



# Cough Audio Analysis for COVID-19 Diagnosis

Teghdeep Kapoor<sup>1</sup> · Tanya Pandhi<sup>1</sup> · Bharat Gupta<sup>2</sup>

Received: 10 March 2022 / Accepted: 23 November 2022 / Published online: 26 December 2022  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2022

## Abstract

Humanity has suffered catastrophically due to the COVID-19 pandemic. One of the most reliable diagnoses of COVID-19 is RT-PCR (reverse-transcription polymer chain reaction) testing. This method, however, has its limitations. It is time consuming and requires scalability. This research work carries out a preliminary prognosis of COVID-19, which is scalable and less time consuming. The research carried out a competitive analysis of four machine-learning models namely, Multilayer Perceptron, Convolutional Neural Networks, Recurrent Neural Networks with Long Short-Term Memory, and VGG-19 with Support Vector Machines. Out of these models, Multilayer Perceptron outperformed with higher specificity of 94.5% and accuracy of 96.8%. The results show that Multilayer Perceptron was able to distinguish between positive and negative COVID-19 coughs by a robust feature embedding technique.

**Keywords** Cough diagnosis · Deep learning · COVID-19 · COVID-19 preliminary diagnosis · CNN · MLP · RNN · LSTM · SVM · Machine learning

## Introduction

Coronavirus is a significant virus that causes illness in both animals and humans. It is a family of RNA viruses that is medium-sized and has a viral RNA genome largest of all known. A new, so far unknown coronavirus, SARS-CoV-2, the cause of COVID-19 disease, belongs to the same subgroup as MERS CoV and SARS-CoV. Coronavirus is known

to common people as COVID-19 was declared a pandemic by WHO (The World Health Organization), on March 11, 2020 [1]. It has forced the world into a mandatory lockdown.

The spread of this virus in the human race has caused 3.35 M deaths in the world as of May 2021 and has brought the economy to a standstill. It has also introduced several challenges worldwide. To date, the mode of transmission of SARS-CoV-2 is unresolved and is a topic of debate among researchers. Most researchers believe that it might be identical to SARS, which transmits through in-person contact or unsanitized surroundings in the form of aerosols and droplets. Studies have accentuated that patients with pulmonary symptoms are at higher risk of transmission [2, 3]. However, studies proved that it is also viable in asymptomatic patients [4]. Therefore, concluding that COVID-19 can spread via symptomatic as well as asymptomatic patients. The major task in fighting COVID-19 in most countries is to find asymptomatic patients who might be potential carriers of coronavirus. Currently, widely used methods for the diagnosis of COVID-19 are RT-PCR (reverse transcription-polymerase chain reaction) and X-ray or CT scans. Since X-rays require a chest scan at a well-equipped medical facility and are quite expensive, RT-PCR is more widely accepted. However, according to the study, this testing is not scalable and sometimes inaccurate [5]. It is also costly, and most countries have faced

---

Teghdeep Kapoor and Tanya Pandhi contributed equally to this work.

---

This article is part of the topical collection “Advances in Applied Image Processing and Pattern Recognition” guest edited by K C Santosh.

---

✉ Teghdeep Kapoor  
teghdeep@gmail.com

Tanya Pandhi  
pandhitanya@gmail.com

Bharat Gupta  
bharat.gupta@jiit.ac.in

<sup>1</sup> Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Sector-62, Noida, Uttar Pradesh 201309, India

<sup>2</sup> Faculty of Computer Science and Engineering, Jaypee Institute of Information Technology, Sector-62, Noida, Uttar Pradesh 201309, India

difficulties buying more test kits. Thus, in near future, there would be a need for an alternate testing method that is simpler, unintrusive, lab-free, and less expensive. Such a method should address all the limitations of current preliminary diagnostic techniques. It must also be based on sound science and identify at-risk individuals effectively.

This research proposes a solution which is a deep neural network that recognizes the differences between COVID-19 positive and negative coughs using audio classification techniques. It takes input as raw audio files and provides a diagnosis of whether that cough comes from a COVID-infected individual. More precisely the contributions of this research paper are as follows:

1. It provides a pre-screening tool for diagnosis of COVID-19 based on deep learning (AI) ubiquitously available to everyone. Its low cost, rapid results, and ease of access to everyone makes it a unique solution that can be employed in offices and various institutions as a pre-screening for entry. It can be used as an aiding tool to increase the diagnostic capability and devise a treatment plan in areas where adequate supplies, healthcare facilities, and medical professionals are not available.
2. We increased our dataset up to 5 times by leveraging data augmentation techniques on the open-source cough audio data set by virufy. Thus, illustrating a potential way to overcome the problem of overfitting in machine learning models due to a shortage of datasets.
3. The research uses features extracted from samples using sound processing techniques. The research constructed four models using two main approaches i.e., the Time Series waveform approach and the Amplitude waveform approach. In the time-series waveform approach, we extracted MFCC's which were fed to MLP, CNN, and RNN with LSTM. Whereas in the amplitude waveform approach we extracted the features from the flattened layer of VGG-19 which were then fed to SVM. Results show that out of all these four models, MLP was most successful in classifying the COVID-19 positive and negative cough with an accuracy of 96%. This shows that the time-series waveform approach was able to learn the robust features and was able to generalize classification better in comparison to the Amplitude waveform approach.
4. Was able to successfully fine-tune multi-layer perceptron to such an extent that it outperformed some of the existing literature [6, 7].
5. Portraying several future directions for our analysis and voice-based diagnosis in the context of COVID-19, which could open the door to pre-screening of COVID-19 and tracking the impact of COVID-19.

## Background

The primary reason behind the intractability of COVID-19 is that there is a significant delay between infection and diagnosis. Two main types of COVID-19 diagnostic techniques: laboratory-based testing and radiography testing.

### Laboratory-Based Testing

Laboratory testing can further be categorized into two kinds: immunoassays and nucleic acid or molecular tests. Immunoassay tests discern virus-associated proteins whereas nucleic acid tests or molecular tests discern the genetic code of the virus. In comparison to Immunoassay tests, nucleic acid tests are sensitive to early detection and for that reason, they are widely being used during this pandemic. The above tests often depend upon classical technologies one of which is RTPCR (reverse transcription-polymerase chain reaction) [8]. To perform laboratory-based testing samples were obtained with throat swabs, nasopharyngeal swabs, deep airway material, or sputum.

Even though this technique is quite sensitive in the early detection of COVID-19, however, there are certain limitations to this technique:

1. Geographical and temporal factors limit the availability of testing in various countries.
2. To fulfill the massive time-sensitive demand, it leads to a scarcity of clinical testing and increases their cost.
3. The need for a personal visit to the medical facility. Such a visit exposes many segments of the community to coronavirus. This can be a major obstacle, according to the study, the aerosol stability of COVID-19 ranges from three hours up to one week on different planes making it highly stable and hence contagious [9].
4. Many reputed newspapers recently highlighted that the turnaround time stretched to 6–7 working days in a few countries due to laboratories being overflowed with COVID tests. As a result, the virus might have already been transmitted to many, by the time a patient is diagnosed and his treatment starts [10, 11].
5. Often medical staff are at higher risk of infection due to these in-person testing techniques. Failure to secure our physicians can further lead to biomedical shortages and increase stress on the already distressed paramedical staff.
6. To protect others from potential exposure, many countries like India have also approved at-home sample collection under the guidelines of ICMR [12]. However, once a patient collects a nasal sample, they need to put

it in a saline solution and ship it overnight to a certified lab authorized to run specific tests on the kit. Hence, this approach also introduces delays and could compromise the quality of samples if the sample is stored for too long.

## Radiography Testing

Experts urge that we need more and faster testing to control the coronavirus and many have suggested that Artificial Intelligence (AI) is the solution. According to the study, multiple diagnoses of COVID-19 in development use AI to quickly analyze X-ray or CT scans have shown that in comparison to laboratorial tests, radiographic tests provide sharpened sensitivity [13, 14]. To manage coronavirus, a Thoracic CT scan—an optional imaging modality—can play a crucial role. This type of CT scan is an important aspect of COVID-19 diagnosis as it has higher precision. To produce high-resolution medical images, firstly X-rays from the patient's thorax cavity are picked up by the radiation detection tools, further, the radiographs generated are remodeled to form the medical images. One should look out for certain patterns in the thorax cavity, which might reveal different symptoms. This is examined by a radiographer, or when integrated with the AI-based analysis of the image, may detect COVID-19 with much higher specificity. This might be more efficient than that of a laboratorial test such as rRT-PCR. Promising results were shown by the study, it was calculated at a 95% confidence interval, having high precision and lower recall of 94 and 37% respectively for a diagnostic test based on radiology [15]. However, these techniques require scanning the chest in a well-equipped and expensive medical laboratory. So, indirectly this method also does not solve the problems faced by office-based tests as accentuated above.

## Cough-Based Testing

Many kinds of research, have been carried out, where various prognostic tools for examination of respiratory infections have been presented which are self-regulating [16–18]. They have used various deep neural networks such as Convolutional Neural Networks (CNNs) to recognize coughs within natural noise and to determine various diseases such as Bronchitis, bronchiolitis, Asthma, COPD, etc. depending on their distinctive cough sound features. Although cough is a frequent medical symptom in many pulmonary diseases, the study has demonstrated that depending on different conditions and locations of the underlying irritants, cough from various pulmonary diseases has unique characteristics [7]. Many types of studies have been done, which show that changes in the character of a coughing sound can indicate conditions of

lung disease [19, 20]. Pathological situations arise as a result of certain conditions such as obstruction, restriction, and integrated patterns. Researchers have made numerous efforts to improve the mechanism of objective classification of coughing, to classify different respiratory infections. Isolation of the cough audio signal helps to distinguish between Covid-19 positive and negative cough based on these features. The analysis of recent neurological symptoms shown by COVID-19 patients developed a link between the brain and COVID-19. This led MIT researchers to evaluate their Alzheimer's biomarkers for COVID-19 diagnosis. To detect Covid-19 coughs, they primarily used vocal cord strength, lung performance, sentiment, and muscular degradation in the human body [21].

## Methodology

### COVID-19 Cough Dataset

In medical research, finding the right amount and standard data is a difficult task. The dataset used in this study was taken from various sources and combined, COVID 19 cough samples were taken from the virufy open-source audio dataset [22]. The dataset consists of 121 sound segments which are digital audio files in.mp3 format out of which 48 are COVID positive and 73 are negative. Within the dataset, out of three, two relevant discrete attributes for the respective domain were selected as shown in Table 1. The cough audio samples were converted from.mp3 format to.wav format. To ensure consistency all over the dataset, preprocessing of three major sound properties(Audio Channels, Sample Rate, and Bit-depth) was done. The audio channels of the cough samples were integrated into mono channels and the sample rates were modified to the default sample rate of 22.05 kHz. In addition to this, to remove the discrepancy in bit depth, the value of each audio file's average amplitude was called down to a range between  $-1$  and  $1$ .

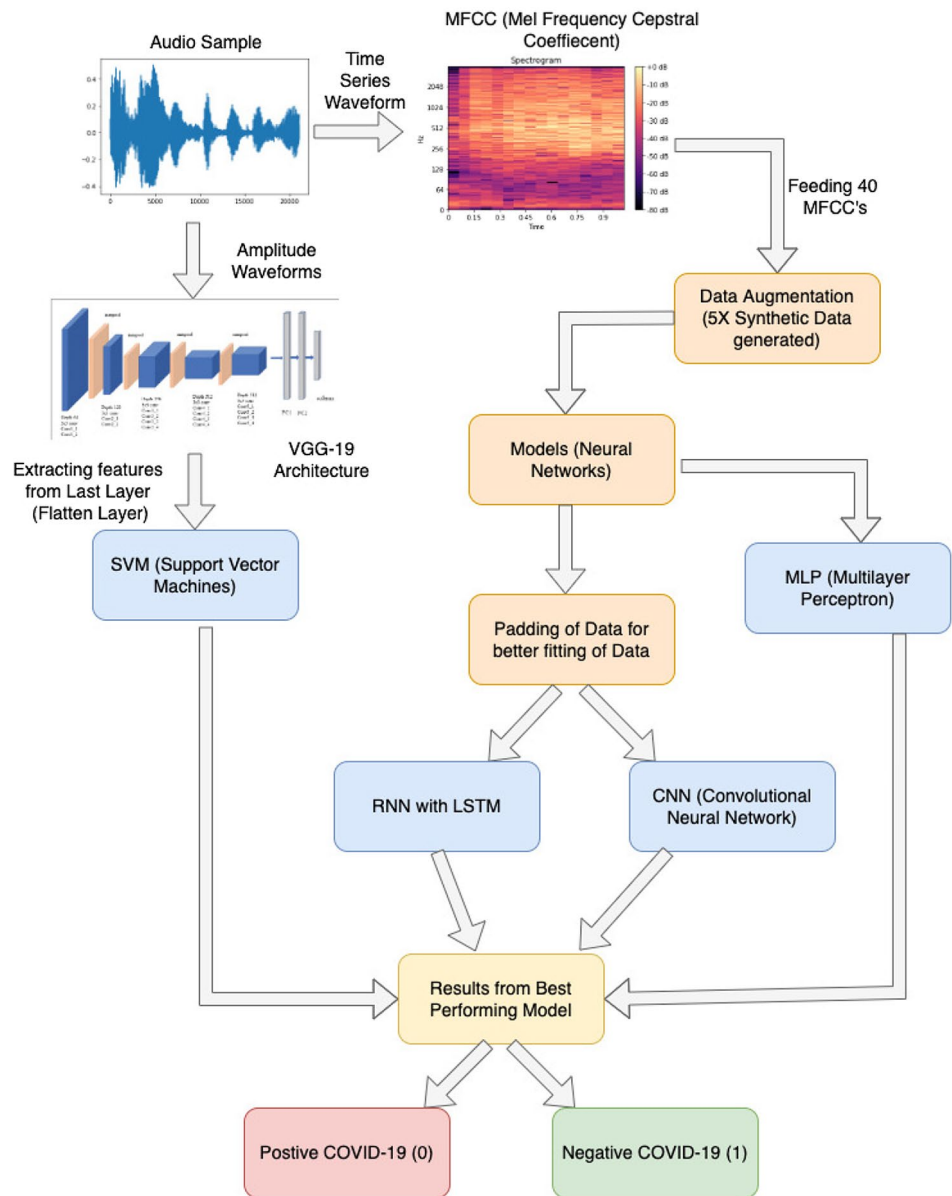
## Proposed Architecture

The architecture is as follows (Fig. 1)

**Table 1** Selected attribute list from the dataset

Attributes	Description	Selected
Patient's gender	Male or female	No
Cough audio sample	Path of the audio file	Yes
COVID status	Positive or negative	Yes

Fig. 1 Proposed architecture



**Data Augmentation**

Some domains have limited access to large data, such as medical image analysis or biomedical audio analysis. As a result, the dataset is not readily available and is quite small in size. This can lead to a problem known as overfitting. Overfitting refers to an event in which a network masters a function with very high variations to the maximum level at which it degrades the performance of the model on unseen data. One of the methods to resolve this problem is data augmentation.

Data Augmentation includes many strategies that improve the diversity and quality of data available for training models so that deep learning models can be built on it without facing the problem of overfitting. Audio augmentation

algorithms are used to generate synthetic audio data. In this study noise injection, shifting time, changing pitch, and speed was applied to the dataset using librosa (library for Recognition and Organization of Speech and Audio). This provides an easy way to manipulate pitch and speed while a Numpy python package was used to handle noise injection and shifting time. As a result, we were able to increase the dataset by 5 folds.

**Feature Extraction**

Past studies have shown that the acoustic of cough sounds may carry important information related to diseases [17]. For extracting these features, in this study two approaches are used. The first one is by extracting MFCC (Mel Frequency

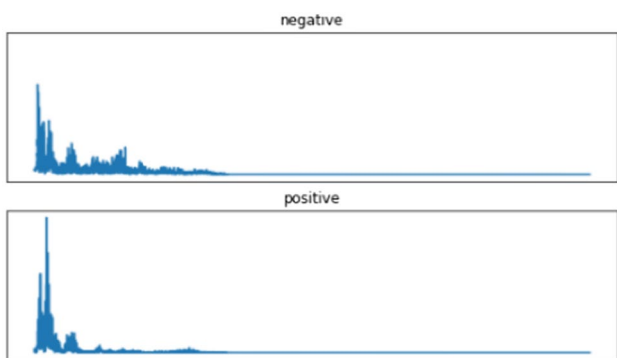


Fig. 2 Fourier transformation of negative and positive cough samples

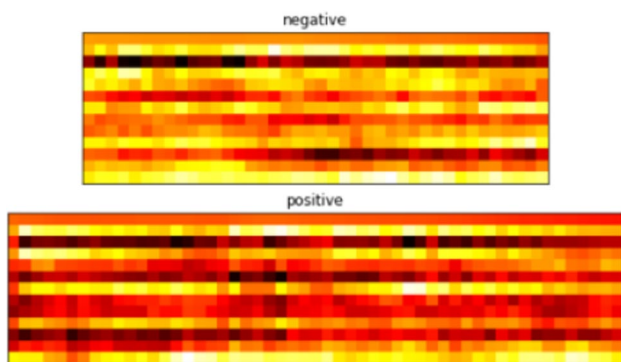


Fig. 3 Mel frequency Cepstrum of negative and positive cough samples

Cepstral Coefficient) from Audio Samples. It has been scientifically proven that humans are more efficacious at identifying minute changes in a speech at lower frequencies. Thus, to leverage this property one can use MFCC’s i.e., Mel frequency cepstrum coefficients. The MFCC converts the standard frequency to the Mel Scale using Eq. (1). It takes into account the human perceptiveness for sensitivity at appropriate frequencies and is therefore suitable for audio classification and sound processing. Mel scale equation is given below:

$$\text{Mel}(f) = 2595 \log(1 + (f/700)) \tag{1}$$

An audio signal’s power spectrum, which is short-term, is represented using the Mel frequency cepstrum (MFC). The first step for obtaining MFC is Fourier transformation. On taking the log of the magnitude of this Fourier spectrum as shown in Fig. 2, and then performing cosine transformation to obtain the spectrum of this log, we observe a crest wherever there is a periodical element in the original time signal [23]. MFCC’s are emanated by the cepstrum visualization of sound samples. They are coefficients that altogether form the

MFC. The study used the librosa python package to calculate a series of 40 MFCCs for each sample as shown in Fig. 3 and stored it in a pandas data frame.

The second approach was extracting important features from the last flattened layer of the VGG-19 model. After that, constructing the VGG-19 model, ImageNet images of size 64\*64 were fed for pre-training. After this, the NumPy array of pixel values was created by converting the PIL image object. Next, with dimensions of [samples, rows, columns, channels], it was expanded to the 4D array from the same 3D array. According to the VGG19 model, pixel values need to be changed. After this, all we need to do is to extract features.

In the VGG19 model, the last layer (1000-dimensional) is removed and the flattened layer results in a 4096-dimensional feature vector representation of an input image. After extracting these features, a 60–40 train test split was performed and then fed into the models.

### Model Architecture

Since the introduction of Neural Networks (NN) for pattern recognition, they have outperformed the results obtained with traditional algorithms. For instance, in the system for urban sound classification conducted, the performance of an SVM was compared with different configurations of neural models like a Deep Neural Network (DNN) a Recurrent Neural Network (RNN), and a Convolutional Neural Network (CNN), obtaining better results using a CNN or a DNN than using an SVM or an RNN [24]. Keeping this in mind, this research used 3 different configurations of neural network and SVM. In the end, the results of each model were compared and the best model was chosen.

### Multilayer Perceptron

Multilayer Perceptrons, or MLP for short, is a long-established neural network. A combination of multiple neurons forms a multilayer perceptron. The feeding of data takes place at the input layer which is then processed by the hidden layers. These hidden layers are used to increase the level of abstraction. After the processing of data from the hidden layers, the output layer gives us the final predictions. The study used Data Augmentation(noise, shift, and stretch) to increase the audio dataset to overcome overfitting. MLP can be constructed using Keras and Tensorflow backend. The model built in this research was sequential and consisted of four layers to increase the level of abstraction. All the four layers - the input layer, two hidden layers, and an output layer are of dense type, which is the standard type in most cases. The number of nodes comprised by each of the three layers

including input and hidden layers was 256, 128, and 64 respectively with an activation function ReLU and a dropout value of 25%. ReLU has proven to perform extremely well with neural network frameworks, it is explained furthermore in Appendix A.2. For better generalization in models, dropout is used which randomly excludes nodes from each epoch which in turn decreases the chance of overfitting. Finally, the output layer has 2 nodes which indicate the number of class labels with softmax. Softmax is the activation function used in the output layer, explained further in Appendix A.1. Softmax transforms the results in the form of probabilities, due to which it is highly used with various machine learning models. The model then, based on the highest probability, classifies the cough into COVID-19 positive or negative.

### Convolutional Neural Networks

Another Deep Learning algorithm implemented in this study is Convolutional Neural Network (CNN). It can take an image as input, allot significance to the various elements in the image, and be able to distinguish one from the other. As a precautionary measure, each recording of the input cough, processed with the MFCC package, was divided into 6-second audio clips and was padded as required. The study used the Convolutional Neural Network again with Keras and TensorFlow as a backend. It is a sequential model that comprises four Conv2D convolution layers out of which two are dense. A pooling layer of the MaxPooling2D type is linked with the final convolutional layer. The pooling layer reduces the parameters as well as the requirements for subsequent computation. This in turn reduced the dimensionality of the model. As a result, it shortens the duration of the training and reduces overfitting. The Max Pooling version has taken the greatest size possible for every window. For convolutional layers, the ReLU activation function was used, it is explained furthermore in Appendix A.2. A dropout value of 50% after the final convolutional layer is applied. The output layer has 2 nodes (number of labels, positive and negative) which are the same as the number of possible classifications. Softmax as the activation function used in the output layer explained further in Appendix A.1. Softmax transforms the results in the form of probabilities, due to which it is highly used with various machine learning models. The model then, based on the highest probability, classifies the cough into COVID-19 positive or negative.

### Recurrent Neural Networks with Long Short-Term Memory

A recurrent neural network (RNN) is a category of neural networks that help in data sequencing. Based on feedforward networks, RNNs show a similar mechanism of action as in

the human brain. To put it simply, there is no alternative algorithm that can produce predictable results in sequential data as accurately as a recurrent neural network can. The model used a sequential model, consisting of two LSTM layers, with four Time distributed layers. All LSTM layers consisted of 128 nodes. After the final LSTM layer, we used a Dropout of the value of 50%. The model has four Time Distributed Layers of dense type with 64, 32, 16, and 8 nodes respectively with an activation function as ReLU (Rectified Linear Activation), it is explained furthermore in Appendix A.2. The output layer has 2 nodes (number of labels, positive and negative) which are the same as the number of possible classifications. Softmax as the activation function used in the output layer, explained further in Appendix A.1. Softmax transforms the results in the form of probabilities, due to which it is highly used with various machine learning models. The model then, based on the highest probability, classifies the cough into COVID-19 positive or negative.

### Support Vector Machines

Support vector machines or also known as SVM, come under the category of data mining techniques that are used for both classification and prediction. It can generalize between two different classes. After providing the SVM model set of labeled training data for every category, it can classify the new text by checking the hyperplane that can distinguish between the two classes. After extracting features from the VGG-19 flatten layer as explained in “[Recurrent Neural Networks with Long Short-Term Memory](#)”. A 70–30 train test split was performed and then fed into a LinearSVM for classification.

## Results

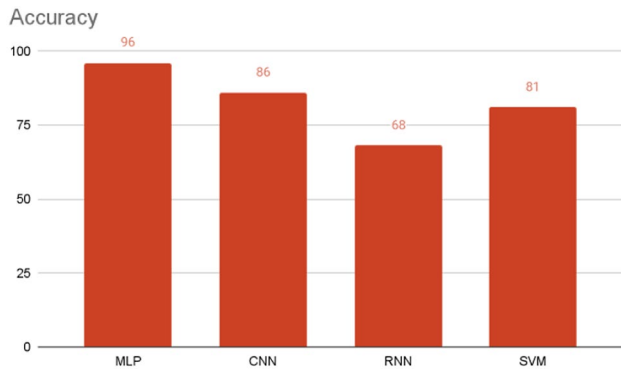
Predictions generated by models were expected to generalize well and could effectively produce the appropriate category label or data classification of previously unknown data. The effectiveness of the classification model was assessed based on the number of precise and false predictions observed by various models implemented on the unseen database. Accuracy, precision, and recall were the three evaluation metrics used which assess the nature of predictions made by the machine learning models developed in this research (Table 2).

### Accuracy

Accuracy is a measurement of the approximate level of quantity rather than the actual value of a quantity. It can

**Table 2** Overall accuracy is achieved by the models in tabular representation

Models	Accuracy (in %)
MLP	96
CNN	86
RNN	68
SVM	81



**Fig. 4** Overall accuracy is achieved by the models in graphical representation. Where T.P. stands for True Positive, T.N. stands for True Negative, F.P. stands for False Positive and F.N. stands for False Negative.

be computed from the confusion matrix using the equation mentioned below.

$$\text{Accuracy} = \frac{\text{T.P.} + \text{T.N.}}{\text{T.P.} + \text{T.N.} + \text{F.P.} + \text{F.N.}} \tag{2}$$

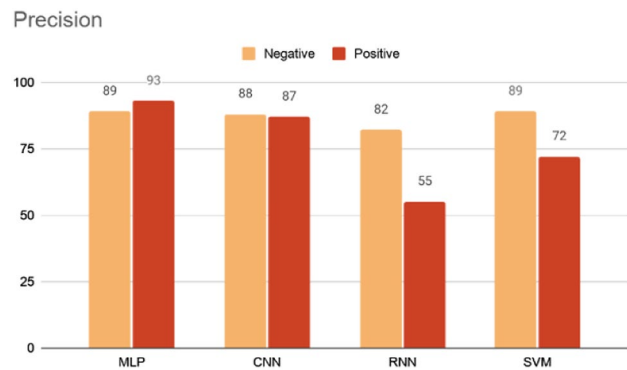
The Fig. 4 shows that Multilayer Perceptron and Convolutional Neural Network performed better than the rest of the models with an overall accuracy of 96 and 86% respectively. SVM performed fairly decent with 81% accuracy whereas Recurrent Neural Network was not able to generalize well and had an accuracy of only 68%.

**Precision**

In pattern detection, data retrieval, and categorization (machine reading), precision is the ratio of relevant instances among the retrieved instances. Precision is also known as a positive predictive value. In this study, that would be the proportion of patients who were positively identified with

**Table 3** The precision achieved by the models in the tabular representation

Models	Precision (in %)	
	Positive	Negative
MLP	93	89
CNN	87	88
RNN	55	82
SVM	72	89



**Fig. 5** The precision is achieved by the models in graphical representation

COVID-19 in all patients who had it. It was computed using the equation given below.

$$\text{Precision} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.P.}} \tag{3}$$

Where T.P. stands for True Positive and F.P. stands for False Positive. The precision of each model achieved in both negative and positive classes in this study was recorded in Table 3.

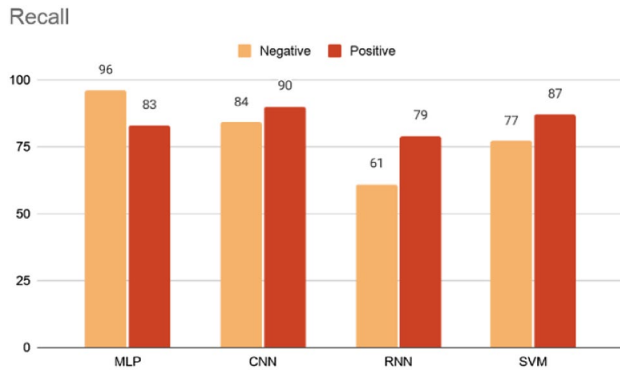
Higher Precision relates to lower false-positive rates. Figure 5 shows that Multilayer Perceptron and Convolutional Neural Networks have lower false-positive rates and can classify covid positive patients very well with a precision of 93 and 87% respectively. RNN has a higher false-positive rate and is prone to false alarms. All the models have a lower false-negative rate and can classify non-covid patients very well.

**Recall**

The recall is the measure of our model that accurately identifies true positives. It is also known as the sensitivity of

**Table 4** The recall achieved by the models in the tabular representation

Models	Recall (in %)	
	Positive	Negative
MLP	83	96
CNN	90	84
RNN	79	61
SVM	87	77



**Fig. 6** The recall is achieved by the models in graphical representation

the model. Therefore, in all patients with actual COVID-19, recall tells us how many did the model accurately identified as COVID-19 positive. It can be computed using the following equation:

$$\text{Recall} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.N.}} \tag{4}$$

Where T.P. stands for True Positive and F.N. stands for False negative. The recall of each model achieved in both negative and positive classes in this study was recorded in Table 4.

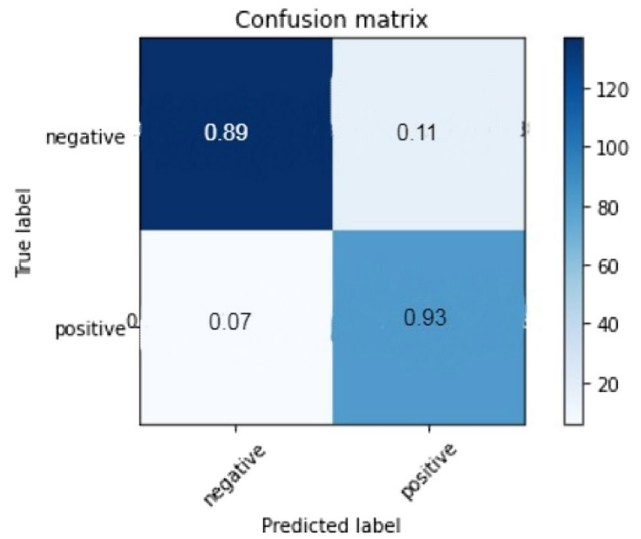
Higher recall relates to higher true positive rates. Figure 6 shows that Convolutional Neural Networks and Support Vector Machines have higher true positive rates for class positive. CNN and SVM correctly identify 90 and 87% of all the positive cases respectively. Multilayer Perceptron and Convolutional Neural Networks have a higher specificity. RNN can only identify 79% of all the positive cases and 61% of all the negative cases.

**Error Analysis**

This part will assist us in identifying and analysing misclassifications made by our top performing model, i.e. MLP-based classifier. This would allow us to improve the efficiency and

Normalized confusion matrix

$\begin{bmatrix} 0.89 & 0.11 \\ 0.07 & 0.93 \end{bmatrix}$



**Fig. 7** The confusion matrix of the multi layer perceptron

**Table 5** Cough audio classification model error distribution

Possible hypothesis for error	Count	% of total errors
Samples with noises	15	62.50
Female samples with low pitch being identified negative	2	8.33
Samples from elderly people	3	12.50
Speed augmented sound samples	4	16.67

accuracy of our model. The confusion matrix in the Fig. 7, displays all classification errors but provides no insights on inaccuracy. As a result, we separate the sound samples to ease our research. After briefly evaluating all samples, we discovered that the following appears to be prevalent across all faulty cases:

1. Audio samples with noises are not being classified properly.
2. Cough samples from females are not being classified accurately.
3. Cough from senior citizen are not being classified well.
4. Cough after augmentation with higher speed were not able to classified properly.



The hypothesis or observation of commonality for all incorrect use cases is provided in the Table 5, which maps the exact distribution of the mistakes.

## Comparing Related Research Works

In the period of last two years, there have been many kinds of research and methods proposed to detect COVID-19 via cough audio samples. However, from the list of works on this topic, we have hand-picked some of the research that focuses mainly on machine learning and deep learning techniques. An SVM model was trained on 16 clinical samples with a specific demographic which were labeled as COVID-19 positive and negative. The SVM and MFCC method reported an accuracy of 95.86% with 98.6 and 91.7% sensitivity and specificity, respectively [25]. Similarly, in [26], they proposed an SVM-based model with features like COMPARE, and eGeMAPS, generating an accuracy of 69%. A deep neural network (DNN), was proposed in [27] that utilized three different types of feature vectors that automatically detect COVID-19 cough sound samples with an overall accuracy of 89.2, 97.5, and 93.8% using time-domain, frequency-domain, and mixed-domain feature vectors, respectively.

This research [28] compares the performance of various machine learning classification algorithms on the dataset of 813 COVID-19 cough samples, out of which Random forest outperformed with an accuracy close to 90%. A similar research [29], developed a Random forest machine learning model with an accuracy of 66.74% using features like MFCC, Spectral bandwidth, and flatness. In [30], A CNN-based deep learning model was proposed which was trained on 545 cough samples of various demographic regions. This model was able to achieve an accuracy of 80%. In [31], the researchers developed an RNN deep learning model on three different datasets consisting of cough, breathing, and voice and achieving an accuracy of 98.2, 97, and 88.2% respectively.

The performance of the proposed system is compared with other related works available in the literature, as listed in Table 6. The comparison table shows that the proposed system performs at par with [27], achieving an accuracy of 96.4%. The Table 6 clearly demonstrates that the proposed system achieves a higher accuracy than the works published in [25, 26, 28–31].

## Conclusion

This paper presents Deep Learning Neural Networks for the initial diagnosis of COVID-19 with cough samples. For model evaluation, higher precision and lower recall

give an extremely accurate result, but it then misses a large number of difficult instances to classify which can't be ignored in COVID-19 diagnosis. Thus, there is a need for models having higher precision and higher recall at the same time for improved generalized classification. Based on the performance metric briefly discussed above, various models developed in this study were analyzed. This analysis revealed that the MLP-based classifier outperformed other Neural Network classifiers developed in this study. These results promise that AI can be used in the clinic and at home as a support system for physicians and the general public in the early detection of COVID-19. It may play an important role in medical diagnosis. This significant achievement supports extensive testing for COVID-19 even in areas where health facilities are not readily available. As a result, it helps to reduce the burden on paramedical staff.

The Trace, Test, and Treat strategy has shown that governments must be able to effectively track the spread of the disease, and isolate infected people. This helps in flattening the curve of infection successfully. However, most countries are not able to do enough rapid tests, which is why the alternative proposed can be very helpful.

## Appendix A

### A.1 Softmax

Softmax is a mathematical function that converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the relative scale of each value in the vector.

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (\text{A1})$$

where,  $\exp(x_i)$  represents standard exponential function for input vector,  $K$  represents number of classes in the multi-class classifier, and  $\exp(x_j)$  represents standard exponential function for output vector.

### A.2 ReLU

The rectified linear activation function or ReLU for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.

$$\text{ReLU}(x) = \max(0.0, x) \quad (\text{A2})$$

where  $x$  is the input to a neuron.

**Table 6** The performance comparison with existing work

Paper name	Dataset	Features used	Model used	Results
"Identifying COVID-19 using spectral analysis of cough recordings: a distinctive classification study [25]"	Virufy Dataset (121 sound samples)	MFCC	SVM	Accuracy—95.86%, Sensitivity—98.6%, and Specificity—91.7%
"An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety [26]"	Organic Dataset (52 sound samples)	COMPARE, and eGeMAPS	SVM	Avg. accuracy—69%, F1 score—eGeMAPS—0.65, ComPARE—0.66
"A study of using cough sounds and deep neural networks for the early detection of COVID-19 [27]"	Virufy Dataset (121 sound samples)	MFCC	DNN	Accuracy and F1 score for time domain, 89.2% and 0.889, for freq. domain, 97.5% and 0.974, for mixed domain, 93.8% and 0.973
"Automated detection of COVID-19 cough [28]"	Combination of Dataset (Virufy, Coswara, Univ. of Cambridge, and Univ. of Lleida)	Time frequency features—spectral energy, instantaneous frequency, instantaneous frequency peak and spectral information	RF	Accuracy—80%, Sensitivity—93.81%, F1 score—0.921, AUC—96.04%
"Coswara—a database of breathing, cough, and voice sounds for COVID-19 diagnosis [29]"	Coswara	Spectral contrast, MFCC, spectral roll-off, spectral centroid, Mean squareenergy	RF	Accuracy—66.74%
"Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data [6]"	Organic Data (crowdsourced—5634 samples)	RMS energy, spectral centroid, roll-off frequencies, zero-crossings, MFCC, duration, tempo onsets, period	CNN	Accuracy—80%, AUC—82%, precision—80%, and recall—72%
"Covid-19 detection system using recurrent neural networks [31]"	Organic Data (UAE Hospital—240 samples)	Spectral centroid, roll-off frequencies, zero-crossing, and MFCC	RNN	Accuracy and AUC of breathing is 98.2, 98.8%, for cough is 97, 97.4%, and for voice is 88.2, 84.4%
"Automatic lung health screening using respiratory sounds [32]"	Publicly available respiratory sounds dataset of size 6800+ clips	LPCC-based features	MLP	Accuracy—99.07%, AUC—99.2%
"COVID-19 artificial intelligence diagnosis using only cough recordings [33]"	MIT Open-source (4256 cough samples were used for training and 1064 for validation)	MFCC	CNN with bio-maker	Accuracy—97%, AUC—97%, sensitivity—98%, and specificity—94%
"Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath [34]"	Coswara	GFCC (Gamma-tone Frequency Cepstral Coefficients), and IMFCC (Improved Multi-frequency Cepstral Coefficients)	DCNN	Accuracy—92.91%, AUC—93.8%, sensitivity—91%, and specificity—96%
"Robust detection of COVID-19 in cough sounds [35]"	Vocaa.ai (1927 cough sound samples)	MFCC	Recur-rence dynamics and variable Markov model	Accuracy—91%, AUC—84%
Proposed method (time series waveform)	Virufy (121 sound samples) and 5 fold data augmentation	MFCC	MLP, CNN, and RNN with LSTM	Accuracy, avg. precision, and avg. recall for MLP, 96.8, 91, 89.5%, for CNN 86, 87.5, 87%, and for RNN with LSTM 68, 68.5, 70%

Table 6 (continued)

Paper name	Dataset	Features used	Model used	Results
Proposed method (amplitude series waveform)	Virufy (121 sound samples)	Transfer learning (extracting features from last layer of VGG-19)	SVM	Accuracy—81%, precision for COVID-19—72% and for non-COVID-19—89%, recall for COVID-19—87%, and for non-COVID-19—77%

**Declarations**

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**References**

- Laboratory testing for coronavirus disease (COVID-19) in suspected human cases. Last accessed 3 June 2021.
- Guo G, Ye L, Pan K, Chen Y, Xing D, Yan K, Chen Z, Ding N, Li W, Huang H, Zhang L, Li X, Xue X. New insights of emerging SARS-CoV-2: epidemiology, etiology, clinical features, clinical treatment, and prevention. *J Front Cell Dev Biol.* 2020;8:410.
- Yu IT, Li Y, Wong TW, Tam W, Chan AT, Lee JH, Leung DY, Ho T. Evidence of airborne transmission of the severe acute respiratory syndrome virus. *J N Engl J Med.* 2004;350(17):1731–9.
- Yang R, Gui X, Xiong Y. Patients with respiratory symptoms are at greater risk of COVID-19 transmission. *J Respir Med.* 2020;165: 105935.
- Kameswari S, Brundha MP, Ezhilarasan D. Advantages and disadvantages of RT- PCR in COVID 19. *Eur J Mol Clin Med.* 2020;7(1):1174–81.
- Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, Xia T, Cicuta P. Mascolo: exploring automatic diagnosis of Covid-19 from crowdsourced respiratory sound data. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 3474–84, 2020.
- Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, John CN, Hussain MI, Nabeel M. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *J Inform Med Unlock.* 2020;20.
- Waltz E. How do Coronavirus tests work? *IEEE Spectrum.* Accessed 2 June 2021.
- Van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, Tamin A, Harcourt JL, Thornburg NJ, Gerber SI, Lloyd-Smith JO: aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *J N Engl J Med.* 2020;382(16):1564–7.
- Delayed RT-PCR reports triggering Covid surge, high transmission rate in Lucknow. *The Times of India.* Last accessed 2 June 2021.
- Gujarat. Why RT-PCR test reports ‘delayed by 5-7 days’; AG says ‘many undergo tests unnecessarily’. *The Indian Express.* Last accessed 2 June 2021.
- Advisory for COVID-19 testing during the second wave of the pandemic. *ICMR Official Advisory.* Last accessed 2 June 2021.
- Ozsahin I, Sekeroglu B, Musa MS, Mustapha MT, Ozsahin DU. Review on diagnosis of COVID-19 from chest CT images using artificial intelligence. *J Comput Math Methods Med.* 2020;10
- Fang Y, Zhang H, Xie J. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *J Radiol.* 2020;296:E115–7.
- Adams HJ, Kwee TC, Kwee RM. COVID-19 and chest CT: do not put the sensitivity value in the isolation room and look beyond the numbers. *Radiology.* 2020:201709
- Bales C, Nabeel M, John CN, Masood U, Qureshi HN, Farooq H, Posokhova I, Imran A. Can machine learning be used to recognize and diagnose coughs? In: *2020 International Conference on e-Health and Bioengineering (EHB)*, vol. 29, p. 1–4, 2020.
- Amrulloh Y, Abeyratne U, Swarnkar V, Triasih R. Cough sound analysis for pneumonia and asthma classification in the pediatric population. In: *IEEE 6th International Conference on Intelligent Systems, Modelling, and Simulation*, p. 127–131, 2020.

18. Infante C, Chamberlain D, Fletcher R, Thorat Y, Kodgule R. Use of cough sounds for diagnosis and screening of pulmonary disease. In: IEEE Global Humanitarian Technology Conference, GHTC, p. 1–10, 2015.
19. Hirschberg J, Szende T. Pathological cry, stridor and cough in infants. Budapest: Akademiai Kiado; 1983.
20. Maryam Z, Fazel ZMH, Mostafa M. Application of intelligent systems in asthma disease: designing a fuzzy rule-based system for evaluating the based on level of asthma exacerbation. *J Med Syst.* 2012;36:2071–83.
21. Laguarda J, Hueto F, Subirana B. A COVID-19 artificial intelligence diagnosis using only cough recordings. In: IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, p. 275–281, 2020.
22. Khanzada A, Wilson T. Virufy COVID-19 Open Cough Dataset. Github, 2020. Last accessed 2 Feb 2021.
23. Nair P. The dummy's guide to MFCC. Medium, 2018. Last accessed 5 June 2021.
24. Chang C, Doran B. Urban sound classification: with Random Forest SVM DNN RNN and CNN classifiers. In: CSCI E-81 Machine Learning and Data Mining Final Project Fall 2016. Harvard University Cambridge; 2016.
25. Melek Manshouri N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. *Cognit Neurodyn.* 2021:1–15.
26. Han J, Qian K, Song M, Yang Z, Ren Z, Liu S, Liu J, Zheng H, Ji W, Koike T, Li X. An early study on intelligent analysis of speech under COVID-19: Severity, sleep quality, fatigue, and anxiety. 2020. arXiv preprint [arXiv:2005.00096](https://arxiv.org/abs/2005.00096).
27. Islam R, Abdel-Raheem E, Tarique M. A study of using cough sounds and deep neural networks for the early detection of Covid-19. *Biomed Eng Ad.* 2022:100025.
28. Tena A, Clariá F, Solsona F. Automated detection of COVID-19 cough. *Biomed Signal Process Control.* 2022;71: 103175.
29. Sharma N, Krishnan P, Kumar R, Ramoji S, Chetupalli SR, Ghosh PK, Ganapathy S. Coswara—a database of breathing, cough, and voice sounds for COVID-19 diagnosis. 2020. arXiv preprint [arXiv:2005.10548](https://arxiv.org/abs/2005.10548).
30. Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, Xia T, Cicuta P, Mascolo C. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. 2020. arXiv preprint [arXiv:2006.05919](https://arxiv.org/abs/2006.05919).
31. Hassan A, Shahin I, Alsabek MB. Covid-19 detection system using recurrent neural networks. In: 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI). IEEE; 2020. p. 1–5.
32. Mukherjee H, Sreerama P, Dhar A, Obaidullah S, Roy K, Mahmud M, Santosh KC. Automatic lung health screening using respiratory sounds. *J Med Syst.* 2021;45(2):1–9.
33. Laguarda J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open J Eng Med Biol.* 2020;1:275–81. <https://doi.org/10.1109/OJEMB.2020.3026928>.
34. Lella KK, Pja A. Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath. *Alex Eng J.* 2022;61(2):1319–34.
35. Mouawad P, Dubnov T, Dubnov S. Robust detection of COVID-19 in cough sounds. *SN Comput Sci.* 2021;2(1):1–13.
36. Santosh KC, Rasmussen N, Mamun M, Aryal S. A systematic review on cough sound analysis for Covid-19 diagnosis and screening: is my cough sound COVID-19? *PeerJ Comput Sci.* 2022;8: e958.
37. Hewage R. Extract features, visualize filters and feature maps in VGG16 and VGG19 CNN models, towards data science, 2020. Last accessed 5 June 2021.
38. Pandhi T, Kapoor T, Gupta B. An improved technique for preliminary diagnosis of COVID-19 via cough audio analysis. In: Santosh K, Hegadi R, Pal U, editors. Recent trends in image processing and pattern recognition. RTIP2R 2021. Communications in computer and information science, vol. 1576. Cham: Springer; 2022.
39. Santosh KC. Speech processing in healthcare: can we integrate? In: Intelligent speech signal processing. New York: Academic Press; 2019. p. 1–4.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.