A MEDICAL MOBILE APPLICATION FOR COVID-19 DIAGNOSIS USING COUGH SOUND

Ibrahim Reda Ibrahim Abueisha, Samah Bayoumi Abid, Salma mohamed kishk, and Zeinab Galal Yahya Salem

Tanta University, Egypt, ibrahim.abueisha97@gmail.com, samahabid98@gmail.com, salma.mohamed2255@gmail.com, Zeinabsalem81@gmail.com

Supervisor: Basma Gh. Elkilany, Assistant Professor

Faculty of Engineering, Egypt, Basma_elkilany@f-eng.tanta.edu.eg

Abstract*– Since the COVID-19 pandemic started, researchers have indicated potential techniques to develop COVID-19 screening tools. One practical and affordable solution is to use cough recordings for COVID-19 detection. Based on the combination of Deep learning and signal processing approaches, we present a mobile application for COVID-19 detection using cough recordings. First, AI model is developed and trained using the COUGHVID cough dataset. Our model uses convolution neural networks and an image classifier, to identify COVID-19 infection from an audio file. It takes a Mel Scale spectrogram as an input, which is an image representation of the audio stream and classifies it into COVID infected or healthy. The testing accuracy for our classification model was 98.7%. Then, we develop a mobile application to receive the cough recoding from the user and display the result. Our application is connected to a server which receives the audio file from the application as an HTTP request, runs the stored Python code on it, and then returns a result as an HTTP response

Keywords-- COVID-19, CNN, Deep learning, Mel-spectrogram, Cough

I. INTRODUCTION

Since the new crown pneumonia pandemic first appeared in Wuhan, China in December 2019, it has quickly spread all over the world. The new crown virus has infected 111M confirmed individuals globally as of mid-February 2021 (2.47M deaths were reported and 62.9M recovered), and it has turned into a significant global health issue [1]. Increasing global research has led to more studies on COVID-19's transmission by droplets or direct contact [2]. The respiratory symptoms of COVID-19 include exhaustion, a dry cough, shortness of breath, joint and muscle discomfort, gastrointestinal complaints, and a loss of smell or taste [3]. The acute respiratory distress syndrome can emerge from either the gaseous or vascular side of the alveolus due to its effects on the vascular endothelium, as shown by chest x-rays or computed tomography (CT) scans in patients with COVID-19 [4,5]. COVID-19 can be identified by medical lab tests that analyze exhaled breaths. Using 28 patients positive for COVID-19 and 12 patients negative for COVID-19, this technique achieved an accuracy rate of 93% [6]. In previous work, machine learning algorithms have been applied to detect COVID-19 using image analysis. Using a Resnet50 architecture, for instance, COVID-19 was accurately recognized from Computerized Tomography (CT) images with 96.23% [7]. With an accuracy rate of 96.7%, the same architecture also identified pneumonia caused by COVID-19 [8]. In addition to, the detection of COVID-19 from x-ray images the accuracy was 96.30% [9].

Recently, there has also been interest in the automatic analysis of cough audio for COVID-19 identification. Coughing is a common sign of many lung conditions, and it can have different effects on the respiratory system. Lung disease can change the way the glottis functions and can restrict or clog the airway, which can affect the acoustics of the vocal audio, including speech, breath, and coughing [10,11]. This increases the possibility of detecting the coughing sounds linked to a certain respiratory condition, like COVID-19. Researchers have discovered that a straightforward binary machine learning classifier, such as coughs collected from crowdsourcing data, can differentiate between healthy and COVID-19 respiratory sounds with an Area under the curve (AUC) above 0.8 [12]. Using a convolutional neural network (CNN) for cough and breath audio, better performance was achieved, with an AUC of 0.846 [13].

According to studies that have investigated using a rapid diagnosis of COVID-19 disease to analyze respiratory and coughing noises. This results from the discovery that, despite being distinct from other respiratory coughs, a dry cough brought on by COVID-19 is observed in many COVID-19 patients, according to WHO [14]. For COVID-19 positive and negative cases, several researchers have developed audio databases with brief recordings of coughing and breathing. This serves as the inspiration for the current research presented in this study, which develops a revolutionary deep-learning architecture based on convolutional neural networks (CNNs). based on the extensive COUGHVID cough dataset.

This paper presents:

• Utilizing deep learning techniques, we present a clear and critical assessment of related works on the diagnosis of COVID-19 using cough and respiratory noises.

• We examine the COUGHVID dataset and suggest adding raw signal and spectral data, class balance, and additional variability to improve the effectiveness of machine learning and deep learning-based solutions.

• Using only cough sound as an input, we offer a novel framework based on deep learning attention-based Convolution Neural Networks (CNN) architecture that can identify and classify the Likely-COVID-19 from NonLikely-COVID-19 situations.

In the rest of the paper, the proposed model is presented as follows. First, section II illustrates the dataset COUGHVID. Then, we apply some pre-processing on these datasets, which is presented in Section III. Section IV present the proposed model, implementation, and results. Finally, Section V presents a discussion and conclusion for our work.

II. DATASET

We used dataset [15] from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, which is a large-scale publicly available dataset with 27550 recordings including 1155 positive cases.

These recordings were collected through a web application from April 1st, 2020, to December 1st, 2020, where users were asked to click on "Record" button to start the recording. After the record process is completed, users were asked to fill in a meta-data questionnaire about age, gender, geolocation information, previous existing respiratory conditions, and COVID-19 status.

The latter includes three classes: Healthy, COVID and Symptomatic. Table 1 shows the distribution of COVID status label. The codec of all audio recordings is Opus, with 48kHz sampling rate. For validation purpose, four physician experts assisted in more than 1000 recording annotations. The items that have been annotated were quality of cough, type of cough, Audible dyspnea, Audible wheezing, Audible stridor, Audible choking, Audible nasal congestion, nothing specific is audible, impression about the patient's infection and an impression about the severity. Interestingly, the meta-data includes an entry called cough detected p, which is a float number between 0 and 1 that indicates the extent to which the recording corresponds to a cough or not (probability that the recording is a cough). This value is the output of a machine learning cough classifier. Other meta-data parameters provided in the dataset description such as reported gender, fever or muscle pain, age and respiratory condition. Fig ١ shows a part of csv column.

Table 1 The distribution of the Dataset.

	Healthy	Sympto-matic	COVID	No Status	Total
No. of Samples	12479	2590	1155	11326	27550

III. PRE-PROCESSING DATA

Since, the COUGHVID dataset has some problems, data pre-processing is highly needed. data should be cleaned, balanced, and converted into Mell spectrogram to represent samples in image format can be used in our end-to-end deep learning model.

A. Data Cleaning

The COUGHVID dataset contains both known and unknown status datums. Since the goal of our study is to execute a supervised learning; we disregard any samples without a COVID-19 status (no status specified) about 11326 samples.

Then, refine detection by keeping the highly detected as a cough. As in COUGHVID dataset, each record entry is

	B	C	D	E	F	G	H	I	J	K
1	datetime	cough_det	SNR	latitude	longitude	age	gender	respirator	fever_mus	status
2	2020-11-2	0.0155	7.326171	48.9	2.4					
3	2020-04-1	0.9609	16.15143	31.3	34.8	15	male	FALSE	FALSE	healthy
4	2020-10-1	0.1643	16.2172			46	female	FALSE	FALSE	healthy
5	2020-04-1	0.9301	20.14606	40	-75.1	34	male	TRUE	FALSE	healthy
6	2020-04-1	0.0482	0	-16.5	-71.5					
7	2020-04-1	0.9968	13.1465			21	male	FALSE	FALSE	healthy
8	2020-04-1	0.0735	23.01471	40.6	-3.6					
9	2020-05-1	0.0306	12.71348	13.8	-89.6		female	FALSE	TRUE	COVID-19
10	2020-10-1	0.7811	12.56641	45.7	4.9	20	male	TRUE	FALSE	healthy
11	2020-05-1	0.0307	0	43.6	-6.9					
12	2020-05-2	0.8937	0			28	female	FALSE	FALSE	healthy
13	2020-04-1	0.9883	14.60385	39.4	67.2	15	male	FALSE	FALSE	healthy
14	2020-07-1	1	9.624196			35	male	TRUE	FALSE	symptoma
15	2020-04-1	0.9959	35.64185			46	female	FALSE	FALSE	healthy
16	2020-04-1	1	6.781094							
17	2020-07-2	0.9712	31.34853							
18	2020-04-2	0.824	25.08306							
19	2020-04-1	0.0576	0	46.8	6.6		male	TRUE	FALSE	COVID-19
20	2020-04-1	0.3009	0							
21	2020-04-1	0.8109	15.69776	41.1	28.8	39	other	FALSE	FALSE	healthy

Fig 1 A Part of The CSV File that contains the information about the Dataset.

associated with a probability p of being a cough sound. This probability is generated by a machine learning model [16].

According to [17], the effectiveness of deep architecture was evaluated using several threshold probabilities that ranged from 0.6 to 0.9. We discovered that selecting p = 0.7 produces the best outcomes. Therefore, we discarded all samples with cough detection probability thresholds below 0.7. After cough detection refinement, samples are downed to 731 COVID, 8967 Healthy and 1936 Symptomatic samples.

The third step is to remove silence from the audio recording (like signal shown in Fig 2). Silence parts should be discarded, and important cough part should be kept. Thus, each audio file is segmented into individual coughs using a hysteresis comparator on the signal power [16]., with RMS energy used as a lower threshold of the hysteresis comparator = 0.1 (consider silence) and a high threshold of the hysteresis comparator = 2, and then resulted segments are collected (result shown in Fig 3). All samples are shrunk to a standard length of 7 seconds. Samples that are longer than 7 seconds are deleted, while those that are shorter are evenly padded with zeros at the beginning and the end (final signal used shown in Fig 4).

B. Data Balancing

In COUGHVID dataset, samples are classified into three groups COVID, Symptomatic and Healthy. Due to the absence of COVID positive instances (731 compared to 9258) and considering that Symptomatic records fall under the cases that the patient must be isolated. We combine the positive COVID and Symptomatic cases under COVID, without any modification on the Healthy samples. Thus, the multi-class classification scheme is converted into a binary classification scheme. As a result, our dataset contains 8967 Healthy samples and 2667 COVID samples.

C. Data Balancing

In COUGHVID dataset, samples are classified into three groups COVID, Symptomatic and Healthy. Due to the absence of COVID positive instances (731 compared to 9258) and considering that Symptomatic records fall under the cases that the patient must be isolated. We combine the positive COVID and Symptomatic cases under COVID, without any modification on the Healthy samples. Thus, the multi-class classification scheme is converted into a binary classification

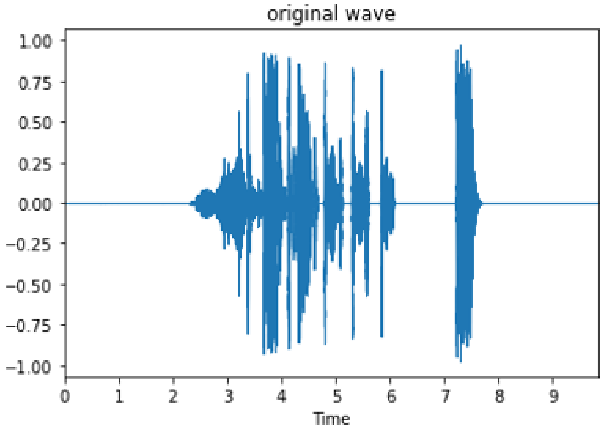

Fig 2 The original signal in Time domain

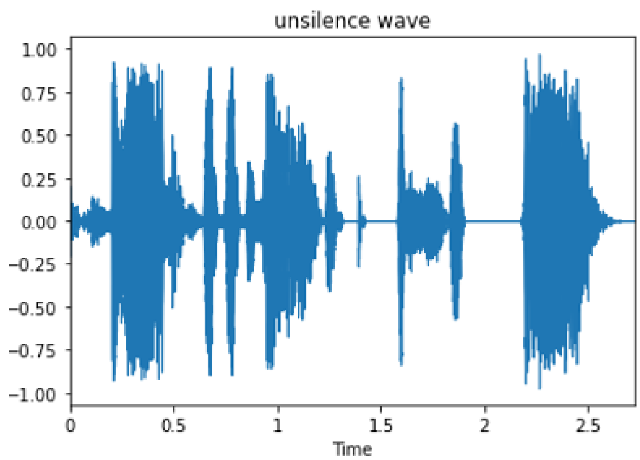

Fig 3 The Unsilenced signal in Time domain

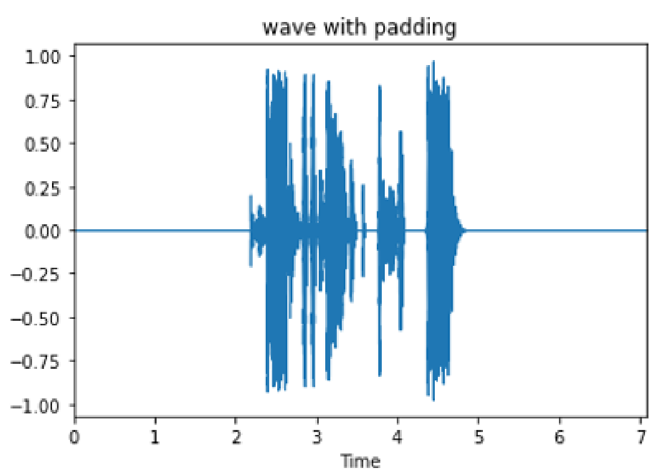

Fig 4 The padded signal in Time domain

scheme. As a result, our dataset contains 8967 Healthy samples and 2667 COVID samples.

To increase our dataset, new samples are created using pitch-shifting augmentation. It is a sound recording technique that raises or lowers the original pitch of a sound by n steps, with step being determined by a semitone. This step is applied only on the COVID samples by the Librosa3 audio processing and analysis Python module, arriving to 5334 COVID samples.

D. Data Balancing

In COUGHVID dataset, samples are classified into three groups COVID, Symptomatic and Healthy. Due to the absence of COVID positive instances (731 compared to 9258) and considering that Symptomatic records fall under the cases that the patient must be isolated. We combine the positive COVID and Symptomatic cases under COVID, without any modification on the Healthy samples. Thus, the multi-class classification scheme is converted into a binary classification scheme. As a result, our dataset contains 8967 Healthy samples and 2667 COVID samples.

To increase our dataset, new samples are created using pitch-shifting augmentation. It is a sound recording technique that raises or lowers the original pitch of a sound by n steps, with step being determined by a semitone. This step is applied only on the COVID samples by the Librosa3 audio processing and analysis Python module, arriving to 5334 COVID samples.

E. Converting to Mel- spectrogram

Samples must be converted to images, So Mel-spectrogram representation is used. As it is common and has good result in audio classification. To get Mel-spectrogram for sample two steps are applied. First, time domain signals are converted into spectrogram (result is represented in Fig 5) using Short Time Fourier transform (STFT). That is resulted from get Fast Fourier transform FFT (get spectrum domain that is shown in Fig 6) for frames along time domain. Second, applying the Mel-scale, introduced by Stevens, Volkmann, and Newmann in 1937 [18], is pitch unit that makes identical pitch distances sound similarly far to the listener. It is applied to the frequencies to convert them to the Mel-scale as following in equation (1).

$$mel = 2955 * \log_{10} \frac{1 + hertz}{700}. \quad (1)$$

Where $hertz$ is the frequencies values in spectrogram in HZ. Therefore, the Mel-spectrogram as shown in Fig 7 is the conversion of spectrogram frequencies to the Mel-scale.

F. Enlarge dataset using spectral augmentation

As a final number of samples is not enough to get good performance and accuracy and to make our model more robust, SpecAugment technique [19] is applied to our dataset. SpecAugment is common in end-to-end problems. SpecAugment uses Mel-spectrograms (like Fig 8) and three-step augmentation method. First, using Time warping, within the time steps, a random point along the horizontal axis passing through the centre of the Mel-spectrogram image is warped to the left or right by a distance selected from a uniform distribution ranging from 0 to the time warp parameter along that line result is represented in Fig 9. Second, Frequency masking is employed as a mechanism of masking f consecutive Mel frequency channels [f0, f0 + f], where f is selected from a uniform distribution ranging from 0 to the frequency mask parameter F result is represented in Fig 10. The third step consists of Time masking, where T

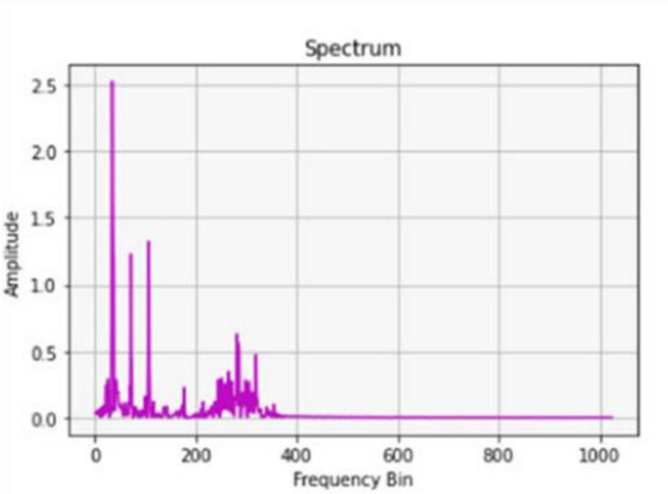

Fig 5 The signal spectrum.

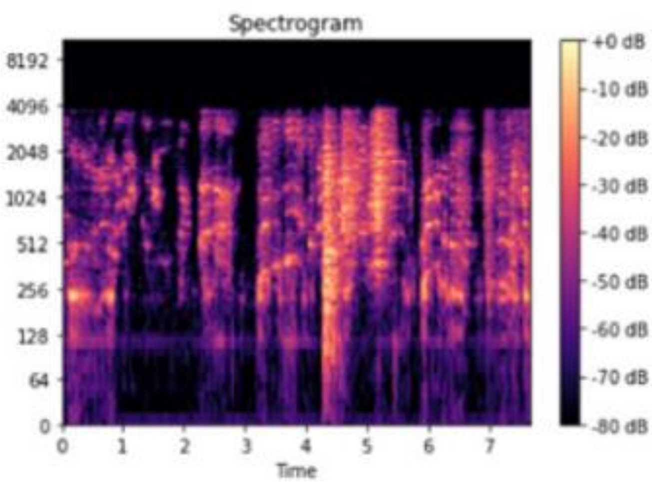

Fig 6 The Signal Spectrogram.

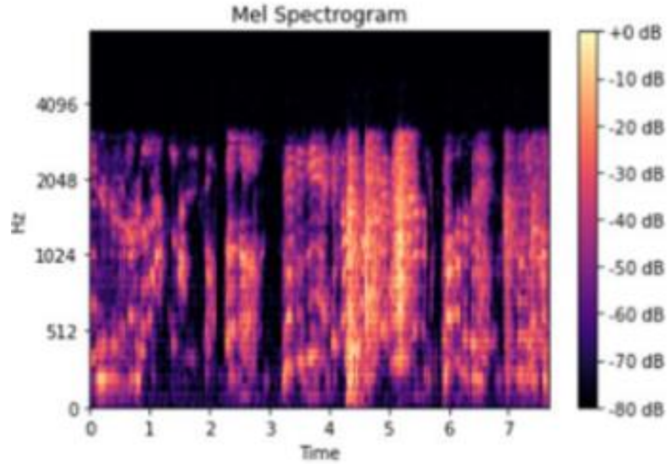

Fig 7 The Signal Spectrogram in Mel scale.

successive time steps [t0, t0 +t) are masked, such that t is taken from a uniform distribution from 0 to the time mask parameter result is represented in Fig 11. Frequency masking

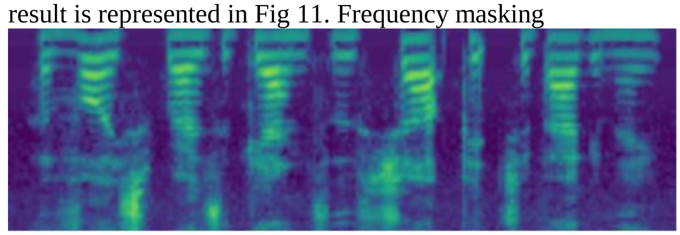

Fig 8 The Original Mel-spectrogram.

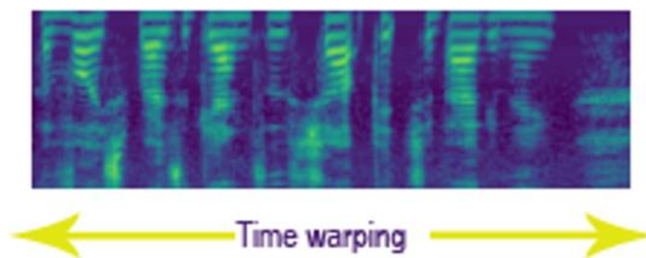

Fig 9 Mel-spectrogram after Time warping.

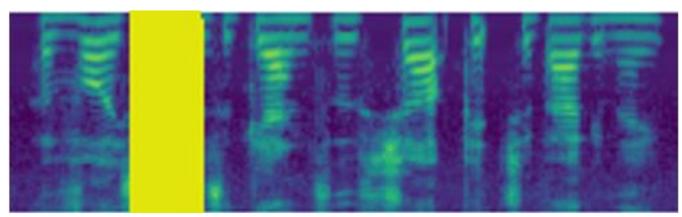

Fig 10 Mel-spectrogram after frequency masking.

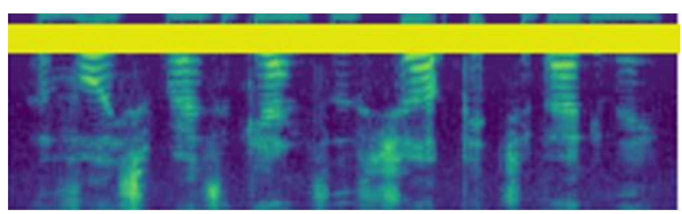

Fig 11 Mel-spectrogram after Time masking.

and time masking is applied to randomly generate two new Mel-spectrograms of COVID samples (beside original Mel-spectrogram) that increase COVID samples triple 16002 and one for Healthy samples COVID (beside original Mel-spectrogram) that increase Healthy samples double 17934.

IV. Implementation and Results

After pre-processing, A CNN is developed to Classify the audio samples into COVID or Healthy, Table 2 shows the specification of the developed CNN. To compile the CNN model, we need to define loss function which is categorical cross-entropy, accuracy metrics which is accuracy score, and an optimizer which is Adam with learning rate = 0.001.

We train the model and save the model in HDF5 format. We train the model for 500 epochs and batch size 256. We use early stop function to stop training before a model starts to be overfit. Or dataset is split into 70% for training, 20% for validation and 10% for test phase.

Fig 12 shows the evaluated accuracies and losses for both training and validation phases. Confusion matrix for our model is shown in Fig 13 where 1 out 1601 positive cases are misclassified and 43 out 1795 negative cases. From confusion matrix it is noticed that f1-score is 0.99 that refers to the

Table 2 Parameters of the modeled CNN.

Layer (type)	Output shape	Param #
Conv2d_1 (Conv2D)	(None, 255, 255, 16)	208
Average_pooling2d (AveragePooling2D)	(None, 254, 254, 16)	0
dropout (Dropout)	(None, 254, 254, 16)	0
Conv2d_2 (Conv2D)	(None, 253, 253, 16)	1040
Average_pooling2d _1 (AveragePooling2D)	(None, 252, 252, 16)	0
dropout_1 (Dropout)	(None, 252, 252, 16)	0
flatten (Flatten)	(None, 1016064)	0
dense (Dense)	(None, 32)	32514080
dropout_2 (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 2)	66
Total params: 32,515,394 Trainable params: 32,515,394 Non-trainable params: 0		

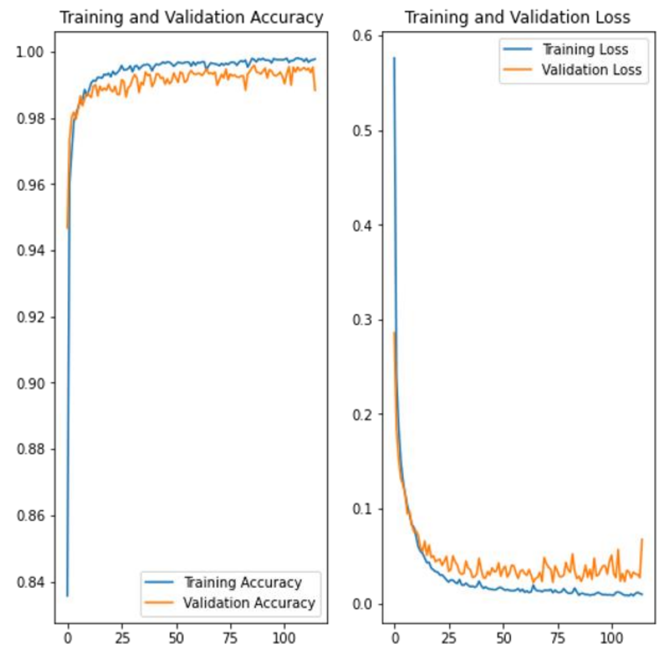

Fig 12 Evaluated accuracies and losses in both training and validation phases.

prefect precision and recall as shown in Fig 14. Obviously, our model has reached an accuracy of almost 98.7% of classification.

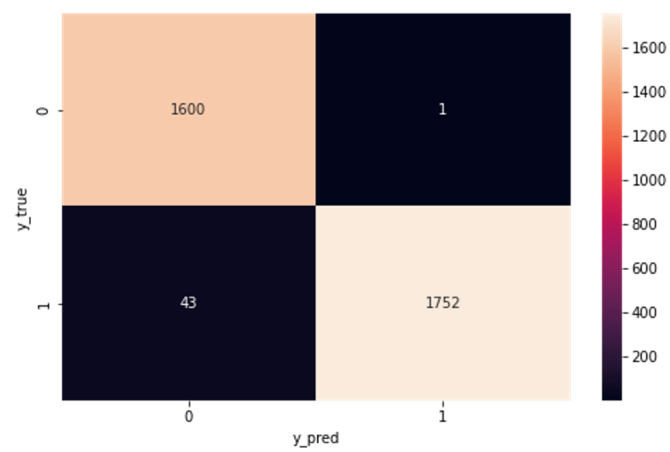

Fig 13 The confusion Matrix.

```
              precision    recall  f1-score   support

           0       0.97      1.00      0.99      1601
           1       1.00      0.98      0.99      1795

    accuracy                           0.99      3396
```

Fig 14 Parameters extracted from the confusion Matrix.

V. Conclusion

In our study, we developed a COVID-19 diagnosis medical mobile application based on cough sound. It is a simple pre-screening tool to determine which people are suspected to be infected with the covid or not before they can go to take the test. Our application mainly consists of the AI model, the mobile application, and the backend server.

In Al part, A CNN was developed which works as an image classifier. It receives an input in the form of a Mel Scale spectrogram and the model can classify the file into one of two classes, either COVID infected or Healthy. Our Dataset was divided into 70% training, 20% validation, and 10% testing after preprocessing the dataset. The model can detect infection of COVID-19 from an audio with high accuracy reaches to 98.7%.

The mobile application was developed using Flutter framework and written in dart language. The application user interface was designed by Figma tool, and then developed in dart language. The application functionality is responsible for getting the input from the user, encoding it in wav format, and dealing with the Backend server and the AI model.

ACKNOWLEDGMENT

We would like to express our very great appreciation to Dr. Basma Elkilany for her valuable and constructive suggestions during the planning and development of this research work.

We also would like to express our sincere appreciation to Eng. Mohamed El-Komy and Eng. Marwa Reda, lecture assistants in our department, who provided us with the needed information and materials and helped us to build the AI model which was more than significant.

REFERENCES

[1] Feng, Ke, et al. "Deep-learning based approach to identify covid-19." *SoutheastCon 2021*. IEEE, 2021.

[2] Wölfel R, Corman VM, et al. "Lirological assessment of hospitalized patients with COVID-2019". *Nature. 2020* Dec;588(7839): E35. doi: 10.1038/s41586-020-2984-3. Erratum for: Nature. 2020 May;581(7809):465-469. PMID: 33303961.

[3] Carfì, Angelo, Roberto Bernabei, and Francesco Landi. "Persistent symptoms in patients after acute COVID-19." *Jama* 324.6 (2020): 603-605.

[4] Marini, John J., and Luciano Gattinoni. "Management of COVID-19 respiratory distress." *Jama* 323.22 (2020): 2329-2330.

[5] Yonker, Lael M., Kunling Shen, and T. Bernard Kinane. "Lessons unfolding from pediatric cases of COVID-19 disease caused by SARS-CoV-2 infection." *Pediatric pulmonology* 55.5 (2020): 1085.

[6] Grassin-Delyle, Stanislas, et al. "Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study." *EBioMedicine* 63 (2021): 103154.

[7] Walvekar, Sanika, and Dr Shinde. "Detection of COVID-19 from CT images using resnet50." *2nd International Conference on Communication & Information Processing (ICCIP)*. 2020.

[8] Sotoudeh, Houman, et al. "Artificial intelligence empowers radiologists to differentiate pneumonia induced by COVID-19 versus influenza viruses." *Acta Informatica Medica* 28.3 (2020): 190.

[9] Yildirim, Muhammed, and Ahmet Cinar. "A Deep Learning Based Hybrid Approach for COVID-19 Disease Detections." *Traitement du Signal* 37.3 (2020): 461-468.

[10] Knocikova, Juliana, et al. "Wavelet analysis of voluntary cough sound in patients with respiratory diseases." *J Physiol Pharmacol* 59.Suppl 6 (2008): 331-40.

[11] Higenbottam, T. "Chronic cough and the cough reflex in common lung diseases." *Pulmonary pharmacology & therapeutics* 15.3 (2002): 241-247.

[12] Han, Jing, et al. "Sounds of COVID-19: exploring realistic performance of audio-based digital testing." *NPJ digital medicine* 5.1 (2022): 1-9.

[13] Coppock, Harry, et al. "End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study." *BMJ innovations* 7.2 (2021).

[14] World Health Organization. WHO Coronavirus (COVID-19) Dashboard [Internet]. Available from: https://covid19.who.int/

[15] Orlandic, Lara, Tomas Teijeiro, and David Atienza. "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms." *Scientific Data* 8.1 (2021): 1-10.

[16] Orlandic, L., Teijeiro, T. & Atienza, D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Sci Data* **8,** 156 (2021). https://doi.org/10.1038/s41597-021-00937-4

[17] Hamdi, Skander, et al. "Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound." *Journal of Intelligent Information Systems* (2022): 1-23.

[18] Imran, Ali, et al. "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app." *Informatics in Medicine Unlocked* 20 (2020): 100378.

[19] Singh, Dilbag, Vijay Kumar, and Manjit Kaur. "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks." *European Journal of Clinical Microbiology & Infectious Diseases* 39.7 (2020): 1379-1389.