons is connected to every neuron in the layer before it. In artificial neural network networks, this layer is the most widely utilized layer. If the preceding layer outputs a (MxN) matrix by combining results from every neuron, the output goes through the dense layer where the count of neurons in a dense layer should be N.

9 Results

In our study, features obtained from the CNN model, in which Mel Spectrograms are given as input, are used for classification. Our model consists of 10 layers. It has been found that performance increases significantly because of feature selection in features. Our model achieved 90% accuracy and a validation accuracy of 80% on subjects diagnosed with clinical testing.

We also tried a VGG-19 model, it is a variant of VGG model which consists of 16 layers of Convolution, 3 layers of fully connected, 5 layers of MaxPool and 1 layers of SoftMax i.e. total 19 layers. These layers are pre-trained on ImageNet Dataset. The model was fed with images of mel-spectrograms in the train-test ratio of 80:20.The results were classified into positive, negative and symptomatic. We achieved 78 % accuracy from our testing.

International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

	precision	recall	f1-score	support
COVID	0.90	0.69	0.78	13
HEALTHY	0.67	0.89	0.76	9
SYMPTOMATIC	1.00	1.00	1.00	8
accuracy			0.83	30
macro avg	0.86	0.86	0.85	30
weighted avg	0.86	0.83	0.83	30

Fig. 5. Precision, Recall and f1-score of CNN Model

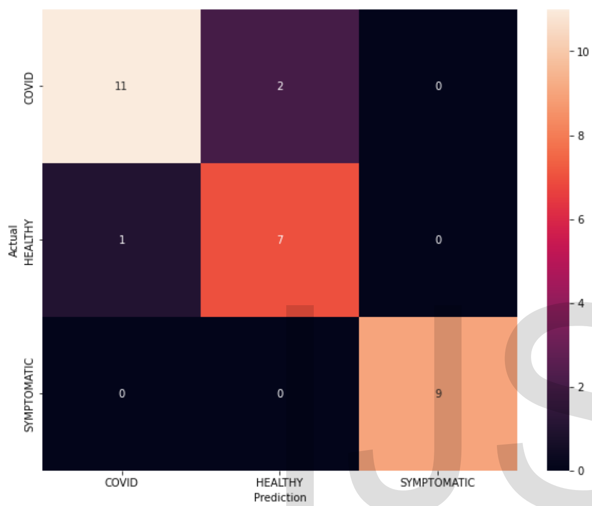

Fig. 6. Confusion Matrix

10 Conclusion

Lack of rapid testing, overcrowding at testing facilities, and cost are some of the major factors behind the rapid spread of the COVID-19 pandemic. To mitigate this, we propose a detection system in this paper for identifying COVID-19. It is a faster and safer alternative to traditional COVID-19 testing. This approach is cost-effective and thus it can be easily implemented and distributed and does not require any extra accessories which increases the ease with which it can be used. Additionally, this approach will also ensure that social distancing is maintained and will decrease overcrowding at testing facilities making it easier for people to know whether they need medical attention or not. The results showcase that the model can diagnose COVID-19 with considerable accuracy. Although the model has a good accuracy rate, it is not meant to compete or replace clinical testing. While we are working to improve our model we hope our system can tackle some of the issues and help in large-scale testing at places where the spread of Covid-19 is rampant.

Acknowledgment

It is our immense pleasure to express our gratitude to Ms. Megha V. Gupta as our Guide for providing us with constructive and positive feedbacks during the preparation of this project. We thank Ms. Arathi Kamble, our project coordinator, for providing all important project information. We express our gratitude to all other staff members of Computer Department for their valuable insights and co-operation. We express our sincere thanks to Dr. Sanjay Sharma, (Head of Computer Department) for his support and co-ordination. We would like to thank Dr. Prashant Deshmukh (Principal), for his kind cooperation, guidance and for providing necessary facilities during the working on this report. Last but not least, we are thankful to our friends whose encouragement and suggestions helped us to enhance our ongoing B.E Project. We are also thankful to our parents for their constant support and best wishes.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[2] YunusEmreErdoğan and Ali Narin (2021) COVID-19 detection with traditional and deep features on cough acoustic signals. doi: 10.1016/j.compbiomed.2021.104765.

[3] Martin Ekman, Antonio Lentini, et al. (2020) Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-PCR. doi: 10.1038/s41467-020-18611-5.

[4] Orlandic L, Teijeiro T, Atienza D (2020) The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale coughs analysis algorithms. arXiv: 2009.11644 . doi: 10.5281/zenodo.4048312.

[5] Chaudhari G, Jiang X, Fakhry A, et al. (2020) Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. arXiv: 2011.13320.

[6] Bader M, Shahin I, Hassan A (2020) Studying the similarity of COVID-19 sounds based on correlation analysis of MFCC. arXiv: 2010.08770 . doi: 10.1109/CCCI49893.2020.9256700.

[7] Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI-enabled preliminary diagnosis for COVID-19 from cough samples via an app (2020). doi: 10.1016/j.imu.2020.100378.

[8] Laguarta J, Hueto F, Subirana B COVID-19 artificial intelligence diagnosis using only cough recordings (2020). doi:10.1109/OJEMB.2020.3026928.

International Journal of Scientific & Engineering Research, Volume 13, Issue 4, April-2022
ISSN 2229-5518

AUTHOR DETAILS

Apoorva Gondane
Dept. Computer Engineering
New Horizon Institute of Technology and Management, Thane, India
apoorvagondane182@nhitm.ac.in

Bhushan Birwadkar
Dept. Computer Engineering
New Horizon Institute of Technology and Management, Thane, India
bhushanbirwadkar182@nhitm.ac.in

Prajwal Aher
Dept. Computer Engineering
New Horizon Institute of Technology and Management, Thane, India
prajwalaher182@nhitm.ac.in

Prof. Megha Gupta
Dept. Computer Engineering
New Horizon Institute of Technology and Management, Thane, India
meghagupta@nhitm.ac.in