INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY
IJARIIT
ISSN: 2454-132X
Impact Factor: 6.078
(Volume 8, Issue 3 - V8I3-1259)
Available online at: https://www.ijariit.com

COVID19 diagnosis using forced coughing and Artificial Intelligence

Vaibhav Dandage
vaibhavdandage@gmail.com
Padmshri VB Kolte College of Engineering, Malkapur, Maharashtra

ABSTRACT

Everyone knows COVID19. No part of the world is untouched by this disease today. The virus spread like wildfire throughout entire world within a year. The main causes of the spread of the virus we can say were incubation period of 11.5 days and it can spread through sneezing, touch and air. Along with that the limitations of testing time period added more scope for the virus spread. In this paper we have proposed a method based on artificial intelligence machine learning, which can be used by anyone to check whether the person has been infected by the COVID disease. The person just needs to provide audio recording of the forced coughing. Within few seconds the person will get the result. This method has a great potential of saving time, money and lives. The similar AI models have been proposed for different disease like pneumonia and Alzheimer.

Keywords: *Artificial Intelligence, COVID19, Corona Virus, AI Model, Machine Learning, MFCC*

1. INTRODUCTION

COVID 19 is caused by Corona virus. It is said that the spread started from bats to humans through food. Some people say it spread through it spread from Wuhan lab in China. Most of the people infected by the corona virus do not show any symptoms and they are called as asymptomatic. In different people show different symptoms like body ach, headache, pneumonia, light temperature. Death rate of the infected had been around 1% and there have been more than few million deaths in the entire world. COVID19 has been designated as pandemic. Virus is probably going to stay here permanently. While writing this report different parts of the world are experiencing waves with respect to different variants of the virus.

1.1 Spread of Corona Virus

The virus can spread very easily through sneezing by infected person, by touch to the objects which came in contact of infected person. The disease also can spread through water and food.

1.2 COVID19 Diagnostic Methods

Mainly below 2 methods are used to diagnose COVID19.
A] RT-PCR test: Which is also identified as molecular test, this test detects genetic material of the Corona virus by a technique called reverse transcription polymerase chain reaction (RT-PCR). A fluid from nostril of the patient is collected by a healthcare professional by inserting a long nasal swab.

B] Antigen test: By this test certain proteins in the virus are detected. Some antigen tests can produce results in minutes. Others may be sent to a lab for analysis. The test results are considered accurate when instructions are carefully followed. But there's also chance of false-negative results — meaning it's possible to be infected by the virus but have a negative result. Depending on the situation, it can be recommended a RT-PCR test to confirm a negative antigen test result.

1.3 Limitations of Traditional COVID19 Diagnosis Methods

Traditional methods are implemented by human and need considerable time for diagnosis. And, those methods need great amount of investment while considering on large scale population testing. There is high probability of spreading virus during manual testing as it involves direct contacts.So, the current method proposed in this paper will help to overcome it to great extent.

2. LITERATURE REVIEW

While carrying out this research we referred various literature available on this subject. As the subject is "COVID 19 diagnosis using forced coughing and artificial intelligence" we referred different literature available on all related aspects. There are different AI biomarkers available to identify different characteristics of different diseases like breathing biomarker, coughing biomarkers, vocal cords biomarker. There have been different AI models to classify different sounds.

There have been many paper publications on above mentioned topics.

There is one IEEE paper published by, Jordi Laguarta, Ferran Hueto, and Brian Subirana, on COVID diagnosis using machine learning AI model based upon Alzheimer's discriminator model which needs patients sounds as input to diagnose Alzheimer as shown in chart 1[1][7][9].

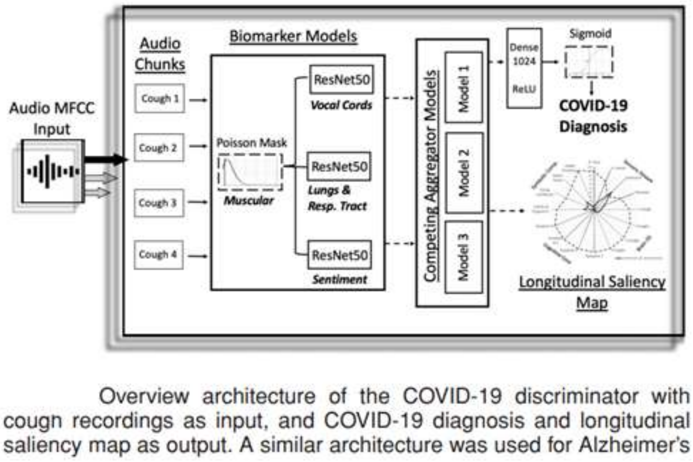

Fig 1: Overview Architecture COVID19 Discriminator Using MFCC Input

There is also another paper, published by Quan Zhou1,Jianhua Shan1,Wenlong Ding1, Chengyin Wang1, Shi Yuan1, Fuchun Sun2, Haiyuan Li3 and Bin Fang-Cough Recognition Based on Mel-Spectrogram & AI on website https://www.frontiersin.org/articles/10.3389/frobt.2021.580080/full[3].

There are many papers available and referred to carry out this research called as CORD19[6] Below table 2[3] shows different methods and their accuracy, which are used to diagnose COVID19 using Artificial Intelligence, Machine Learning and Coughing audio recording of the person.

3. METHODOLOGY

As shown below in Figure-1, the work-flow of our COVID-19 diagnosis using forced coughing and artificial intelligence is presented.

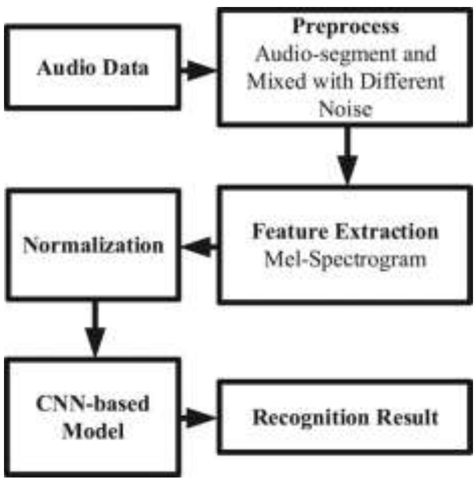

Fig 2: Flowchart of Process of COVID-19 diagnosis using forced coughing and artificial intelligence

We have taken audio data from IISc Bangalore Coswara project available publicly on GitHub. This audio data has been used to train our AI model. There is an algorithm proposed in this paper, which can be used to diagnose the patient from provided coughing audio recording as an input to AI model.

Each sound type produces different structure in Mel-Spectrogram as shown below. As each sound has different pattern in Mel-Spectrogram, we can use this to train our AI model, as shown in chart 2[3]. The features of Mel spectrogram will be used to train our AI model.

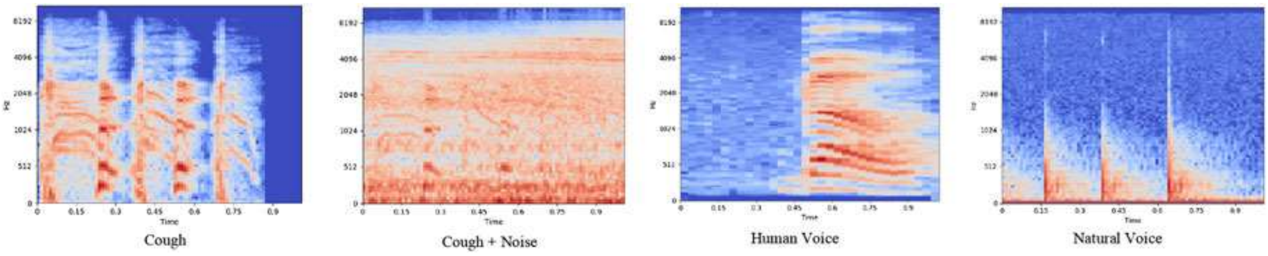

Fig 3: Mel Spectrogram of Different Sounds

As shown below in chart 3, the trained AI model can be used to classify different audio signal. There have been pre-proposed methods [7][8][9] for this type of audio classification.

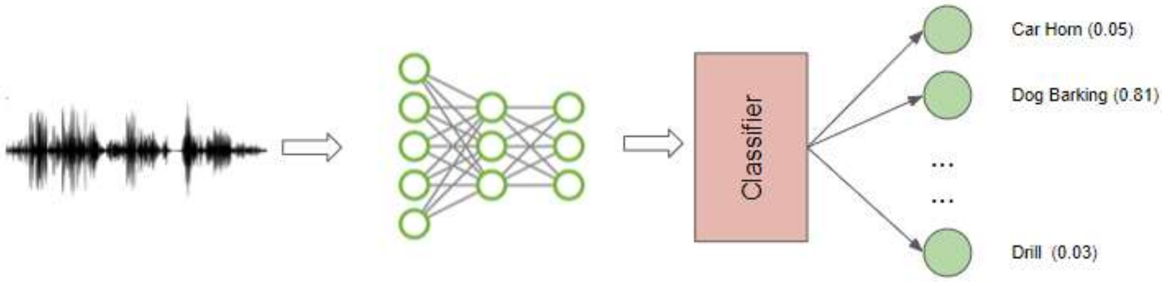

Fig 4: Flowchart to Classify Different Audio Recordings Using Artificial Intelligence

4. DATA AVAILABILITY & PREPROCESSING

The data used to carry out this research was taken from publicly available sources. Below are more details.

4.1 Data Availability

Data Availability There was a project started by IISc Bangalore India, called as Coswara Project [2]. For which they developed a website https://coswara.iisc.ac.in/ where people can submit their information like forced cough recordings, symptoms if they have any, breathing sound. Which is recorded and made publicly available on website https://github.com/iiscleap/Coswara-Data. We have used this freely publicly available data to train our AI model.

4.2 Data Preprocessing

Data Preprocessing The audio data needs to be preprocessed before providing as input to train AI model. So, we'll extract Mel Frequency Cepstral Coefficient features also called as MFCC features from each coughing audio file. Then use these MFCC features to train our AI model. In chart you can see how different types of sounds have different patterns in Mel-Spectrogram.

5. ALGORITHM

The research work has been carried out by using below algorithm.

Step 1: Extract MFCC's for every audio file available with forced coughing recording

mfccs = librosa.feature.mfcc(y=librosa_audio_data, sr=librosa_sample_rate, n_mfcc=40)

```
mfccs

array([[-467.07614  , -291.45007  , -226.93472  , ..., -368.057    ,
        -376.71964  , -380.64713  ],
       [  83.975426 ,  129.96451  ,  146.9947   , ...,   92.737656 ,
          85.98154  ,   83.28431  ],
       [  18.041386 ,  -21.119074 ,  -36.52266  , ...,  -43.74014  ,
         -42.21115  ,  -41.4625   ],
       ...,
       [   2.7200918 ,    4.382388 ,    3.3810616 , ...,    2.0435476 ,
          -5.698351  ,  -10.461082 ],
       [   4.539895  ,    1.3840971 ,   -5.2722774 , ...,   -2.3307931 ,
          -7.3472958 ,  -10.098457 ],
       [   2.0889082 ,   -2.376497  ,  -12.439749  , ...,   -1.4136873 ,
          -0.74377954,   -0.63850975]], dtype=float32)
```

Step 2: Converting extracted features to Panda's data frame

extracted_features_df=pd.DataFrame(extracted_features,columns=['feature','class'])

```
extracted_features_df
```

	feature	class
0	[-1131.371, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,...	positive_moderate
1	[-420.36105, 45.441746, -43.810143, 17.733166,...	positive_asymp
2	[-311.62018, 97.83663, -64.45676, 32.80387, -3...	positive_asymp
3	[-1131.371, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,...	positive_moderate
4	[-305.90717, 60.513542, -38.131687, 7.659088, ...	positive_moderate
...	...	...
96	[-1131.3708, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...	positive_moderate
97	[-1131.3708, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...	positive_asymp
98	[-288.2159, 90.11927, -32.00084, 19.935781, -1...	positive_moderate
99	[-464.63422, 67.42508, -16.147852, 0.35573447,...	positive_mild
100	[-405.6892, 62.791126, -27.14041, 2.4412513, -...	positive_moderate

101 rows × 2 columns

Step 3: Split the dataset into independent and dependent dataset
X=np.array(extracted_features_df['feature'].tolist())
y=np.array(extracted_features_df['class'].tolist())

Step 4: Train Test Split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)

Step 5: Training AI model
model=Sequential()
model.fit(X_train, y_train, batch_size=num_batch_size, epochs=num_epochs, validation_data=(X_test, y_test), callbacks=[checkpointer], verbose=1)
test_accuracy=model.evaluate(X_test,y_test,verbose=0)

Step 6: Provide audio input to trained AI model
filename="D:/aws/20210930/MBs72wy9MATqbupZbIaYYCzj7dg1/cough-heavy.wav"
prediction_feature=features_extractor(filename)

Step 7: AI model will display the COVID19 infection status
predict_x=model.predict(X_test)

6. RESULTS AND DISCUSSION

A] Accuracy

We were able to achieve 57.14285969734192 % accuracy using 100 input samples. In this project we were able to get the expected result for the data using which we trained the AI model.

Table-1: AI Model Results with AI Model Training Data

Sr. No.	Sample audio file name	Result with traditional methods	Result with our AI model
1	0Js6ZUZQ9NUnu568Fh7B6mZ1R8o1	positive_moderate	positive_moderate
2	0MVnLUuWMBdzmvRDgLvi4TRl4Zf1	positive_asymp	positive_asymp
3	1aPZz1jeoBcw9B046d504mn4cq82	positive_asymp	positive_asymp
4	1NQvmLMrJyTwrmbNwAm6wDT4wpz2	positive_moderate	positive_moderate
5	2lsppspH44WH3Scc1ojisHaXJzp1	positive_moderate	positive_moderate
6	2oEWhawfKmanWlvgQfkdkmgoJ9k1	positive_moderate	positive_moderate
7	2tjg2NfgJPaYdgTZRDUBhWm8F133	positive_mild	positive_mild
8	4RTWSyPKTdc7aRr4ocbrk5m153M2	positive_moderate	positive_moderate
9	4SjSzRW7o2P3lHyxHBfNRfooMun2	positive_moderate	positive_moderate
10	4Zg768BEWlXeVeUynZPu5jF1XOX2	no_resp_illness_exposed	no_resp_illness_exposed
11	5tVbqoDLs6PZKArvhcMTZcgxruo1	positive_mild	positive_mild
12	6aOGyUJRIcXSEWNRCkzUodJimux1	positive_mild	positive_mild

13	6WWaE2bjq1YgmugnDuBgA3kLEoL2	healthy	healthy
14	8Ul16g4L9nP9lcp16BKh1X6cfhb2	positive_moderate	positive_moderate
15	8Vz9jwmadkRfjbwrHwGZ2MsUtcz1	positive_mild	positive_mild
16	9sAOoNazDOXPars5nzlLzOwuYVW2	positive_mild	positive_mild

While training the model we observed that as we increase the training with large numbers of epochs with training input data the model accuracy increases. We were able to achieve 57.14285969734192 % accuracy with 100 epochs. If we increase the count of input samples the accuracy can be easily increased to 80%.

Below graph shows how the test data loss will be minimized as we increase the number of epochs.

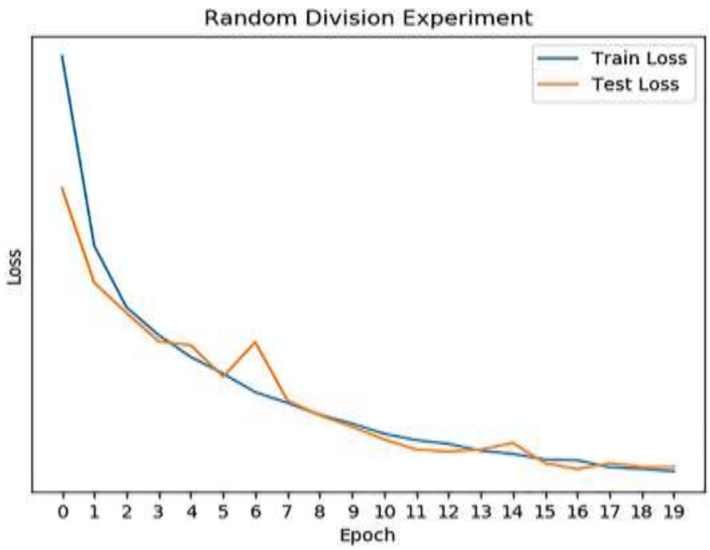

Fig 5: Loss Inversely Proportional to Epochs

B] Comparison with other models

In this project we created an artificial intelligence model which can be used to diagnose COVID19 patients by using their forced cough recordings. Although this methos will provide very good results there will be scope to add many more extensive measures for the accuracy improvement. Table 2[3] shows different methods and their accuracy which exists today.

Table-2[3]: Comparison with Other Models

Methods	Random division recognition task				No-leakage division recognition task			
	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)
Mel-spectrogram + CNN	98.18	99.18	99.28	99.23	95.18	93.33	100	96.55
Mel-spectrogram + BP	94.34	87.50	100	93.33	91.44	93.75	93.75	93.75
MFCC + CNN	97.43	88.88	100	94.12	94.04	100	88.88	94.11
MFCC + BP	96.12	97.19	93.87	97.19	93.45	90.91	100	95.23
MFCC + SVM	95.76	96.99	94.57	95.77	93.29	93.56	91.79	92.67
MFCC + K-means	52.93	42.86	53.09	47.43	50.34	42.44	44.96	43.66
MFCC + Naive-bayes	88.57	95.31	83.83	89.20	78.81	82.43	73.87	77.92
MFCC + LightGBM	95.73	98.46	93.29	95.80	89.89	88.17	89.38	88.77

5. CONCLUSION

There has been extensive work on diagnosing COVID19 using various methods in the field of Artificial Intelligence. And those methods are getting better day by day. In this research project we created an artificial intelligence model which can be used to diagnose COVID19 patients by using forced cough recordings from the patients. This model will be able to provide good results and by adding more and more biomarkers and also considering various symptoms reported by the patient's accuracy can be increased considerably. Table 1[3] shows different methods and their accuracy which exists today.

6. ACKNOWLEDGEMENT

Prof. Manjiri Karande, PVBKCOE, Malkapur and HOD DCE Mr. Farfat Sir PVBKCOE, Malkapur who helped me in conducting this research. I take this opportunity to express my gratitude towards them.

7. REFERENCES

[1] Jordi Laguarta , Ferran Hueto, and Brian Subirana – "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings", IEEE OPEN JOURNAL OF ENGINEERING IN MEDICINE AND BIOLOGY, VOL. 1, 2020
[2] Arup Anupam, N Jagan Mohan, Sudarsan Sahoo, Sudipta Chakraborty, "PRELIMINARY DIAGNOSIS OF COVID-19 BASED ON COUGH SOUNDS USING MACHINE LEARNING ALGORITHMS", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)

[3] Quan Zhou1,Jianhua Shan1,Wenlong Ding1, Chengyin Wang1, Shi Yuan1, Fuchun Sun2, Haiyuan Li3 and Bin Fang- “Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network,” https://www.frontiersin.org/articles/10.3389/frobt.2021.580080/full

[4] Ahmed Fakhry, Xinyi Jiang, Jaclyn Xiao, Gunvant Chaudhari, Asriel Han, Amil Khanzada, "Virufy: A Multi-Branch Deep Learning Network for Automated Detection of COVID-19" https://arxiv.org/ftp/arxiv/papers/2103/2103.01806.pdf

[5] Md. Afzal Hossan, Sheeraz Memon, Mark A Gregory, "A novel approach for MFCC feature extraction", Publisher: IEEE

[6] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk Rodney Kinney, Ziyang Liu, William Merrill, etc. “CORD-19: The Covid-19 Open Research Dataset”, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7251955/

[7] MIT, “Artificial intelligence model detects asymptomatic Covid-19 infections through cellphone-recorded coughs” https://news.mit.edu/2020/covid-19-cough-cellphone-detection-1029

[8] Feng Rong, "Audio Classification Method Based on Machine Learning”, Publisher: IEEE

[9] Krish Naik, “Audio-Classification” – Published on – “GitHub”

[10] Longitudinal Speech Biomarkers for Automated Alzheimer's Detection - https://www.researchgate.net/figure/Impact-of-Poisson-mask-on-AD-performance-Baseline-is-a-ResNet50-trained-on-the-AD-task_tbl1_343728225

