COVID-19 Detection System Using Recurrent Neural Networks

Abdelfatah Hassan[1], Ismail Shahin[2], Mohamed Bader Alsabek[3]
Electrical Engineering Department
University of Sharjah
Sharjah, United Arab Emirates
[1]U17200027@sharjah.ac.ae, [2]ismail@sharjah.ac.ae, [3]U16104773@sharjah.ac.ae

***Abstract*— Lately, an immense amount of work has been done by people working on the frontlines, such as hospitals, clinics, and laboratories, alongside researchers and scientists who are also making considerable efforts in the fight against the COVID-19 epidemic. Due to the unconscionable dissemination of the disease, the implementation of Artificial Intelligence (AI) has made a significant contribution to the digital health district by applying the basics of Automatic Speech Recognition (ASR) and deep learning algorithms. In this study, we highlight the importance of speech signal processing in the process of early screening and diagnosing the COVID-19 virus by utilizing the Recurrent Neural Network (RNN) and specifically its significant well-known architecture, the Long Short-Term Memory (LSTM) for analyzing the acoustic features of cough, breathing, and voice of the patients. Our results show a low accuracy in the voice test compared to both coughing and breathing sound samples. Moreover, our results are preparatory, and there is a possibility to enhance the accuracy of the voice tests by expanding the data set and targeting a larger group of healthy and infected people.**

Keywords*— *Artificial Intelligence; Breathing; coughing; COVID-19; Deep Learning; Recurrent Neural Network.

I. Introduction

Until this moment and according to statistics from the World Health Organization (WHO), the number of COVID-19 confirmed cases had exceeded 15 million, with casualties of 600,000 people in more than 200 countries across the world. The WHO has stated the pivotal COVID-19 symptoms, namely, body aches, high body temperature, heavy coughing, and severe predicament in breathing [1]. Scientists claim that audio sounds generated by the respiratory system can be diagnosed and analyzed in order to determine the presence of the disease. The preposterous and escalated spread of the COVID-19 virus has led to a massive collaboration among various fields to control the dissemination of the virus and prevent its collateral damages daily. Moreover, an innovative and contemporary way solution is set to take its part of the combat against COVID-19 by implementing the Artificial Intelligence (AI) and deep learning algorithms in the digital-health district. Lately, there has been massive implementations of the AI alongside with the deep learning methods in the fields of Automatic Speaker Recognition (ASR) and Speech-Audio analysis that could be pragmatic in the process of early detection and screening of COVID-19, by performing analysis on the extracted features of coughing, breathing and speech sound using the Recurrent Neural Network (RNN) as classifier.

This paper is organized as follows: Section II covers the related work. Section III discusses the speech corpus and feature extraction. Section IV illustrates the architecture of the RNN network. Section V discusses the experimental results. Finally, section VI gives the concluding remarks of this work.

II. Literature Review

Numerous number of studies have been made for the promotion of automated screening and diagnosing that is based on the analysis of chest CT-images [2], [3], [4], [5]. AI can be clenched and enforced in the e-health district to aid the early detection of COVID-19 by analyzing these three main sounds, namely, coughing, breathing, and voice [1]. Further, respiratory sounds can carry a variety of indications on the human being's health state, which can be recognized and diagnosed by the implementation of machine learning algorithms [6]. Thus, since the massive outbreak of the COVID-19 virus, scientists and researchers are now considering the detection of COVID-19 from respiratory sounds [7]. In [8] and [9], a low consuming power and a modern wearable system is proposed for the detection of both asthma and wheezing based on the analysis of their sound features and the frequencies of their respiratory sounds. In [10], Convolutional Neural Networks (CNN) are utilized to detect different types of coughs based on the analysis of their extracted sound features. Besides, there has been a proposed system for predicting COVID-19 using deep learning algorithms and several classifiers such as CNN, Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANN) [11]. The detection of COVID-19 patients' health state can be obtained utilizing their speech signals. Therefore, a health state detection system can be used to observe and analyze the sleep-quality, severity of illness, fatigue, and anxiety [12]. Cough has always been a symptom of many diseases. In fact, it is possible to distinguish between coughs and assess the type of illness by testing the auditory features using multiple classifiers [13]. Moreover, in [14] a proposed system composed of a novel multi-pronged mediator AI architecture model is set to differentiate between the different types of coughs.

In this paper, we propose a system for detecting the COVID-19 based on speech and sound analysis of the different extracted acoustic features using RNN, specifically the LSTM architecture.

III. Speech Corpus and Features Extraction

A. *Speech Corpus*

The collected dataset can easily influence the robustness and performance of our detection system. Thus, it is vital to have an appropriate dataset to train and test the system. Our speech corpus was collected from 60 Healthy speakers (40 male and 20 female), and 20 COVID-19 patients (12 male and 8 female). Each participant was asked to record a sample of cough sounds, breathing sounds, and voice. Therefore, three samples were obtained from each participant. The total number of acoustic data used in our system was 240 ((60 healthy participants × 3 recordings) + (20 COVID-19 participants × 3 recordings)). Samples of COVID-19 patients were recorded in different United Arab Emirates hospitals.

Note that all the dataset samples were captured using a mobile microphone. Besides, we have manually pre-processed (clearance of silence) all user samples using software called PRAAT. Fig. 1a and Fig 1b display the cough signal of COVID-19 infected patients and healthy people in time and frequency domain, respectively.

B. *Extraction of Features*

The sound waves have a set of parameters called speech features. The determination of these features is an important step that affects system accuracy. In the present work, various features are extracted from the captured dataset. A list of the extracted features is provided below:

- **Spectral Centroid (SC):** The mean of the spectral energy, that illustrates the signal frequency changes and signal phase content over time. Also, SC allows us to find the exact location of the dominant formant on each sub-band [15], [16].
- **Spectral Roll-off (SR) :** The skewness of the signal's spectrum and the 90th percentile of the power spectral distribution. The SR can describe the slope of the signal's spectrum. Besides, it is used to differentiate between voiced and unvoiced sounds [16], [17].
- **Zero-Crossing Rate (ZCR):** The number of times the audio signal sign changes from positive to negative and vice versa. The ZCR can estimate the dominant frequency component of the signal [18], [19].
- **Mel-Frequency Cepstral Coefficients (MFCC):** MFCC is an essential feature that is mostly employed in the emotion recognition field because it provides a high-level representation of human auditory perception [20], [21], [22], [23]. Moreover, the MFCCs are computed using a psycho acoustically motivated filter bank followed by a logarithmic compression and Discrete Cosine Transform (DCT). The MFCCs can be obtained as given in the following formula [24]:
$$C(n) = \sum_{m=1}^{M} [\log Y(m)] \cos\left[\frac{\pi n}{M}\left(m - \frac{1}{2}\right)\right] \quad (1)$$
where the output of the M-channel filter bank is denoted by *Y(m)*, and the index of the cepstral coefficient is referred to *n*.
- **Δ MFCC:** The first order time derivative of MFCC. For computing the differential coefficients, we use the following expression [24], [25]:
$$D_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2} \quad (2)$$
where D_t is the delta coefficient of the frame *t,* and c_{t+n} to c_{t-n} are the static coefficients.
- **Δ^2 MFCC:** The second order time derivative of MFCC. These features can be obtained from Δ MFCC using Eq. (2).

Feature extraction is considered a significant step since one of our objectives is to find a suitable feature extraction method that improves our system's accuracy. Fig. 2 shows the extracted acoustic features from COVID-19 and non-COVID-19 cough sounds. Further, it can be seen that the COVID-19 and non-COVID-19 uncorrelated from features perspective.

IV. Network Architecture

A. *Recurrent Neural Network (RNN)*

The RNN is utilized mainly to predict the future data sequence through the use of previous data samples. The RNN is very commonly used in modeling sequence data such as speech or text. However, these networks have not been broadly used since they are considered difficult to train, such that it captures the long-term dependencies [26]. RNN output is calculated by iterating the following equations from time *t* = 1 to *t* = *T*:

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (3)$$

$$y_t = W_{hy}h_t + b_y \quad (4)$$

where the input is denoted by x, the output sequence is referred to y, and h is the hidden vector sequence. Besides, W refers to weight matrices, b refers to the bias vector, and $\mathcal{H}$ is the hidden layer function [27].

B. *Long Short-Term Memory (LSTM)*

RNN suffers from the vanishing gradient problem, which rises with the length of the training sequences. Therefore, LSTM is utilized to overcome this problem. LSTM stores data information for a long period, and it is easier to remember the past data in the memory. Besides, for the version of LSTM used in our system, the standard formulation of a single LSTM cell unit can be given by the following equations [27]:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (6)$$

$$\tilde{C}_t = tanh(W_C[h_{t-1}, x_t] + b_C) \quad (7)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (8)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t * \tanh(C_t) \quad (10)$$

where *σ, i, f, o, C,* $\tilde{C}$ is the sigmoid function, input gate, forget gate, output gate, memory cell content, and new memory cell content, respectively. The sigmoid function is utilized to form three gates in the memory cell, while the tanh function is used to enlarge the memory cell output [27].

V. Results and Experiments

A. *Training and Testing Procedure*

In this work, the RNN architecture that we are using is LSTM. Fig. 3 displays the structure of this LSTM. We used the extracted features from our collected database for network training and testing. Besides, 70% of the data was used for training and 30% for testing. Training and testing phases are crucial for our system. The training sets are represented in a vector and passed to the LSTM network. These measurements are compared with goal classes, and the weights during the preparation process are modified. The test signals are then sent to the network, and their aim values are calculated depending on the weights trained [28]. Fig. 4 illustrates the accuracy and loss of one trained and tested model. The training and testing were made using python, where TensorFlow has been utilized as the deep learning library. Table I shows the parameters that have been set to create our deep learning models.

B. *Feature Extraction Importance*

The selection of features to be utilized is essential to the overall performance of the system. Thus, in this work, six acoustic features were obtained from the collected sounds and passed to our system classification network. In Fig. 5, we illustrate the feature importance based on system accuracy. To obtain these results, we used only cough samples. Furthermore, it can be seen that the MFCC features have the highest accuracy. Therefore, we will continue our system analysis on the entire captured samples with MFCCs, since they enhance system accuracy.

C. *System Overall Performance*

To evaluate our system, we use the performance metrics of precision, recall, F1-score, Area Under Curve (AUC), and accuracy. The normalized confusion matrix and performance metrics for our COVID-19 detection system are reported in Fig. 6 and Table II, respectively. These results were obtained based on, namely, cough sounds, breath sounds, and voices. When comparing the three types of sounds, the best accuracy is achieved for breathing sound, reaching up to 98.2%. Then, for cough sounds, an accuracy of 97% is attained. When it comes to voices, the accuracy of the system is only 88.2%. Our analysis shows that we can rely in the first place on collecting cough and breathing sounds to make a COVID-19 detection system.

VI. Concluding Remarks

This paper has provided a groundbreaking and modern approach for early diagnosis of COVID-19. It also illustrates the mechanism of the proposed COVID-19 detection system. This analysis is done by evaluating the different acoustic features of cough sound, breathing sound, and voice. Based on the obtained results, it can be stated that Patients' voice has shown such inconvenient accuracy compared to cough and breathing sounds. The reason behind these in-efficient preliminary results is that due to time constraints, the collected data set is comparatively small and lacks control group data on healthy subjects and other patients suffering from different kinds of respiratory illnesses. Eventually, we can conclude that the analysis and observations based on the patient's cough and breathing are the most effective factors to diagnose infection. Due to the consequences that some people may experience from quarantine, the production of chatbots is soon possible to provide mental support and aid the control of anxiety disorders [29].

Acknowledgments

The authors would like to thank the University of Sharjah in the United Arab Emirates to fund this work through the competitive research project entitled "Emirati-Accented Speaker and Emotion Recognition Based on Deep Neural Network, No. 19020403139" and the spotlight project called "Capturing Emirati-Accented Speech Corpora for Applications of Speech Signal Processing" which is supported by the College of Engineering.

References

[1] [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019. [Accessed 28 June 2020].

[2] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia and J. Xia, "Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT," *Radiology,* pp. 1-16, March 2020.

[3] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis and A. Mohammadi, "COVID-CAPS: A Capsule Network-based Framework for Identification of COVID-19 cases from X-ray Images," pp. 1-5, April 2020.

[4] L. Wang and A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," pp. 1-12, May 2020.

[5] M. Farooq and A. Hafeez, "COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs," 2020.

[6] R. X. A. Pramono, S. Bowyer and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS ONE,* vol. 12, no. 5, pp. 1-43, 2017.

[7] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng and X. Li, "COVID-19 and Computer Audition: An Overview on What Speech & Sound Analysis Could Contribute in the SARS-CoV-2 Corona Crisis," pp. 1-8, March 2020.

[8] S. H. Li, B. S. Lin, C. H. Tsai, C. T. Yang and B. S. Lin, "Design of wearable breathing sound monitoring system for real-time wheeze detection," *Sensors ,* vol. 17, no. 1, pp. 1-15, 2017.

[9] D. Oletic and V. Bilas, "Energy-efficient respiratory sounds sensing for personal mobile asthma monitoring," *IEEE Sensors Journal,* vol. 16, no. 23, pp. 8295-8303, 2016.

[10] C. Bales, M. Nabeel, C. N. John, U. Masood, H. N. Qureshi, H. Farooq, I. Posokhova and A. Imran, "Can Machine Learning Be Used to Recognize and Diagnose Coughs?," pp. 1-10, May 2020.

[11] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons and Fractals,* pp. 1-14, June 2020.

[12] G. Deshpande and B. Schuller, "An Overview on Audio, Signal, Speech, & Language Processing for COVID-19," pp. 1-5, May 2020.

[13] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta and C. Mascolo, "Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data," June 2020.

[14] A. Imran, I. Q. H. N. Posokhova, U. Masood, S. Riaz, K. Ali, C. N. John, I. Hussain and M. Nabeel, "AI4COVID-19: AI enabled preliminary diagnosis for

COVID-19 from cough samples via an app," *Informatics in Medicine Unlocked,* pp. 1-11, May 2020.

[15] F. Alías, J. C. Socoró and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Applied Sciences,* vol. 6, no. 5, pp. 1-44, 2016.

[16] N. Kamarudin, S. A. Al-Haddad, S. J. Hashim, M. A. Nematollahi and A. R. B. Hassan, "Feature extraction using Spectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent Automatic Identification," *2014 IEEE Student Conference on Research and Development, SCOReD 2014,* pp. 1-6, 2014.

[17] E. Mezghani, M. Charfeddine, C. Ben Amar and H. Nicolas, "Multifeature speech/music discrimination based on mid-term level statistics and supervised classifiers," *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA,* pp. 1-8, 2016.

[18] Y. A. Ibrahim, J. C. Odiketa and T. S. Ibiyemi, "Preprocessing technique in automatic speech recogntion for human computer interaction: an overview," *Annals. Computer Science Series,* vol. 15, no. 1, pp. 186 - 191, 2017.

[19] M. Kos, Z. Kačič and D. Vlaj, "Acoustic classification and segmentation using modified spectral roll-off and variance-based features," *Digital Signal Processing: A Review Journal,* vol. 23, no. 2, pp. 659-674, 2013.

[20] I. Shahin, "Studying and enhancing talking condition recognition in stressful and emotional talking environments based on HMMs, CHMM2s and SPHMMs," *Journal on Multimodal User Interfaces,* vol. 6, no. 1, pp. 59-71, 2012.

[21] I. Shahin, "Employing both gender and emotion cues to enhance speaker identification performance in emotional talking environments," *International Journal of Speech Technology,* vol. 16, no. 3, pp. 341-351, 2013.

[22] I. Shahin, "Employing second-order circular suprasegmental hidden markov models to enhance speaker identification performance in shouted talking environments," *Eurasip Journal on Audio, Speech, and Music Processing,* vol. 2010, 2010.

[23] I. Shahin, "Employing Emotion Cues to Verify Speakers in Emotional Talking Environments," *Journal of Intelligent Systems,* vol. 25, no. 1, pp. 3-17, 2016.

[24] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication,* vol. 52, no. 1, pp. 12-40, 2010.

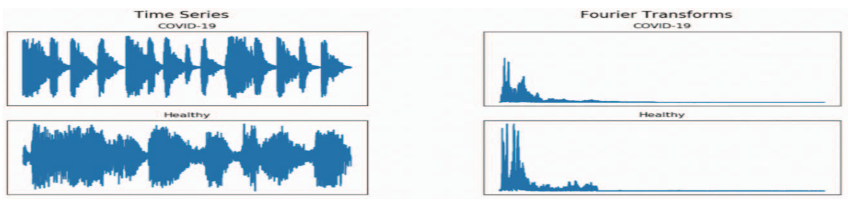
[25] R. Ranjan and A. Thakur, "Analysis of feature extraction techniques for speech recognition system," *International Journal of Innovative Technology and Exploring Engineering,* vol. 8, no. 7C2, pp. 197-200, 2019.

[26] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," *IEEE Access,* vol. 7, pp. 19143-19165, 2019.

[27] A. N. D. L. S. A. Shewalkar, "Performance Evaluation of Deep neural networks Applied to Speech Recognition: Rnn, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research,* vol. 9, no. 4, pp. 235-245, 2019.

[28] P. E. James, M. H. Kit, C. A. Vaithilingam and A. T. W. Chiat, "Recurrent neural network-based speech recognition using MATLAB," *International Journal of Intelligent Enterprise,* vol. 7, pp. 56-66, January 2020.

[29] A. S. Miner, L. Laranjo and A. B. Kocaballi, "Chatbots in the fight against the COVID-19 pandemic," *npj Digital Medicine,* vol. 3, no. 1, pp. 1-4, May 2020.



(a) (b)

Figure 1. (a) Raw data of contaminated and uncontaminated cough sounds; (b) spectrum of contaminated and uncontaminated cough sounds

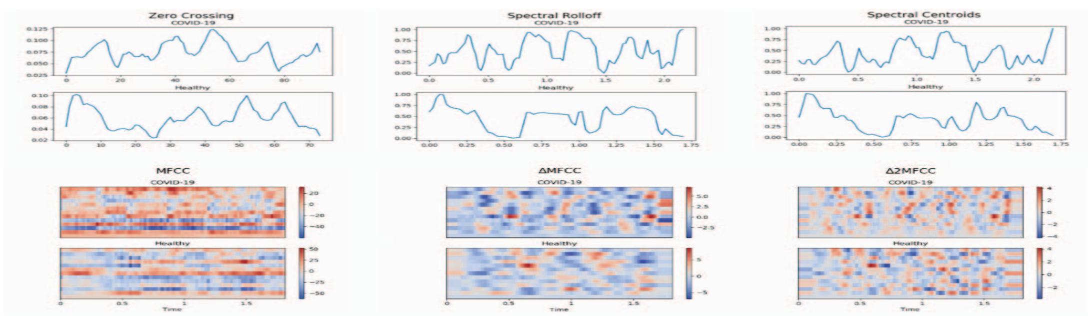


Figure 2. Six different extracted features from COVID-19 and non-COVID-19 cough sounds

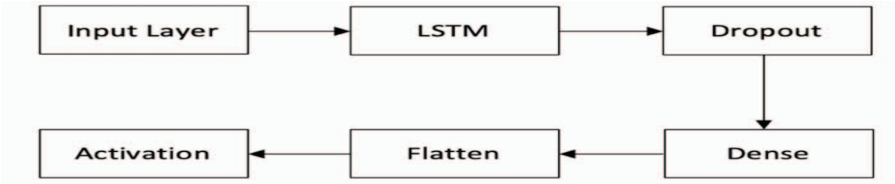
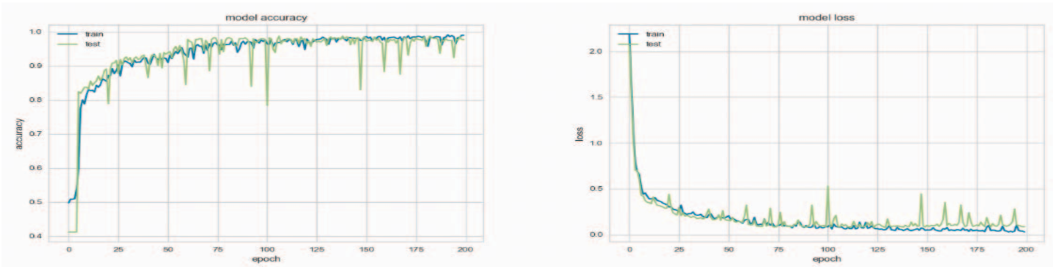


Figure 3. Long Short-Term Memory network architecture



(a) (b)

Figure 4. The Plot of (a) model accuracy on training and testing dataset; (b) model loss on training and testing dataset

TABLE I. CLASSIFICATION NETWORK PARAMETERS.

Parameters	layers	Activation	Learning rate	epoch	Optimizer	fully connected units	fully connected layers	LSTM units	Dropout
Value	2	ReLU	10^{-3}	200	Adam	64, 32, 16	1, 2, 3	512	0.5

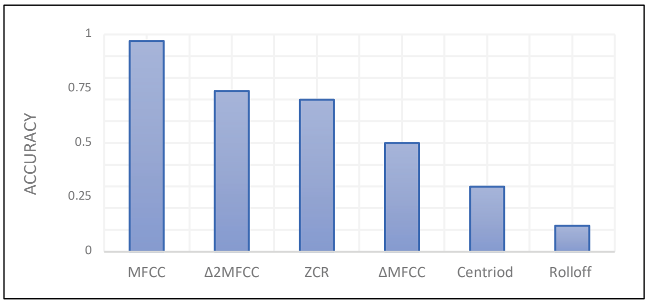
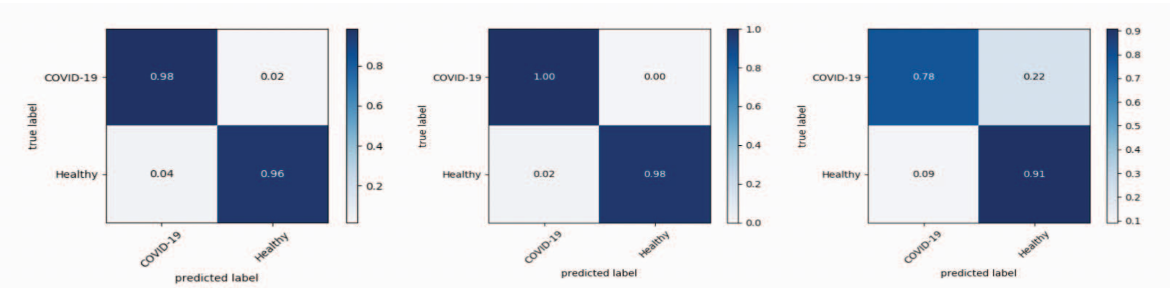


Figure 5. Feature importance of cough sounds based on classification accuracy



(a) (b) (c)

Figure 6. Normalized confusion matrix for COVID-19 detection system based on (a) cough sound; (b) breathing sound; (c) voice

TABLE II. PERFORMANCE METRICS FOR COVID-19 DETECTION.

Sample	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	Accuracy (%)
Cough sounds	99.3	96.4	97.9	97.4	97
Breathing sounds	100	97.7	98.8	98.8	98.2
Voices	94.3	90.8	92.5	84.4	88.2