International Journal of Speech Technology (2022) 25:641–649
https://doi.org/10.1007/s10772-021-09878-0

An adaptive speech signal processing for COVID-19 detection using deep learning approach

Kawther A. Al-Dhlan[1]

Received: 9 December 2020 / Accepted: 29 July 2021 / Published online: 21 August 2021

Abstract
Researchers and scientists have been conducting plenty of research on COVID-19 since its outbreak. Healthcare professionals, laboratory technicians, and front-line workers like sanitary workers, data collectors are putting tremendous efforts to avoid the prevalence of the COVID-19 pandemic. Currently, the reverse transcription polymerase chain reaction (RT-PCR) testing strategy determines the COVID-19 virus. This RT-PCR processing is more expensive and induces violation of social distancing rules, and time-consuming. Therefore, this research work introduces a generative adversarial network deep learning for quickly detect COVID-19 from speech signals. This proposed system consists of two stages, pre-processing and classification. This work uses the least mean square (LMS) filter algorithm to remove the noise or artifacts from input speech signals. After removing the noise, the proposed generative adversarial network classification method analyses the mel-frequency cepstral coefficients features and classifies the COVID-19 signals and non-COVID-19 signals. The results show a more prominent correlation of MFCCs with various COVID-19 cough and breathing sounds, while the sound is more robust between COVID-19 and non-COVID-19 models. As compared with the existing Artificial Neural Network, Convolutional Neural Network, and Recurrent Neural Network, the proposed GAN method obtains the best result. The precision, recall, accuracy, and F-measure of the proposed GAN are 96.54%, 96.15%, 98.56%, and 0.96, respectively

Keywords COVID-19 · Automatic speech recognition · Generative adversarial network · Mel-frequency cepstral coefficients

1 Introduction

COVID 19 is a respiratory contaminant due to the most severe respiratory disease, Covid 2 (SARS-CoV-2) (Trouvain & Truong, 2015). Many worldwide have an infection rate between 1 and 10% in many countries, and the condition has not been officially reported (James, 2015). Figure 1 shows the Evolution of COVID-19 cases and deaths up to august 2020. This development direction began on January 4, 2020, and has constrained numerous nations to take serious control estimates across country lockdowns and scaling-up of the confinement offices in emergency clinics (Sakai, 2015; Schuller et al., 2014). Lockdown process is valuable because it gives excellent time and scope of testing for a maximum number of patients. Reverse transcription polymerase chain reaction (RT-PCR) is one of the best methods for analyzing and detecting COVID 19 within 48 h (Ghosh et al., 2015, 2016a, 2016b; Usman, 2017).

The testing interaction incorporates (i) avoid social distance, it grows the chances for effectively spreading the infection, (ii) the expense of having chemical reagents and widgets, (iii) testing time is high, and (iv) obstacles in huge-scale spread. Attempts to predict a more significant number of COVID-19 cases have led to productive recommendations on innovative solutions for medical services (Botha et al., 2018; McKeown et al., 2012; Porter et al., 2019; Windmon et al., 2018). In particular, progress needs to be made to test simpler, less expensive, and more accurate diagnosis approaches (Breathing sounds for COVID-19, 2020; Indian Institute of Science, 2020; Menni et al., 2020). A few countries have changed the essential, policymaking, and economic restructuring of medical services. The attention is also focused on the purpose of diagnosis tools, innovation arrangements that can be facilitated quickly for pre-screening, and exploring less expensive options than RT-PCR

✉ Kawther A. Al-Dhlan
K_Aldhlan@Hotmail.Com

[1] Information and Computer Science Department, University of Ha'il, Hail, Kingdom of Saudi Arabia

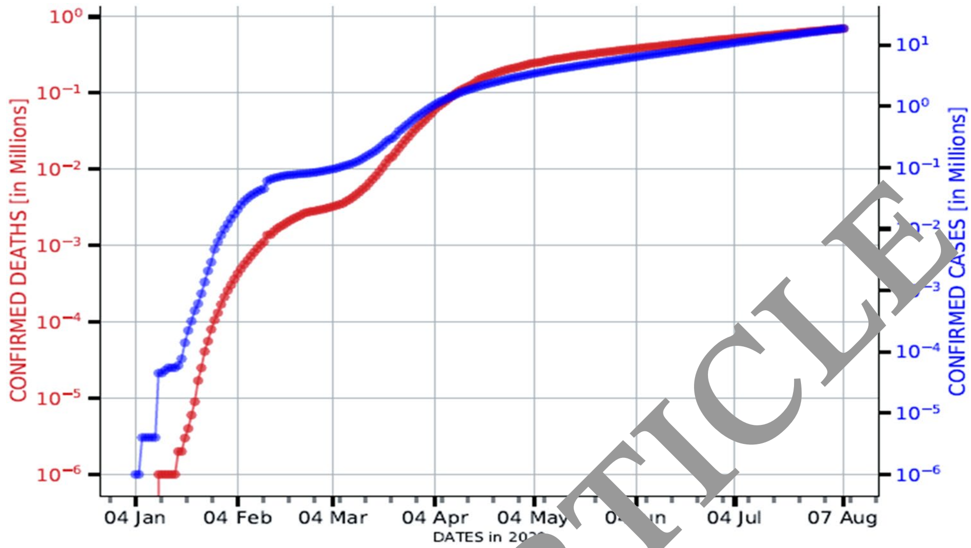


Fig. 1 Ratio of COVID-19 cases up to August 2020. *Source* https://arxiv.org/pdf/2005.10548.pdf

test, which will overcome the chemical testing method's drawbacks.

COVID 19 identification and testing development are being carried out in various laboratories around the world. The WHO and the CDC have identified speech loss as one of the main symptoms of this infectious illness, presenting as difficult coughing, a dry cough, and chest pain up to 14 days after exposure to the virus. Clinical testing projects that incorporate structural and physiological (Huber & Stathopoulos, 2015) improvements in the unpredictable respiratory system are speech breathing models. Based on our observations, we believe that speech signals might blame the shift in COVID 19 detection.

Bringing together an enormous data set of breathing sounds and respiratory disease skills from clinical experts can evaluate the expected effect of utilizing breath sounds to recognize COVID-19 indications using deep learning methods (Thorpe et al., 2021). This work's primary purpose is to supplement existing clinical testing methods by replacing them with low cost, fast process, and high accuracy. This research work provides efforts in this direction.

1.1 Dataset

First, to generate data on healthy and unhealthy sound samples, including COVID-19 identification. The generated samples are analyzed using the proposed generative adversarial network method. It has built on assistive mathematical models that identify biomarkers from sound models. Progress should be made when creating task data at this stage.

1.2 Literature survey

Several studies have proposed sound features that detect symptoms and vocal signals in respiratory diseases in recent years.

As the examination has focused on expanded COVID 19, ongoing works have started researching the utilization of deep neural networks by people to characterize sick dependent on cough sounds. Venkata Srikanth and Strik (2019) use Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architectures for breath occasion discovery as a likely pointer of COVID-19 recognition. As of late, Basheer et al. (2020) used the CNN architecture to perform direct COVID-19 symptomatic groupings dependent on cough sounds. The work in Chon et al. (2012) uses a learning step technique of deep finding out how to do a similar analysis to our own, with an F1 score of 0.929, which is not at all like the methods discussed in this article.

More recently, microphones in devices, for example, cell phones and wearable devices, have been abused for voice examination. In Rachuri et al. (2010), the microphone audio is utilized to comprehend the client's current circumstance. This data is assembled to briefly look at the environmental factors in places around the city alone. In COVID-19 recognition (Nandakumar et al., 2015), a sensor recognizes clients' feelings through the telephone's receiver wild Gaussian compound models. In Oletic and Bilas (2016), Pramono et al. (2017), Praveen Sundar et al. (2020), the authors distinguished COVID-19 in the investigation using sound samples based on different machine learning methods.

2 Proposed COVID-19 detection using speech signal

The generative adversarial network with speech signal-based COVID-19 detection system is shown in Fig. 2. The proposed system consists of two stages, pre-processing and classification. The Least Mean Square filter removes the artifacts or noise from the input speech signal in the pre-processing step. After completing the pre-processing process, the GAN classifier analyses the filtering signal to classify COVID-19 and non-COVID-19 signals.

2.1 Noise reduction using LMS

Typically, all biomedical signals contain noise or artifacts. Hence, before classifying the signals, we need to remove the noise or artifacts for accurate results. In this research work, the Least-Mean-Square (LMS) filtering method is used to remove the noise. As compared with other filters, the LMS decreases the variance of weights to stabilize the signal using the Lagrangian approach. This Lagrangian method has a nonlinear transformation rule, and it differentiates the input and output derivatives, which solves the optimization problem of the LMS algorithm. The LMS pre-processing steps are discussed below.

2.1.1 LMS algorithm

Parameters: M = Filter length μ = Step size Initialization, W= weight:
W (0) = 0
Computation: For n = 1, 2, 3, · · ·
i) Filter output y(n) = w T (n)x(n)
ii) Error output
$$e(n) = d(n) - y(n) \quad .... (1)$$
iii) Weight update equation
$$w(n + 1) = w(n) + \mu e(n)x(n) \quad .... (2)$$

The optimization issues is overcome using the strategy of Lagrange multipliers. The equation of Lagrangian is given in Eq. (3).

$$\mathrm{L}(\mathrm{w}(\mathrm{n}+1)) = ||\S\mathrm{w}(\mathrm{n}+1)||^2 + \mathrm{Re}\left[\lambda * \mathrm{e}^{[\mathrm{n}+1]}(\mathrm{n})\right] \tag{3}$$

where w(n + 1) = tap weight vector, §w(n + 1) = w(n + 1) – w(n) in the tap-weight vector w (n + 1) with respect to its old worth w(n).

Here λ* is known as the Lagrange multiplier, in this way getting the famous variation rule in (3) with the standardized advance size gave by $\mu = \hat{\mu}/||x(n)||^2$. The last restriction is unnecessarily obstructive in operational applications; therefore, an additional interesting solution is derived when we relax it.

2.2 GAN classifier

This section discusses the Generative Adversarial Network method's working function based on COVID-19 detection from the speech signal. The optimal threshold value of COVID-19 is above 1.2 Hz, and non-COVID-19 is below 0.60 Hz. The investigation model's unsupervised learning piece is developed for the Deep Convolution Generative Adversarial Network (GAN) design or DCGAN.DCGAN contains two main blocks known as generators and discriminators, and these blocks are trained using min–max arrangement. The Generator receives the samples from random distributions variance of output conditions. The discriminator takes samples from either the output of the generator or actual speech samples from the dataset. During training, the discriminator utilizes the cross-entropy loss function to distinguish the number of classified models completely in genuine models, and the Generator classifies the number of good ones. The mathematical calculation of real (y) and predicted ($\hat{y}$) values are defined in Eq. (4).

$$L(w) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y}_n + \left(1 - y_n\right)\log\left(1 - \hat{y}_n\right)\right] \tag{4}$$

where w = weights of learned vectors, N = size of samples.

For this calculation, 1 represents the real sample, and 0 represents the generated samples. The prediction of discriminator ($\hat{y}_r$) is computed using Eq. (5).

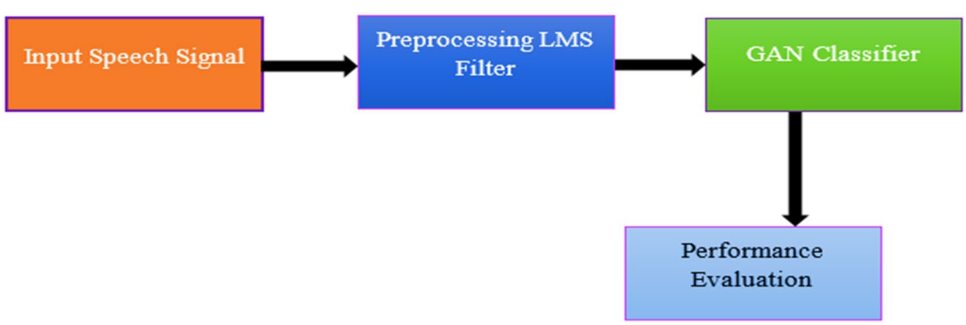


Fig. 2 Block diagram of COVID-19 detection

$$L_r(W) = -\frac{1}{N}\sum_{n=1}^{N} \log \hat{y}_r, n \tag{5}$$

All the correct predictions are considered as zero for this case. Similarly, the $\hat{y}_g$ discrimination represents prediction. Therefore, the correct prediction of the cross-entropy function is simplified by using Eq. (6)

$$L_f(W) = -\frac{1}{N}\sum_{n=1}^{N} 1 - \log \hat{y}_g, n \tag{6}$$

The generator also uses cross-entropy loss, which should be interpreted in terms of fallen generator outputs into the real sample. The cross-entropy loss of the Generator is computed using Eq. (7).

$$L_g(W) = -\frac{1}{N}\sum_{n=1}^{N} \log \hat{y}_g, n \tag{7}$$

If the generator has low loss, the proposed system gives the discriminator results as accurate.

This process leads the Generator to produce output and looks like an actual sample of well-trained iterations shown in Fig. 3. Both the activation of the valence classifier cross-entropy misfortune function to reduce the loss. The cross-entropy function is discussed by Eq. (7): the valence activation classifier network, and the discrimination share layer model, which learns the characteristics. The convolution filter is effectively used for the valence classification task to activate the classification network to distinguish between actual and generated speech samples.

Figure 4 discusses the overall process for describing the proposed Deep Convolution Generative Adversarial Network with record cough-breath sound, extract audio features, split the training/testing ratio, and performance validation. The testing and training ratio is 80:20. The classification response of the proposed COVID-19 detection system's performance is validated using precision, recall, and accuracy. Compared to other deep learning methods, GAN does not require labeled data; they can be trained using unlabeled data to learn the data's internal representations. So the performance is automatically improved.

Precision It is the fraction of relevant speech samples among the retrieved speech samples. The mathematical formula of precision is shown in Eq. (8).

$$\text{Precision } (P) = \frac{T_p}{T_p + F_p} \tag{8}$$

Recall It is the fraction of retrieved relevant speech samples among all relevant speech samples. The mathematical formula of recall is shown in Eq. (9).

$$\text{Recall (R)} = \frac{T_p}{T_p + F_n} \tag{9}$$

Accuracy Accuracy is the ratio of correctly classify the COVID-19 samples from the total number of samples. The following Eq. (10) is used to compute the accuracy.

$$\text{Accuracy} = \frac{T_p + T_n}{\left(T_p + T_n + F_p + F_n\right)} \tag{10}$$

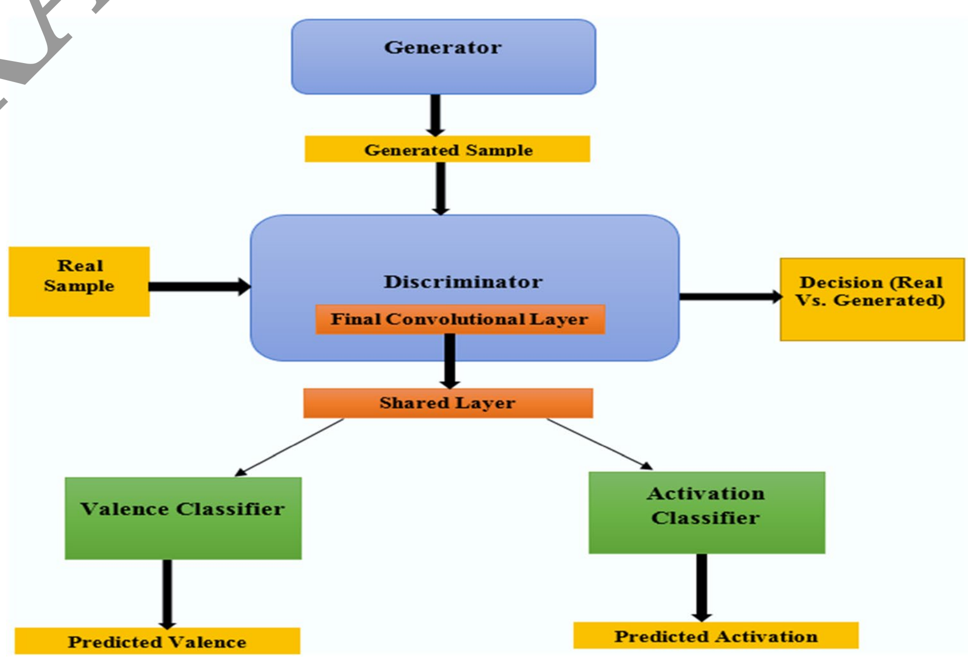


Fig. 3 The architecture of GAN classifier

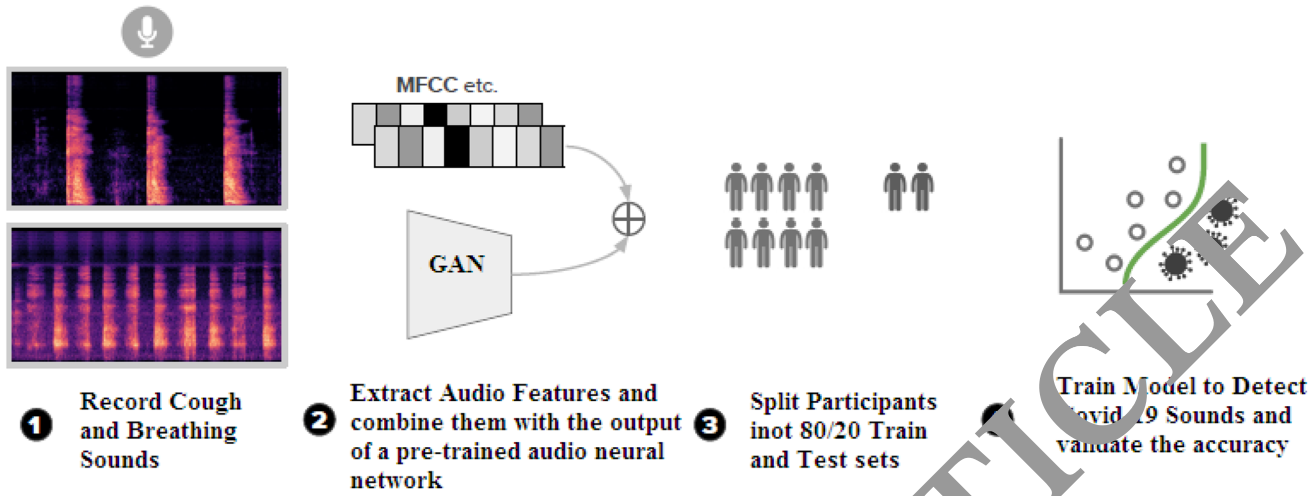


Fig. 4 Overall process of proposed method

where T_p = true positive, T_n = true negative, F_p = false positive, F_n = false negative.

3 Simulation results and discussion

Simulation results and performance analysis of the proposed COVID 19 detection system are discussed in this section. This work aims to classify speech samples from normal and abnormal people, include to identifying COVID-19 patients.

The input speech signal of the proposed COVID-19 detection is depicted in Fig. 5. The input signal's frequency range is 8 kHz.

Time-domain representation of proposed Generative Adversarial Neural Network-based COVID-19 detection is shown in Fig. 6.

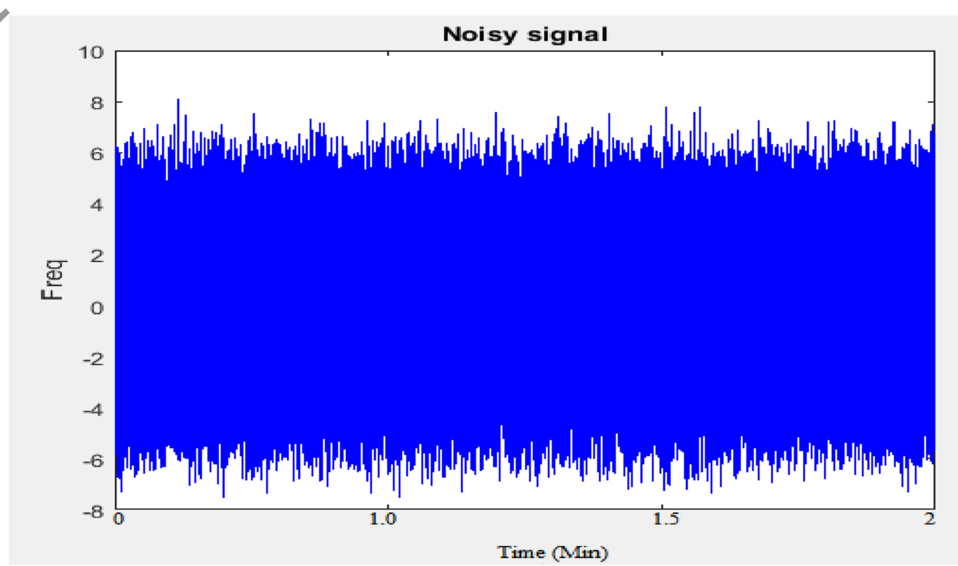
The proposed Generative Adversarial Neural Network-based time-domain representation of the noise signal of COVID-19 detection is shown in Fig. 7.

The proposed Generative Adversarial Neural Network-based time and frequency response of the filtered signal COVID-19 detection is shown in Fig. 8.

Figure 9 shows the Spectrogram of the pre-processed speech signal. The Spectrogram splits the Window that allows overlapping elements in each section with windows notation.

Figure 10 shows the simulation results of validation accuracy and loss in training. The proposed COVID-19 detection

Fig. 5 Noisy signal



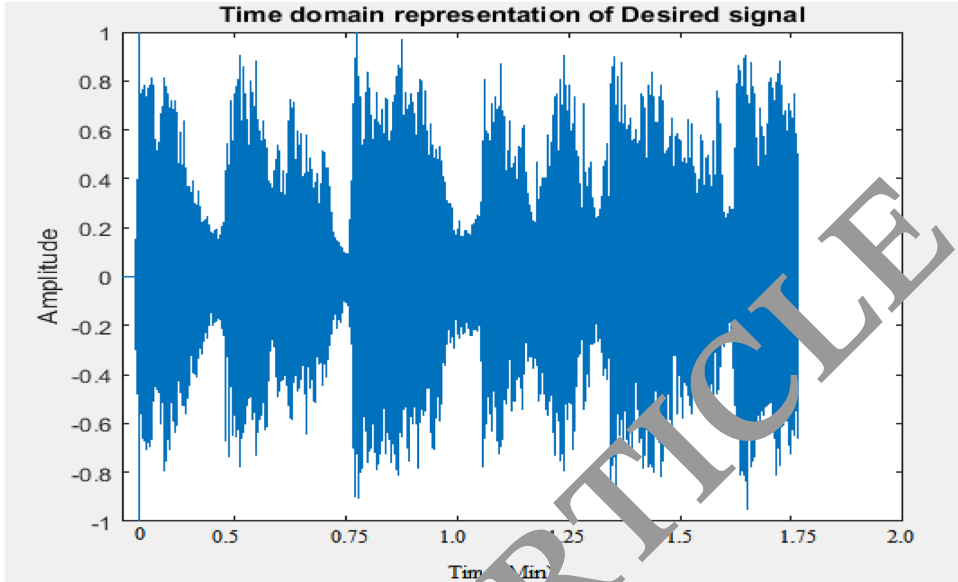


Fig. 6 Time domain representation of the desired signal

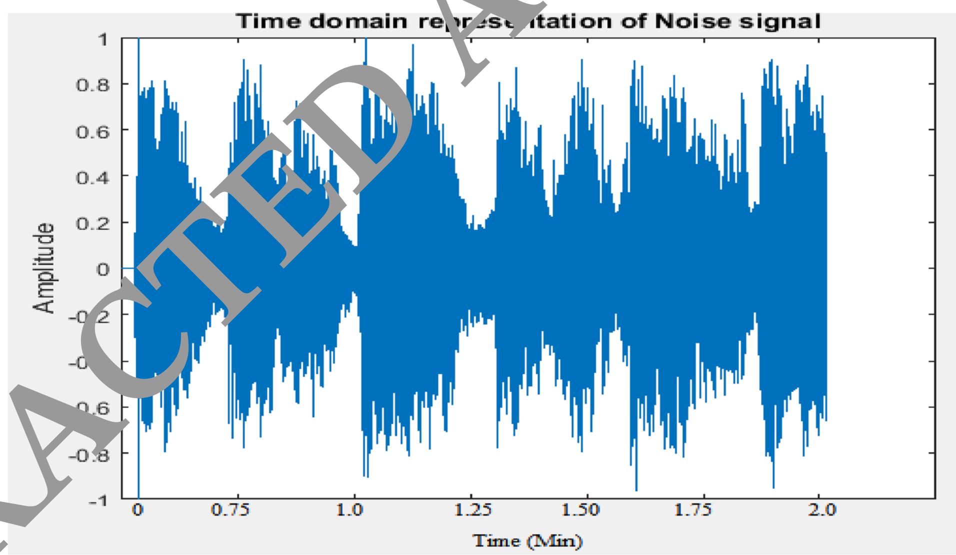


Fig. 7 Time domain representation of noise signal

system reduces the validation loss and increases the validation accuracy, making the model learning low mean squared error.

Figure 11 and Table 1 discuss the performance analysis of the proposed COVID-19 classification system with existing methods. As compared with existing methods, the proposed GAN method achieves a good result. The precision, recall, accuracy and F-measure are 96.54%, 96.15%, 98.56% and 0.96% respectively.

4 Conclusion

This research work introduces Generative Adversarial Network for the detection of COVID-19 symptoms from a speech signal. Typically, speech signals contain intrinsic information regarding the physiological as well as emotional conditions of humans. Accurate measurement of such physiological parameters using speech signals has facilitated real-time, remote monitoring of infected/symptomatic individuals and early detection of COVID-19 symptoms, resulting in containing the spread of the

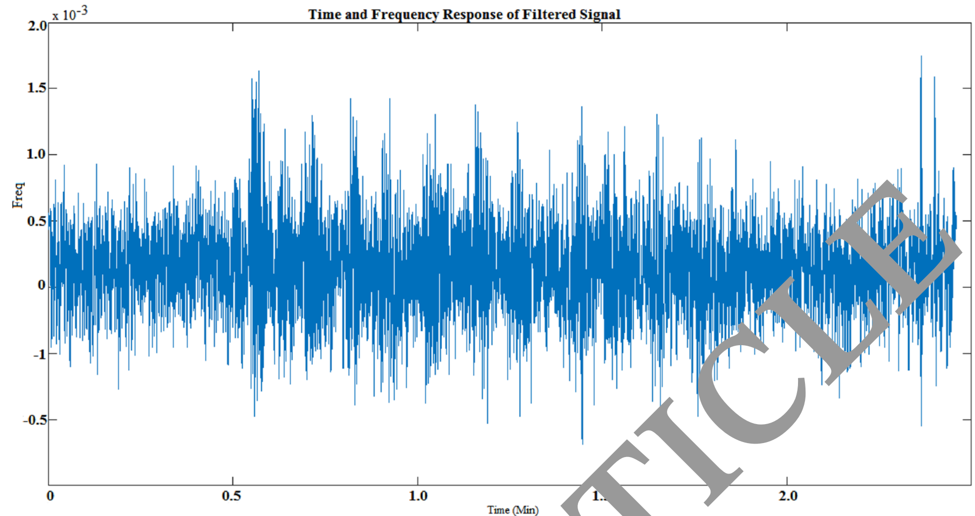


Fig. 8 Time and frequency response of a filtered signal

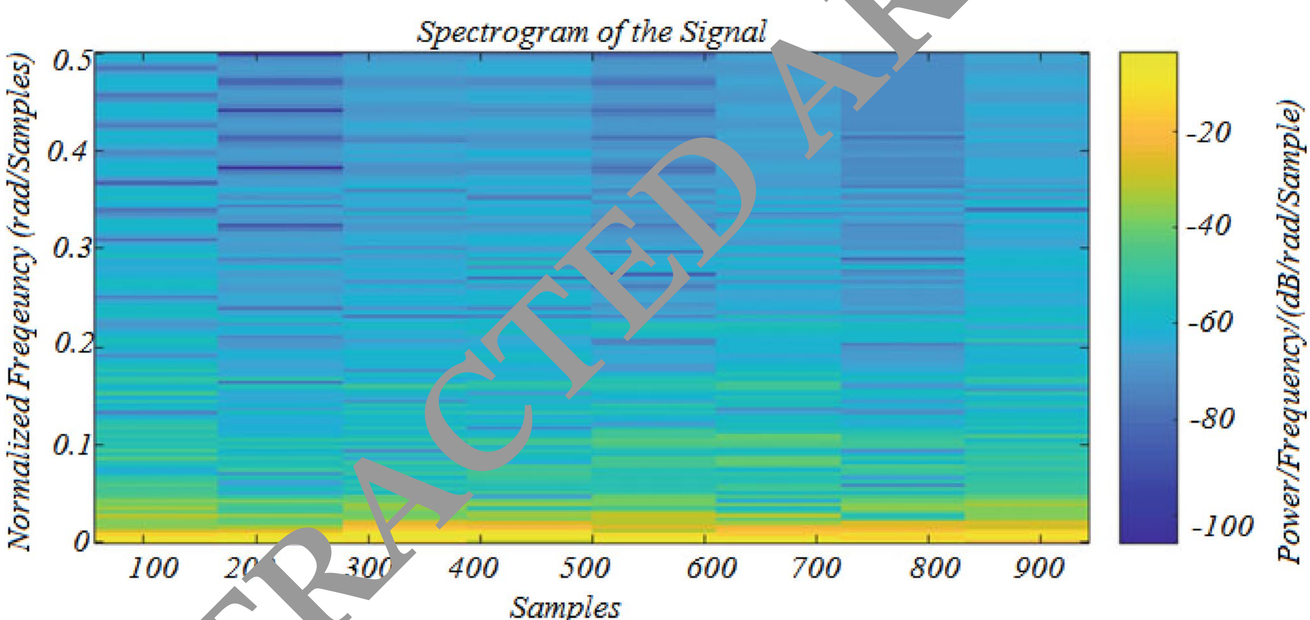


Fig. 9 Spectrogram of a speech signal

infection. The reverse transcription-polymerase chain reaction (RT-PCR) testing strategy is used to determine the COVID-19 virus. This RT-PCR processing is more expensive and inducing social distancing rules violation, and time-consuming. Therefore, this research work introduces the Generative Adversarial Network (GAN) based deep learning method to detect COVID-19 from speech signals quickly. As compared with existing methods, the proposed GAN method achieves a good result. The precision, recall, accuracy, and F-measure are 96.54%, 96.15%, 98.56%, and 0.96, respectively.

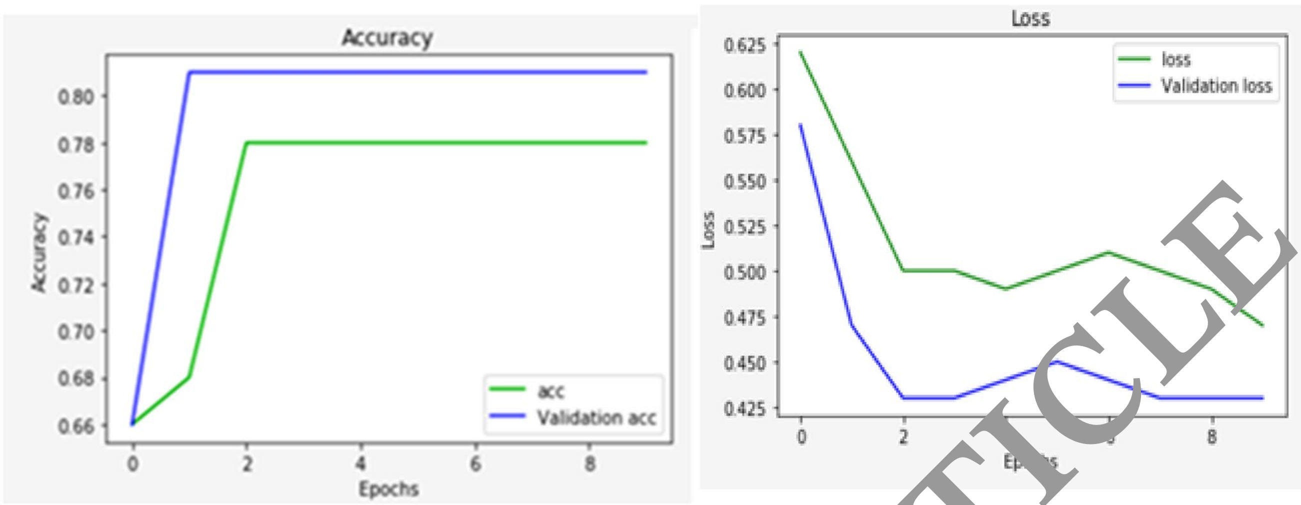


Fig. 10 Validation accuracy and loss during the training

Fig. 11 Performance analysis of classification ratio

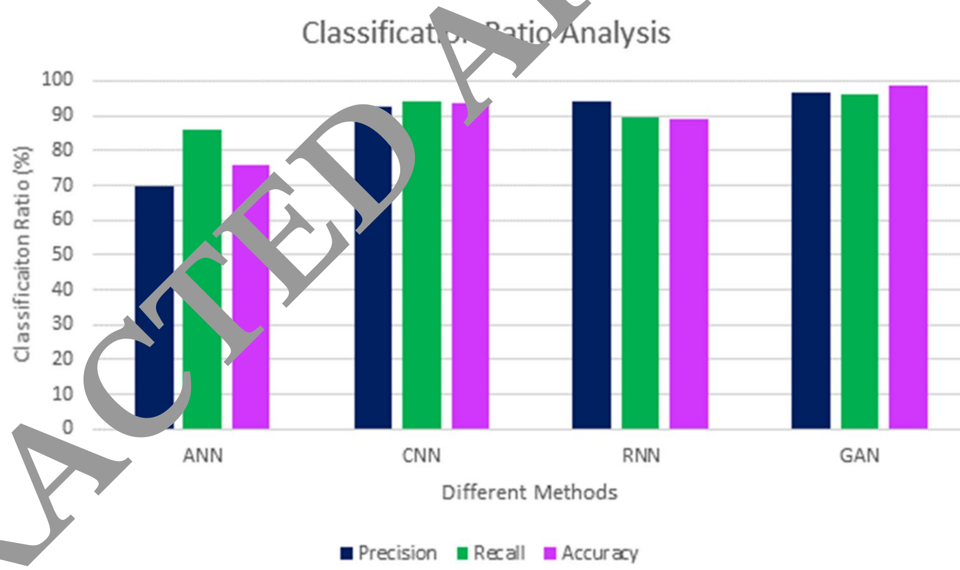


Table 1 Performance evaluation of classification ratio

Method	Precision (%)	Recall (%)	Accuracy (%)	F-measure (%)
ANN	[illegible]	86.10	75.883	0.86
CNN	92.65	94.12	93.47	0.89
RNN	94.16	89.65	89.13	0.91
GAN	96.54	96.15	98.56	0.97

References

Basheer, S., Anbarasi, M., Sakshi, D. G., et al. (2020). Efficient text summarization method for blind people using text mining techniques. *International Journal of Speech Technology, 23*, 713–725. https://doi.org/10.1007/s10772-020-09712-z

Botha, G., Theron, G., Warren, R., Klopper, M., Dheda, K., Van Helden, P., & Niesler, T. (2018). Detection of tuberculosis by automatic cough sound analysis. *Physiological Measurement*. https://doi.org/10.1088/1361-6579/aab6d0

Breathing sounds for COVID-19. Retrieved from May 8, 2020, from https://breatheforscience.com/

Chon, Y., Lane, N. D., Li, F., Cha, H., & Zhao, F. (2012). Automatically characterizing places with opportunistic crowdsensing using smartphones. In: *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp)*. Pittsburgh, PA, pp. 481–490.

Ghosh, S., Laksana, E., Morency, L.-P., & Scherer, S. (2015). *Learning representations of effect from speech*. CoRR, vol. abs/1511.04747.

Ghosh, S., Laksana, E., Morency, L.-P., & Scherer, S. (2016a). Representation learning for speech emotion recognition. In: *Proceedings of Interspeech* 2016.

Ghosh, S., Laksana, E., Morency, L.-P., & Scherer, S. (2016b). An unsupervised approach to glottalin verse filtering. In: *Proceedings of EUSIPCO* 2016.

Huber, J. E., & Stathopoulos, E. T. (2015). *Speech Breathing Across the Life Span and in Disease, Ch. 2* (pp. 11–33). Wiley.

Indian Institute of science—Coswara: A sound-based diagnostic tool for covid19. Retrieved from May 8, 2020, from https://coswara.iisc.ac.in/

James, A. P. (2015). Heart rate monitoring using human speech spectral features. *Human-Centric Computing and Information Sciences, 5*(1), 1–12.

McKeown, G., Valstar, M., Cowie, R., Pantic, M., & Schroder, M. (2012). The Semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing, 3*(1), 5–17.

Menni, C., Valdes, A. M., Freidin, M. B., Sudre, C. H., Nguyen, L. H., Drew, D. A., Ganesh, S., Varsavsky, T., Cardoso, M. J., El-Sayed Moustafa, J. S., Visconti, A., Hysi, P., Bowyer, R. C. E., Mangino, M., Falchi, M., Wolf, J., Ourselin, S., Chan, A. T., Steves, C. J., & Spector, T. D. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*. https://doi.org/10.1038/s41591-020-0916-2

Nandakumar, R., Gollakota, S., & Watson, N. (2015). Contactless sleep apnea detection on smartphones. In: *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. Florence, Italy, pp. 45–57.

Oletic, D, & Bilas, V. (2016). Energy-efficient respiratory sounds are sensing for personal mobile asthma monitoring. *IEEE Sensors Journal, 16*(23), 8295–8303.

Porter, P., Abeyratne, U., Swarnkar, V., Tan, J., Ng, T.-W., Brisbane, J. M., Speldewinde, D., Choveaux, J., Sharan, R., Kosasih, K., et al. (2019). A prospective multicentrestudy was testing the diagnostic accuracy of an automated cough sound centered analytic system for the identification of common respiratory disorders in children. *Respiratory Research, 20*(1), 81.

Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0177926

Praveen Sundar, P. V., Ranjith, D., Karthikeyan, T., et al. (2020). Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture. *International Journal of Speech Technology, 23*, 287–296. https://doi.org/10.1007/s10772-020-09686-y

Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow. P. J., Longworth, C., & Aucinas, A. (2010). EmotionSense: A mobile phones-based adaptive platform for experimental social psychology research. In: *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp)*. Copenhagen, Denmark, pp. 281–290.

Sakai, M. (2015). Modeling the relationship between heart rate and features of vocal frequency. *International Journal of Computer Applications, 120*(6), 32–37.

Schuller, B., Friedmann, F., Eyben, F. (2014). The Munich biovoice Corpus: effects of physical exercising, heart rate and skin conductance on human speech production. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation*, Reykjavik, Iceland, 26–31 May 2014, pp. 1506–1510.

Thorpe, W., Kurver, M., King, G., & Salome, C. (2001). Acoustic analysis of cough. In: *Proceedings of the seventh Australian and New Zealand Intelligent Information Systems Conference*. IEEE, pp. 391–394.

Trouvain, J., & Truong, K. P. (2015). Prosodic characteristics of reading speech before and after treadmill running. In: *16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, September 6–10, 2015.

Usman, M. (2017). On the performance degradation of speaker recognition system due to variation in speech characteristics caused by physiological changes. *International Journal of Computing and Digital Systems (IJCDS), 6*(3), 119–126.

Venkata Srikanth, N., & Strik, H. (2019). *Deep sensing of breathing signal during conversational speech.*

Windmon, A., Minakshi, M., Bharti, P., Chellappan, S., Johansson, M., Jenkins, B. A., & Athilingam, P. R. (2018). Tussiswatch: A smartphone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and congestive heart failure. *IEEE Journal of Biomedical and Health Informatics, 23*(4), 1566–1573.

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