

Research Article

Speech as a Biomarker for COVID-19 Detection Using Machine Learning

Mohammed Usman ¹, Vinit Kumar Gunjan ², Mohd Wajid ³, Mohammed Zubair ¹,
and Kazy Noor-e-alam Siddiquee ⁴

¹Department of Electrical Engineering, King Khalid University, Abha 61411, Saudi Arabia

²Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India

³Department of Electronics Engineering, ZHCET, Aligarh Muslim University, Aligarh 202002, India

⁴Department of Computer Science and Engineering, University of Science and Technology, Chittagong, Bangladesh

Correspondence should be addressed to Kazy Noor-e-alam Siddiquee; knalam@ustc.ac.bd

Received 15 February 2022; Revised 7 March 2022; Accepted 21 March 2022; Published 18 April 2022

Academic Editor: Arpit Bhardwaj

Copyright © 2022 Mohammed Usman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The use of speech as a biomedical signal for diagnosing COVID-19 is investigated using statistical analysis of speech spectral features and classification algorithms based on machine learning. It is established that spectral features of speech, obtained by computing the short-time Fourier Transform (STFT), get altered in a statistical sense as a result of physiological changes. These spectral features are then used as input features to machine learning-based classification algorithms to classify them as coming from a COVID-19 positive individual or not. Speech samples from healthy as well as “asymptomatic” COVID-19 positive individuals have been used in this study. It is shown that the RMS error of statistical distribution fitting is higher in the case of speech samples of COVID-19 positive speech samples as compared to the speech samples of healthy individuals. Five state-of-the-art machine learning classification algorithms have also been analyzed, and the performance evaluation metrics of these algorithms are also presented. The tuning of machine learning model parameters is done so as to minimize the misclassification of COVID-19 positive individuals as being COVID-19 negative since the cost associated with this misclassification is higher than the opposite misclassification. The best performance in terms of the “recall” metric is observed for the Decision Forest algorithm which gives a recall value of 0.7892.

1. Introduction

The most basic functions of the human body are usually monitored by measuring the vital signs—temperature, heart (pulse) rate, respiratory (breathing rate), and blood pressure [1]. These are usually measured using medical devices, but nowadays, easy to use and low-cost devices and smart gadgets are available which allow measuring temperature, pulse rate, and blood pressure at home, even by nonmedical professionals. Various types of sensors, present in such devices, sense some signal generated by the human body, process the signal, and provide a reading of the vital sign in a simple and easy to interpret format. It is therefore necessary

to connect the devices/sensors at appropriate points on the human body to obtain the desired measurements. The placement of such sensors on the body is invasive and intrusive and causes inconvenience to the patient/individual. This is particularly true when monitoring professional athletes and sports persons while they are performing intense exercise/training. Moreover, it does not allow measurement of body parameters without attaching the device/its sensors (probes) to the body or from a remote location; i.e., the patient has to be at the same location as the medical device. This article investigates the use of speech as a biomedical signal to detect COVID-19 based on a statistical analysis of speech and binary classification using machine

learning. Speech characteristics of an individual get altered as a result of physiological and emotional changes [2–6]. Other factors that can cause physiological changes in the body are changes in health conditions, aging, stress, pollution exposure, and physical activity. There is significant evidence from the literature that clearly establishes a correlation between the characteristics of human speech and the physiological parameters of the speaker. A correlation between heart-related parameters such as heart rate, electrocardiogram (ECG) features, and the influence of heart function on speech characteristics has been shown in [7–13]. In [14], the variation of speech characteristics due to physical activity is demonstrated, and the effect of physical activity and fitness level on heart rate is shown in [15]. Physiological changes due to physical activity also depend on the regularity, duration, and intensity of the activity performed [16]. A correlation between speech and blood pressure is established in [17], and a method to detect emotions from speech is described in [15]. It is also established in the literature that tiredness can also affect an individual's speech and can cause speech to become slurred (dysarthria) [18]. Noncontact methods to measure physiological parameters based on image and video processing have also been investigated and reported in the literature [19,20]. Noncontact methods based on speech, images, and video can facilitate remote monitoring, telemedicine, and smart healthcare which are expected to play a major role in future healthcare infrastructures.

The COVID-19 pandemic era has necessitated and triggered an enormous amount of research into such noncontact-based diagnostic methods to detect COVID-19 using machine learning and deep learning [21–30]. A review of COVID-19 diagnostic methods along with prevention tools and policy decisions for COVID-19 management is provided in [31]. Artificial intelligence-based COVID-19 diagnosis tools are not without pitfalls. A critique of AI-based tools being given emergency authorization by regulatory bodies indicates that many such tools have been developed using small or low-quality datasets [32], concluding that AI could be useful in dealing with the COVID-19 pandemic but requires more detailed investigation and validation. Several research studies based on artificial intelligence are ongoing not only to detect COVID-19 but also to predict and understand the effects of the pandemic and be prepared for eventualities. A method to detect COVID-19 using machine learning on symptoms is proposed in [33]. COVID-19 detection based on the application of AI on X-ray and computerized tomography (CT) images is reported in [34–41]. In [42–44], detection of COVID-19 by applying machine learning to routine blood examination data has also been reported. AI-based systems to predict the deterioration of COVID-19 patients toward severe disease have been presented in [45–47], and prediction of mortality risk among COVID-19 infected individuals using AI is also available in the literature [48–50]. The COVID-19 pandemic has indeed necessitated and highlighted the need for interdisciplinary and transdisciplinary approaches to diagnose, treat, and manage not just COVID-19 patients but also to address medical problems in general [51]. The challenges

involved in the use of AI for COVID-19 are elaborated in [52]. Several research groups [53–56] are actively investigating the use of speech sounds, cough sounds, and breathing/respiratory sounds to detect COVID-19 by analyzing these sounds using artificial intelligence algorithms.

2. Materials and Methods

2.1. Data Used in This Study. Speech recordings used in this study comprise two categories—speech from healthy individuals with no known preexisting medical conditions at resting heart rate and speech from asymptomatic COVID-19 positive individuals. Heart rate and blood oxygen saturation level (SpO₂) are measured simultaneously at the time of recording the speech using an off-the-shelf pulse oximeter. It should be noted that the pulse rate measured by the pulse oximeter is exactly equal to the heart rate [57]. The total number of speech recordings of healthy individuals at resting heart rate is 84. All the healthy volunteers are in the age group of 25–45 years. Speech samples along with heart rate and SpO₂ measurements were also obtained from 22 individuals who had tested COVID-19 positive following contact tracing, but with no conspicuous symptoms. The youngest in this category is 32 years and the oldest is 57 years. It should be noted that obtaining speech data of COVID-19 patients was challenging and hence the relatively small set of samples.

Speech recording was made using a Logitech headphone equipped with a noise cancellation microphone. While the samples of all the healthy individuals were recorded using the same microphone in the same environment, the speech samples of COVID-19 individuals were recorded with different microphones of the same make and model (Logitech H540) and under different ambient conditions for each. Hence, any variations in speech characteristics arising due to the difference in recording device and ambiance are not taken into consideration. It is reasonable to ignore these variations since the recordings were made in quiet rooms using microphones of the same make and model and therefore have the same technical specifications. Of course, the acoustic effects of the room and background noise, albeit small, are not taken into consideration as it was not possible to bring the COVID-19 patients to the laboratory settings where the recording of healthy individuals was made. Each individual was asked to read the sentence “A quick brown fox jumped over the lazy dogs” which was recorded by turning the microphone “ON”: for 5 seconds. The recording was made in stereo format at a sampling rate of 16000 samples per second (sps) which is the standard sampling rate for wideband representation of speech [58]. The recording is quantized using 2¹⁶ quantization levels resulting in an audio bit rate of 256 kbps and stored on a computer in uncompressed .WAV format. Heart rate and SpO₂ level are also concurrently measured at the time of speech recording using a pulse oximeter. The attributes of the data are highlighted in Table 1.

2.2. Preprocessing of Speech Data. Unwanted components such as DC bias, which usually gets introduced by PC audio cards [59] and silence intervals due to pauses made by the

TABLE 1: Attributes of data used in this study.

	Age group (years)	No. of recordings	Sampling rate (sps)	Quantization depth (bits)	Audio bit rate (kbps)	Audio format	Other parameters measured
Healthy	25–45	84	16000	16	256	.wav	Heart rate, SpO ₂
COVID+	32–57	22					

speaker, are removed by preprocessing each of the speech recordings. DC bias is removed using a 1st-order infinite impulse response (IIR) filter, whereas silence intervals are removed by applying a voice activity detection (VAD) mechanism which extracts speech frames containing voice activity. VAD also mitigates noise effects by applying a posteriori signal to noise ratio (SNR) weighting to emphasize reliable segments of voice activity even at low SNR. DC bias removal and VAD are applied as per the implementation provided in [60]. A block diagram of preprocessing steps is shown in Figure 1.

The features of speech that have been used in this study are the short-term Fourier Transform (STFT) coefficients. Features are defined as characteristics of a signal that enables some algorithm to detect an inherent pattern associated with the signal [61]. The premise of detecting COVID-19 from speech features stems from the fact that speech is produced by moving air from the lungs through the vocal cavity. Since there is an interaction between the lungs and heart for the oxygenation of the blood, cardiovascular responses are influenced by activities such as reading and speaking [62]. It is shown in [63] that breathing pattern is affected by the process of speech production. Changes in breathing patterns, in turn, have an effect on the heart rate, and this effect is termed respiratory sinus arrhythmia (RSA) [64]. Several techniques for feature extraction have been proposed in the literature for various applications but predominantly for speaker/speech recognition and speech enhancement. Linear prediction coefficients (LPC), linear prediction cepstral coefficients (LPCC), perceptual linear prediction (PLP), Mel frequency cepstral coefficients (MFCC), Mel frequency discrete wavelet coefficients (MFDWC), feature extraction using principal component analysis (PCA), and wavelets based features are some of the common features that have been reported in the literature [65–70]. STFT represents the time-varying spectral properties of a signal, and for this study, STFT coefficients with a high spectral resolution are used in order to capture subtle differences between closely spaced frequency components. The high spectral resolution is achieved by computing the STFT of long segments of speech, i.e., segment length greater than 500 ms. The high spectral resolution is achieved at the expense of temporal resolution. Since a correlation between physiological parameters and spectral features of speech is evident from existing literature, STFT coefficients with high spectral resolution have been used in this study. The STFT coefficients are used as input features to machine learning algorithms to classify the speech signal as that of COVID-19 positive or not.

2.3. Statistical Modeling of Speech Features. The most common symptoms of COVID-19 are fever, tiredness, and dry cough, and these may not be conspicuous until about 14

days after getting infected with an average of 5–6 days for the symptoms to become conspicuous. In this article, it is shown using statistical modeling of speech features that it is possible to detect COVID-19 from an individual’s speech much before the symptoms become conspicuous so that the person can be quarantined, tested, and provided with medical support at an early stage. At their onset, while symptoms may not be conspicuous to the affected individual or to observers, physiological changes occur in the individual that cause variations in speech characteristics which can be analyzed by artificial intelligence (AI) algorithms. Signal processing and AI can be applied to speech to detect physiological changes which have a direct or indirect relation to one or more of the COVID-19 symptoms. The existence of a correlation between speech characteristics and physiological, psychological, and emotional conditions is well established in the literature. It is therefore possible to detect COVID-19 infection from speech samples of individuals and this possibility is investigated in this article. The relationship between the most common symptoms of COVID-19 and affected physiological parameters is illustrated in Figure 2.

A statistical analysis of speech spectral features is performed by applying maximum likelihood estimation (MLE) to obtain the best statistical distribution along with the distribution parameters that best characterize speech STFT coefficients, statistically. It has been shown in [58] that for speech samples at resting heart rate, STFT coefficients having high spectral resolution are accurately modeled by a Laplacian distribution (LD) with the estimated LD parameters exhibiting small RMS error. The Laplacian distribution is defined as

$$p(\mathbf{x}) = \frac{1}{2b} \exp\left(\frac{-|\mathbf{x} - \boldsymbol{\mu}|}{b}\right) \quad (1)$$

parameter and $b > 0$ is the scale parameter. The procedure to estimate μ and b , the RMS error associated with the estimation of b , and its lower bound defined as the Cramer-Rao bound (CRB) are provided in [58].

2.4. Binary Classification of Speech Samples Using Machine Learning. From the statistical analysis of speech STFT coefficients at the high spectral resolution, it is evident that the RMS error of fitted LD increases as a result of COVID-19 infection. Based on this finding, binary classification of speech signals as COVID-19 positive or COVID-19 negative is investigated by using STFT coefficients as input features to machine learning algorithms. In order to train and develop the AI models, speech samples of healthy as well as COVID-19 positive individuals are used. The trained AI model can

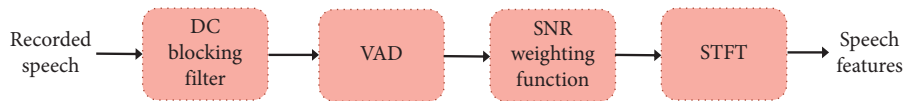


FIGURE 1: Speech preprocessing.

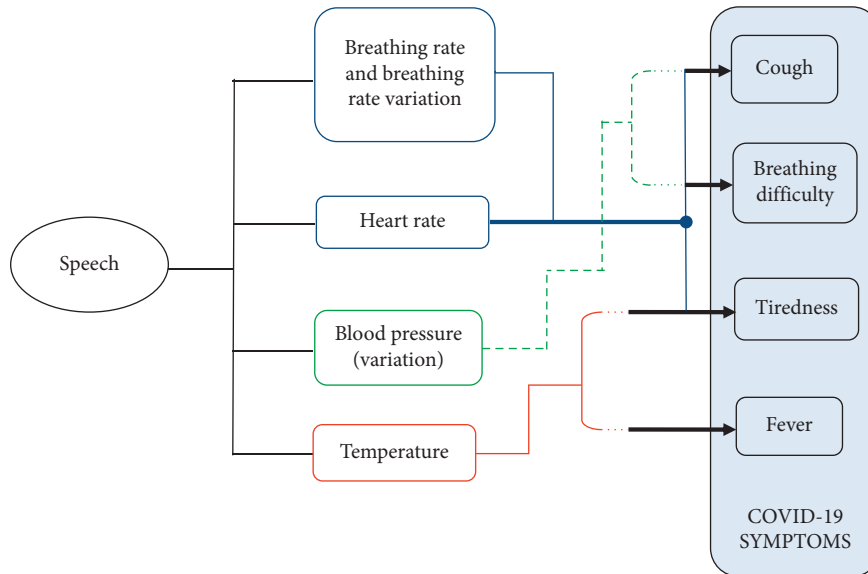


FIGURE 2: Biological parameters correlated to speech and COVID-19 symptoms.

then be incorporated into a mobile “app” for early detection of COVID-19, once the desired level of accuracy is achieved and regulatory approvals are obtained. If speech can be used to detect COVID-19, the functionality of the “smartphone” which already has wide proliferation and ubiquitous presence can be extended to alleviate the challenges posed by the pandemic. The results reported in the literature [71–78] are quite promising, providing exciting and interesting answers, giving confidence that research on this topic can lead to the development of mobile applications which can be used not only to detect COVID-19 from human sounds but also for other medical diagnostic/monitoring purposes. COVID-19 diagnosis using only cough recordings is presented in [71]. However, it uses biomarker information such as muscular degradation, vocal cords, sentiment, and lungs/respiratory tract function along with the cough recordings for diagnosis. The relation between COVID-19 symptoms and respiratory system function is highlighted in [72] along with a survey of AI-based COVID-19 diagnoses using human audio signals. Cough and respiratory sounds are used to classify COVID-19 and non-COVID-19 individuals in [73]. Furthermore, it is shown that cough from COVID-19 can be distinguished from healthy individuals’ cough as well as cough of asthmatic patients. A project in progress [74] investigates the detection of COVID-19 from human audio sounds using AI. A news feature article in Nature [75] highlights research interest and progress among academic as well as commercial organizations to use the human voice for various diagnostic purposes including COVID-19. AI4COVID-19 is an app that runs an AI algorithm in the cloud to detect COVID-19

from cough sounds and reports promising results, encouraging further collection of labeled cough sounds [76]. An overview of the possibilities, challenges, and use cases of computer audition is presented in [77], which clearly highlights the potential of using sound analysis using AI for COVID-19 diagnosis. A support vector machine (SVM) based method to detect COVID-19 from speech signals is presented in [78] which combines voice signals and symptoms reported by the patient. In contrast to the research available in the literature, this article uses only speech signals without any side information such as symptoms or other biomarker information.

The STFT coefficients are labeled as coming from the speech of COVID-19 negative, i.e., healthy (Class 0) and COVID-19 positive (Class 1) individuals. Microsoft Azure Machine Learning Studio (MAMLS) cloud platform is used in this study to perform binary classification, and the performance of classification is analyzed and compared for five state-of-the-art classification algorithms available in MAMLS. Machine learning techniques produce a model for the data by learning the statistical relationship between input data (e.g., STFT coefficients extracted from speech signals) and output data (e.g., class label). The hyperparameters of the produced model are tuned optimally to minimize the classification error in an independent test dataset, resulting in a generalized model that can perform well on the test data set as well. The tuning is performed manually by adjusting the model hyperparameters until the highest value for the “recall” metric is achieved. Good performance on only the training dataset would result in an overfitting solution. A

brief description of the five algorithms used for binary classification is provided here for completeness. The block diagram of the methodology used in the work is shown in Figure 3. An overview of the used ML algorithms follows.

Boosted Decision Tree (BDT) is an ensemble learning technique, wherein the succeeding tree corrects the errors of the previous tree to minimize classification error. The complete ensemble of trees is used for correctly predicting the binary class to which the input data belongs [79]. Another classification algorithm based on ensemble learning is the Decision Forest (DF) algorithm, wherein the most popular class is selected depending on the vote from each of the generated trees [80]. Neural Networks (NNs) are a network of interconnected layers of processing units called neurons. A typical NN consists of neurons aggregated into three layers. The first layer is formed by the input feature set which is linked to the output layer via an interconnection of several hidden layers in the middle. Each neuron processes its input variables and its output is passed to the neuron in the subsequent layer [81]. Logistic Regression (LoR) is a statistical technique for analyzing data when a dichotomous outcome is determined by one or more independent variables [82]. Support Vector Machines (SVMs) are based on the principle of recognizing patterns in a multidimensional hyperplane to estimate the maximum margin between samples of binary classes using a multidimensional input feature space [83].

These algorithms have relatively fast training and good performance and are robust to overfitting and have therefore been chosen in this study. The performance of classification models based on each of these algorithms is evaluated using the evaluation metrics listed in Table 2. These evaluation metrics are standard in machine learning literature [84].

The input features used for binary classification are the STFT coefficients of speech from each of the 84 healthy individuals and the 22 COVID-19 positive individuals. Each individual’s speech sample comprises 8 segments and STFT coefficients obtained from each speech segment are used as input features for binary classification. As mentioned in Section 4, STFT coefficients are complex numbers; hence, each speech segment comprises “real” and “imaginary” parts of STFT coefficients. The number of frequency points used in the computation of STFT coefficients is 8192, which is obtained as the next power of 2 greater than the segment length. Thus, for each individual speech sample, a matrix of 8192 rows \times 8 columns is generated. The real and imaginary parts of the complex STFT coefficients are separated resulting in two separate matrices having dimensions of 8192 rows \times 8 columns each. Class label is assigned to each row of these matrices as “Class 0” for healthy individuals’ speech samples and “Class 1” for COVID-19 positive individuals. Thus, there are 8192 \times 84 = 688,128 rows of STFT coefficients (real part) labeled as Class 0 and 8192 \times 22 = 180,224 rows of STFT coefficients (real part) labeled as Class 1. Correspondingly, an equal number of rows are available under each class label containing the “imaginary part” of STFT coefficients.

Each row of the real/imaginary part of the STFT coefficients matrix corresponds to a frequency point in the STFT

computation, and each column represents a segment of speech. The rows are treated as examples and columns as features since each column of the STFT matrix represents the time-localized spectral features of the speech signals. Every real and imaginary part “ x ” of STFT coefficients is normalized to lie in the interval [0,1] using a MinMax normalizer as follows:

$$\text{Normalized value} = \frac{x - \min(x)}{[\max(x) - \min(x)]}. \quad (2)$$

Since the statistical distribution for both the real and imaginary parts of speech STFT coefficients is Laplacian, these are treated together without distinction in the context of this work. Thus, the entire dataset comprises 1,736,704 labeled rows, half of which comprise the real part of STFT coefficients and the other half comprise the imaginary part of STFT coefficients, which is saved in .csv format. For binary classification using machine learning, only the rows corresponding to the real part of STFT coefficients are utilized to reduce the time taken for training and cross-validation. The data is split in an 80 : 20 ratio; i.e., 80% of the rows are used for training and the remaining 20% are used for testing. Since the dataset used in this study is highly imbalanced—Class 0 constitutes nearly 80% of the dataset and Class 1 constitutes a little over 20% of the dataset—data splitting is performed with “stratification.” Stratification ensures that each subset of split data has the same class distribution as the entire dataset. The ratio of healthy: COVID-19 + samples in terms of speech recordings is 84 : 22 = 3.8181 : 1. In terms of the STFT coefficients also, this ratio remains the same. Since only the real part of STFT coefficients has been used for binary classification, the ratio healthy: COVID-19 + samples in terms of STFT coefficients is 688128:180224 = 3.8181 : 1. Since stratification has been used, both training and testing data contain STFT coefficients of “healthy” and “COVID+” individuals in the same proportion as 84 : 22 = 3.8181 : 1. Furthermore, the train-test split with stratification at the STFT level ensures permutation of the labeled STFT coefficients across all “individuals”—Class 0 as well as Class 1. Even though the number of speech samples used in this study is small, the number of frequency points (rows) of STFT is large due to the high spectral resolution adopted. The performance evaluation metrics are computed following a 10-fold cross-validation process. Furthermore, as in the data splitting process, stratification is used in the cross-validation process as well to ensure that the class distribution of the training data set is maintained in each fold of cross-validation.

3. Results

3.1. Classification of COVID-19 Samples Based on Statistical Distribution Fitting. LD fitting based on MLE is applied to all the speech samples used in this study—healthy individuals without COVID-19 as well as COVID-19 positive individuals. A comparison of statistical properties of speech STFT coefficients of the two categories of speech samples is performed. The statistical distribution of speech STFT coefficients of healthy individuals, i.e., not

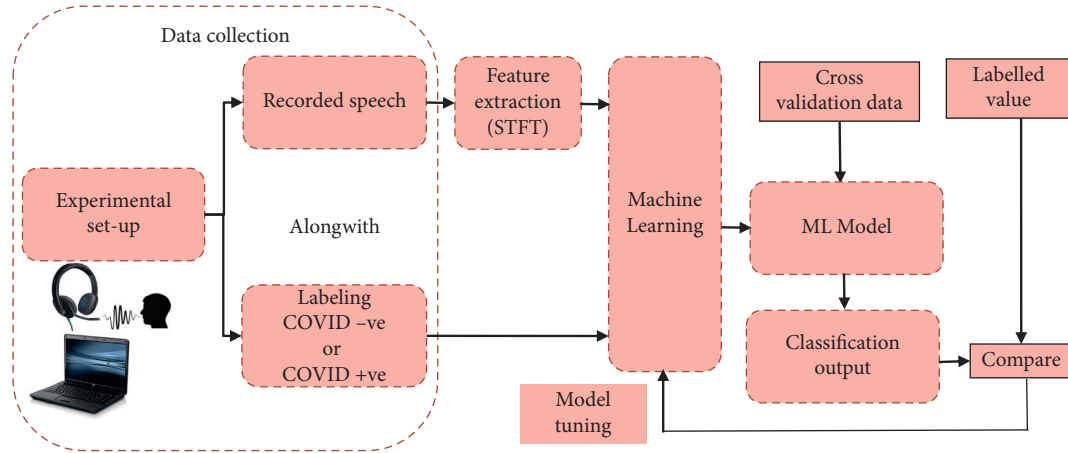


FIGURE 3: Block diagram of the methodology used.

TABLE 2: Evaluation metrics for binary classification.

Evaluation metric	Definition	Notations
Binary classification		
Precision (PRE)	$PRE = t_p / t_p + f_p$	tp-Total no. of true positive samples
Recall (REC)	$REC = t_p / t_p + f_n$	tn-Total no. of true negative samples
Accuracy (ACC)	$ACC = t_p + t_n / t_p + t_n + f_p + f_n$	fp-Total no. of false positive samples
F1-score	$F1 = 2 * P_{RE} * R_{EC} / P_{RE} + R_{EC}$	fn - Total no. of false negative samples
Area under RoC curve	(AUC) $AUC = \int_0^1 RoC$	RoC - receiver operating characteristic curve

infected by COVID-19, is shown in Figure 4 and that of a COVID-19 positive individual is shown in Figure 5.

It is found from Figure 4 that spectral features of the speech of healthy individuals are accurately modeled by LD, with small RMS error as has been established in the literature [58]. In the case of speech samples of COVID-19 positive individuals without any conspicuous symptoms, while the statistical distribution of the STFT coefficients is still closely modeled by LD, the RMS error of the fitted distribution has a nearly 10-fold increase as compared to non-COVID-19 individuals. This increase in the RMS error of the fitted distribution indicates a variation of speech characteristics as a result of COVID-19 infection. The PDFs are obtained by plotting the envelope of histograms of STFT coefficients. The “estimated” PDF represents the fitted distribution based on estimated Laplacian distribution parameters “ μ ” and “ b ” and the “actual” PDF is the actual distribution of the STFT coefficients. The Laplacian distribution is therefore a suitable distribution for speech STFT coefficients as the RMS error between the actual and fitted distributions is small.

3.2. Performance Evaluation of Binary Classification. The model hyperparameters for each algorithm are tuned and optimized to achieve the best performance in binary classification in terms of the “Recall” metric. The performance evaluation metrics for the five binary classification algorithms used in this study, along with their optimal parameterization, are listed in Table 3. The best performance in terms of precision, recall, accuracy, and F1 score is achieved for the DF algorithm. For the classification application

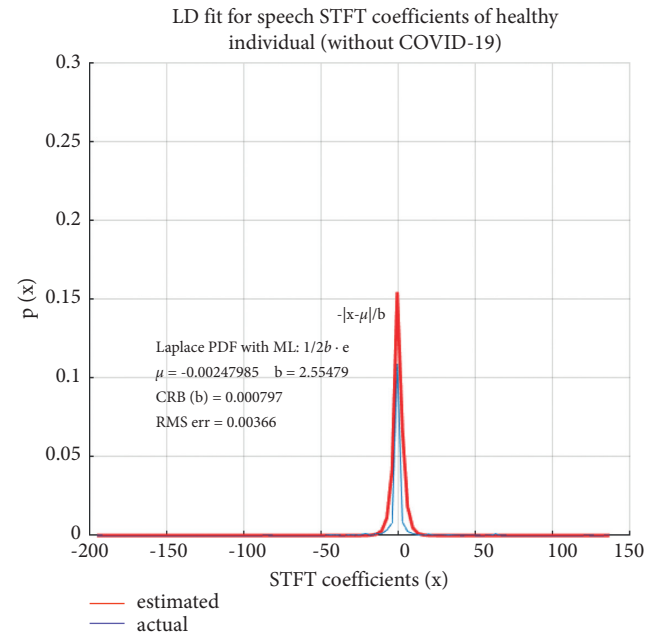


FIGURE 4: Statistical properties of speech STFT coefficients of a healthy person (without COVID-19).

considered in this work, classifying speech spectral features as COVID-19 positive or not, the cost associated with misclassification is very high for “false negative” as compared to “false positive.”

Since “recall” provides a measure of correctly predicted positives against the total number of positive

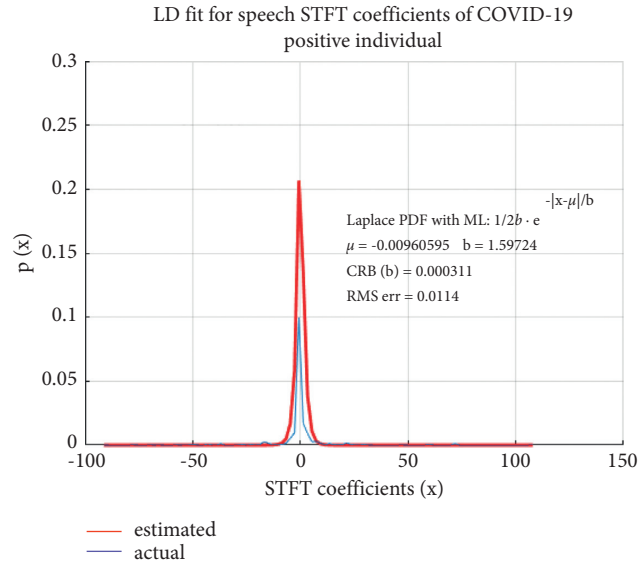


FIGURE 5: Statistical properties of speech STFT coefficients of an infected person (with COVID-19).

examples, it is important for our classification problem to have a high value for this metric. This will minimize misclassifying a COVID-19 positive example as not being COVID-19 positive. While other evaluation metrics have also been determined, “recall” is the more important metric in the context of this work. In Table 3, the values within brackets are the standard deviations of the metrics. Small values for standard deviation indicate that the models are verified with an unbiased dataset which has been achieved by using stratification in the train-test split as well as in cross-validation.

4. Discussion

4.1. Statistical Distribution of COVID-19 Positive and COVID-19 Negative Speech Samples. The average RMS error for the fitted LD averaged over all the speech samples belonging to each of the two categories is shown in Table 4. This increase in RMS error of the fitted LD in COVID-19 positive samples is attributed to the physiological changes associated with COVID-19 infection which affect the characteristics of speech.

Since the symptoms are not conspicuous among the samples used in this study, the distribution of STFT coefficients of speech is still Laplacian, albeit with a higher RMS error. STFT coefficients being complex numbers, the above findings are valid for both the “real” as well as “imaginary” parts of STFT coefficients and the same has also been shown in [58]. It remains to be seen if the distribution deviates significantly from being Laplacian or even ceases to be Laplacian when the symptoms become more pronounced and conspicuous. The increase in RMS error indicates such a trend. It should be noted that the data used in this study is unbalanced—the dataset from COVID-19 positive individuals is smaller than that of healthy individuals. The results discussed in this section clearly indicate that the statistical properties of speech spectral features are altered as a result of

COVID-19 infection. However, as it was not possible to obtain samples of COVID-19 positive individuals whose symptoms are more pronounced and conspicuous, this shall be a subject of future investigation, once such samples are obtained. Due to the prevailing COVID-19 restrictions, access to such individuals has not been possible.

4.2. COVID-19 Detection on Test Data Using the Binary Classification Models. Finally, the optimally parameterized classification algorithms discussed in Section 2.4 have been tested on the test dataset. As discussed in Section 3.2, the algorithms have been parameterized to optimize the “recall” performance metric. The classification results of 20 samples (rows) from the test data are shown in Table 5 which contains the “actual” and “predicted” classes for the 20 test samples by each of the five classification algorithms. It can be observed from Table 5 that the misclassification of COVID-19 positive as “not positive”, i.e., class 1 being misclassified as Class 0, is the lowest for the DF algorithm. The misclassified values are highlighted in bold and underlined.

For future investigation, concurrently at the time of recording speech samples of individuals, biomedical parameters such as heart rate (pulse oximeter), oxygen saturation (pulse oximeter), blood pressure (digital BP monitor), and temperature (infrared thermometer) have also been measured. These shall be used for future research to develop machine learning-based regression algorithms to predict these biomedical parameters from speech signals. The variations of these parameters among COVID-19 negative and COVID-19 positive individuals shall be analyzed to improve the accuracy of detecting COVID-19 from speech samples. The devices used to measure the biomedical parameters are shown in Figure 6. The e-health sensor platform shown in Figure 6 facilitates direct recording of the biomedical parameter to a PC, thus avoiding the manual entry of data.

A limitation of the work presented in this article is that it cannot distinguish between similar symptoms which may

TABLE 3: Performance metrics for binary classification algorithms.

Classification algorithms	Optimal Parameterization	Performance metrics Mean value (standard deviation)				
		A _{CC}	P _{RE}	R _{EC}	F1 score	AUC
BDT	No. of Leaves: 16 Learning rate: 0.05 No. of trees: 100	0.724 (0.048)	0.714 (0.037)	0.7037 (0.063)	0.7088 (0.052)	0.717 (0.053)
DF	Random split Count: 128 Maximum Depth: 32 No. of decision trees: 16	0.7317 (0.021)	0.7421 (0.017)	0.7892 (0.081)	0.7649 (0.025)	0.755 (0.017)
NN	Learning rate: 0.001 No. of hidden Nodes: 314	0.711 (0.031)	0.7271 (0.043)	0.7188 (0.018)	0.7229 (0.029)	0.7616 (0.095)
LoR	Optimization Tolerance: 1e-06 L1 regularization weight: 1 Memory size for L-BFGS: 18	0.6741 (0.019)	0.6805 (0.024)	0.6161 (0.027)	0.6467 (0.019)	0.6874 (0.065)
SVM	Lambda - 0.001	0.694 (0.017)	0.673 (0.074)	0.6027 (0.019)	0.6359 (0.011)	0.6619 (0.037)

TABLE 4: Average RMS error of the fitted LD distributions.

Category	Average RMS error of fitted LD
Without COVID-19	0.00354
With COVID-19	0.01271

TABLE 5: Test evaluation for binary classification

Test sample	Actual class	Predicted class				
		BDT	DF	NN	LoR	SVM
1	0	0	1	0	0	1
2	1	1	1	<u>0</u>	1	1
3	1	1	1	<u>1</u>	<u>0</u>	<u>0</u>
4	0	1	0	1	1	1
5	1	1	1	<u>0</u>	1	1
6	0	0	0	0	0	0
7	1	<u>0</u>	<u>0</u>	1	1	1
8	1	<u>1</u>	<u>1</u>	1	<u>0</u>	<u>0</u>
9	1	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u>	<u>1</u>
10	0	<u>1</u>	<u>0</u>	<u>0</u>	1	1
11	1	<u>0</u>	1	1	1	1
12	0	0	1	1	0	0
13	0	1	0	1	1	0
14	1	1	1	1	<u>0</u>	1
15	0	0	0	0	0	1
16	0	1	1	0	0	1
17	1	1	1	<u>0</u>	<u>0</u>	<u>0</u>
18	1	<u>0</u>	1	<u>1</u>	<u>1</u>	<u>1</u>
19	0	<u>0</u>	0	1	0	1
20	1	1	1	1	0	1

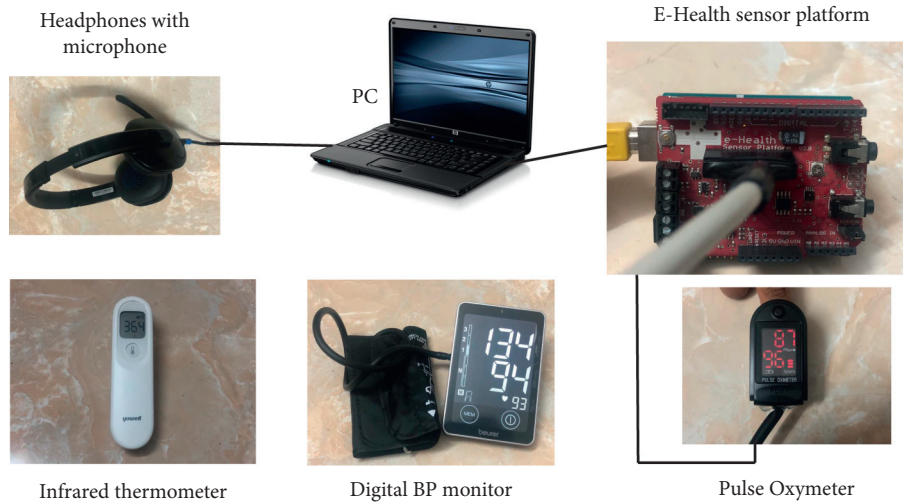


FIGURE 6: Devices used for measuring biomedical parameters along with speech for future analysis.

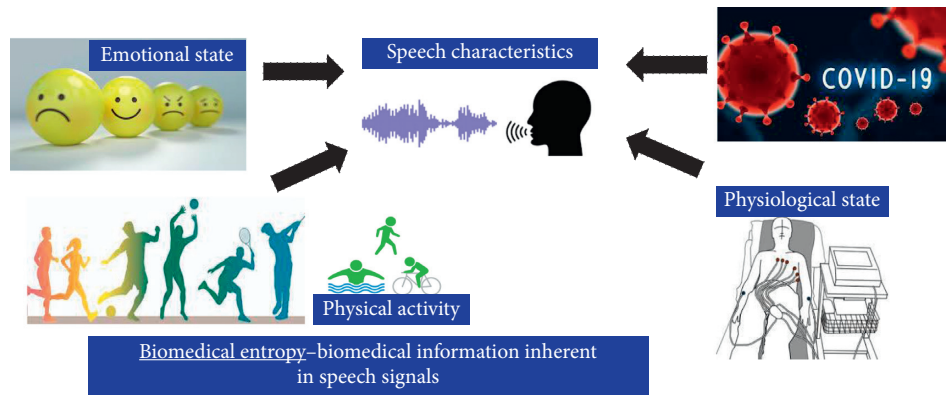


FIGURE 7: Possible applications of the proposed research.

appear due to multiple different causes such as influenza or myocarditis. It requires further research involving speech data from patients with various illnesses that have symptoms similar to COVID-19. This shall be a subject of future work.

5. Conclusions

This article investigates the statistical properties of speech spectral features for samples taken from healthy as well as asymptomatic COVID-19 positive individuals. While the statistical distribution for both is Laplacian, the RMS error of the fitted Laplace distribution is higher in the case of asymptomatic COVID-19 positive speech samples. This indicates that spectral properties of speech get altered as a result of physiological changes caused due to COVID-19 infection. It is therefore deduced that there is an associated entropy in speech which can be used to detect COVID-19. STFT coefficients of speech are then used as input features of machine learning-based classification algorithms and the classification performance of five state-of-the-art classification algorithms has been evaluated. All the five classification algorithms exhibit a moderate level of performance having their evaluation metrics values around 70% of their

maximum values. The best performance is observed for the DF algorithm which has the highest value for the “recall” metric with a value of 0.7892. “Recall” is the metric used while training the model hyperparameters as a higher recall value means minimizing misclassification of the “false negative” category. The cost of misclassifying a COVID-19 positive sample as a COVID-19 negative is high and hence the choice of recall as the evaluation metric is to be maximized while tuning the model parameters. It is also noted from Table 5 that the misclassification of Class 1 (COVID-19 positive) as Class 0 (COVID-19 negative) is least for the DF algorithm when tested on previously unseen test data. The results obtained are promising and provide evidence that COVID-19 infection can be detected from speech signals of individuals.

Speech can be used as a biomedical signal to diagnose various physical and emotional disorders. It can be used to monitor the performance/health conditions of individuals while performing physical activity. The focus of this work, however, is to detect COVID-19 infection by analyzing a person’s speech signal. This is possible because speech characteristics of individuals get altered by these conditions as depicted in Figure 7.

The results presented are concurring with similar approaches available in published literature. For example, 100% sensitivity is reported in [71] for asymptomatic cases, but it uses additional biomarker information along with cough sounds. A maximum “recall” value of 0.72 is reported in [73,78] while, in [76], the highest accuracy of 92.85% is reported for binary classification using deep transfer learning.

The results presented in this work can be improved by using a larger dataset comprising different classes of human vocal sounds which should also include samples of individuals of different languages, dialects, and other health conditions. It was intended to collect large datasets by encouraging community participation but that could not be achieved due to regulatory procedures and limitations of funding. Hence, the results presented in this work are based on a small dataset but the findings are encouraging. Future work shall consider using the magnitude of STFT coefficients rather than just the real/imaginary part and also consider the use of other types of audio signal features such as MFCC as input features for ML-based classification. The detection of COVID-19 using speech can facilitate real-time, remote monitoring of infected yet asymptomatic individuals. This will allow early detection of COVID-19 symptoms and help manage the ongoing COVID-19 situation better. It should be noted that the fundamental idea presented in this article is not limited to detecting COVID-19 symptoms only but has broader applications in medical diagnosis and patient monitoring/care. AI can detect changes in human vocal sounds not discernible to the human ear. Smartphone apps that use AI algorithms to analyze human vocal sounds for diagnosis, screening, and monitoring can be extremely useful and are expected to play a vital role in future healthcare technologies.

Data Availability

The speech recordings and the features data used to support the findings of this study have not been made available because the participants have not consented to their data being shared with any third party.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are thankful to the Institute of Research and Consulting Studies at King Khalid University for supporting this research through grant number # 33-91-S-2020.

References

- [1] Vital Signs (Body Temperature, Pulse Rate, Respiration Rate, Blood Pressure) [Internet], “Johns Hopkins Medicine,” 2021, <https://www.hopkinsmedicine.org/health/conditions-and-diseases/vital-signs-body-temperature-pulse-rate-respiration-rate-blood-pressure>.
- [2] A. Reynolds and A. Paivio, “Cognitive and emotional determinants of speech,” *Canadian Journal of Psychology/Revue canadienne de psychologie*, vol. 22, no. 3, pp. 164–175, 1968.
- [3] L. A. Ramig, “Effects of physiological aging on vowel spectral noise,” *Journal of Gerontology*, vol. 38, no. 2, pp. 223–225, 1983.
- [4] J. Trouvain and K. P. Truong, *Prosodic Characteristics of Read Speech before and after Treadmill Running*, International Speech Communication Association (ISCA), Baixas, France, 2015.
- [5] T. D. Borkovec, R. L. Wall, and N. M. Stone, “False Physiological Feedback and the maintenance of speech anxiety,” *Journal of Abnormal Psychology*, vol. 83, no. 2, pp. 164–168, 1974.
- [6] ScienceEncyclopedia, “Speech - the Physiology of Speech - Air, Vocal, Words, and Sound - JRank Articles [Internet],” 2019, <https://science.jrank.org/pages/6371/Speech-physiology-speech.html>.
- [7] D. Skopin and S. Baglikov, “Heartbeat feature extraction from vowel speech signal using 2D spectrum representation,” in *Proceedings of the 4th International Conference Information Technology*, p. 6, Doha, Qatar, June 2009.
- [8] B. Schuller, F. Friedmann, and F. Eyben, “Automatic recognition of physiological parameters in the human voice: heart rate and skin conductance,” in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7219–7223, IEEE, Vancouver, Canada, Vancouver 2013.
- [9] A. Mesleh, D. Skopin, S. Baglikov, and A. Quteishat, “Heart rate extraction from vowel speech signals,” *Journal of Computer Science and Technology*, vol. 27, no. 6, pp. 1243–1251, 2012.
- [10] J. Kaur and R. Kaur, “Extraction of heart rate parameters using speech analysis,” *International Journal of Science and Research*, vol. 3, no. 10, pp. 1374–1376, 2014.
- [11] M. Sakai, “Modeling the relationship between heart rate and features of vocal frequency,” *International Journal of Computer Application*, vol. 120, no. 6, pp. 32–37, 2015.
- [12] R. F. Orlikoff and R. J. Baken, “The effect of the heartbeat on vocal fundamental frequency perturbation,” *Journal of Speech, Language, and Hearing Research*, vol. 32, no. 3, pp. 576–582, 1989.
- [13] B. Schuller, F. Friedmann, and F. Eyben, “The munich bio-voice corpus: effects of physical exercising, heart rate, and skin conductance on human speech production,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pp. 1506–1510, Reykjavik: European Language Resources Association (ELRA), Reykjavik, Iceland, May 2014.
- [14] M. Usman, “On the performance degradation of speaker recognition system due to variation in speech characteristics caused by physiological changes,” *International Journal of Computing and Digital Systemss*, vol. 6, no. 3, pp. 119–127, 2017.
- [15] A. P. James, “Heart rate monitoring using human speech spectral features,” *Human-centric Computing and Information Sciences*, vol. 5, no. 1, pp. 1–12, 2015.
- [16] D. A. Burton, K. Stokes, and G. M. Hall, “Physiological effects of exercise,” *Continuing Education in Anaesthesia, Critical Care & Pain*, vol. 4, no. 6, pp. 185–188, 2004.
- [17] M. Sakai, “Feasibility study on blood pressure estimations from voice spectrum analysis,” *International Journal of Computer Application*, vol. 109, no. 7, pp. 39–43, 2015.

- [18] D. Griswold and M. Abraham, "Slurred Speech From Anxiety: Causes and Treatments [Internet]. CalmClinic," 2021, <https://www.calmclinic.com/anxiety/symptoms/slurred-speech>.
- [19] C. G. Scully, J. Jinseok Lee, J. Meyer et al., "Physiological parameter monitoring from optical recordings with a mobile phone," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 303–306, 2012.
- [20] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.
- [21] Y. Mardian, H. Kosasih, M. Karyana, A. Neal, and C.-Y. Lau, "Review of current COVID-19 diagnostics and opportunities for further development," *Frontiers of Medicine*, vol. 8, 2021.
- [22] N. M.-H. Tayarani, "Applications of artificial intelligence in battling against covid-19: a literature review," *Chaos, Solitons & Fractals*, vol. 142, Article ID 110338, 2021.
- [23] A. S. Albahri, R. A. Hamid, J. k. Alwan et al., "Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review," *Journal of Medical Systems*, vol. 44, no. 7, p. 122, 2020.
- [24] N. Alballa and I. Al-Turaiki, "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: a review," *Informatics in Medicine Unlocked*, vol. 24, Article ID 100564, 2021.
- [25] J. Bullock, A. Luccioni, K. H. Pham, C. Sin Nga Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *Journal of Artificial Intelligence Research*, vol. 69, pp. 807–845, 2020.
- [26] I. E. Agbehadji, B. O. Awuzie, A. B. Ngowi, and R. C. Millham, "Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing," *International Journal of Environmental Research and Public Health*, vol. 17, no. 15, p. 5330, 2020.
- [27] T. Aishwarya and V. Ravi Kumar, "Machine learning and deep learning approaches to analyze and detect COVID-19: a review," *SN Computer Science*, vol. 2, no. 3, p. 226, 2021.
- [28] H. Swapnarekha, H. S. Behera, J. Nayak, and B. Naik, "Role of intelligent computing in COVID-19 prognosis: a state-of-the-art review," *Chaos, Solitons & Fractals*, vol. 138, Article ID 109947, 2020.
- [29] L. Wynants, B. Van Calster, G. S. Collins et al., "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *BMJ*, vol. 369, Article ID m1328, 2020.
- [30] W. T. Li, J. Ma, N. Shende et al., "Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 247, 2020.
- [31] M. Allam, S. Cai, S. Ganesh et al., "COVID-19 diagnostics, tools, and prevention," *Diagnostics*, vol. 10, no. 6, p. 409, 2020.
- [32] The Lancet Digital Health, "Artificial intelligence for COVID-19: saviour or saboteur?" *Lancet Digit Heal*, vol. 3, p. e1, 2021.
- [33] Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *Npj Digital Medicine*, vol. 4, no. 1, p. 3, 2021.
- [34] M. Irfan, M. A. Iftikhar, S. Yasin et al., "Role of hybrid deep neural networks (HDNNs), computed tomography, and chest X-rays for the detection of COVID-19," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, p. 3056, 2021.
- [35] S. Masoud Rezaeijo, M. Ghorvei, and M. Alaei, "A machine learning method based on lesion segmentation for quantitative analysis of CT radiomics to detect COVID-19," in *Proceedings of the 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1–5, IEEE, Mashhad, Iran, December 2020.
- [36] S. M. Rezaeijo, R. Abedi-Firouzjah, M. Ghorvei, and S. Sarnameh, "Screening of COVID-19 based on the extracted radiomics features from chest CT images," *Journal of X-Ray Science and Technology*, vol. 29, no. 2, pp. 229–243, 2021.
- [37] A. Borkowski, "Using artificial intelligence for COVID-19 chest X-ray diagnosis," *Federal Practitioner*, vol. 37, no. 9, pp. 398–404, 2020.
- [38] L. Brunese, F. Martinelli, F. Mercaldo, and A. Santone, "Machine learning for coronavirus covid-19 detection from chest x-rays," *Procedia Computer Science*, vol. 176, pp. 2212–2221, 2020.
- [39] Y. Karbhari, A. Basu, Z. W. Geem, G.-T. Han, and R. Sarkar, "Generation of synthetic chest X-ray images and detection of COVID-19: a deep learning based approach," *Diagnostics*, vol. 11, no. 5, p. 895, 2021.
- [40] Y. Qiblawey, A. Tahir, M. E. H. Chowdhury et al., "Detection and severity classification of COVID-19 in CT images using deep learning," *Diagnostics*, vol. 11, no. 5, p. 893, 2021.
- [41] S. Chattopadhyay, A. Dey, P. K. Singh, Z. W. Geem, and R. Sarkar, "COVID-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer," *Diagnostics*, vol. 11, no. 2, p. 315, 2021.
- [42] M. Kukar, G. Gunčar, T. Vovko et al., "COVID-19 Diagnosis by Routine Blood Tests Using Machine Learning," *Scientific Reports*, vol. 11, Article ID 10738, 2020.
- [43] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, "Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study," *Journal of Medical Systems*, vol. 44, no. 8, p. 135, 2020.
- [44] M. AlJame, I. Ahmad, A. Imtiaz, and A. Mohammed, "Ensemble learning model for diagnosing COVID-19 from routine blood tests," *Informatics in Medicine Unlocked*, vol. 21, Article ID 100449, 2020.
- [45] F. E. Shamout, Y. Shen, N. Wu et al., "An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department," *Npj Digital Medicine*, vol. 4, no. 1, p. 80, 2021.
- [46] Z. Dai, D. Zeng, D. Cui et al., "Prediction of COVID-19 patients at high risk of progression to severe disease," *Frontiers in Public Health*, vol. 8, Article ID 574915, 2020.
- [47] M. Carlile, B. Hurt, A. Hsiao, M. Hogarth, C. A. Longhurst, and C. Dameff, "Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department," *Journal of the American College of Emergency Physicians Open*, vol. 1, no. 6, pp. 1459–1464, 2020.
- [48] C. Hu, Z. Liu, Y. Jiang et al., "Early prediction of mortality risk among patients with severe COVID-19, using machine learning," *International Journal of Epidemiology*, vol. 49, no. 6, pp. 1918–1929, 2021.
- [49] A. López-Escobar, R. Madurga, J. M. Castellano et al., "Risk score for predicting in-hospital mortality in COVID-19 (RIM score)," *Diagnostics*, vol. 11, no. 4, p. 596, 2021.
- [50] Y. Gao, G.-Y. Cai, W. Fang et al., "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19," *Nature Communications*, vol. 11, no. 1, p. 5033, 2020.

- [51] F. Tretter, O. Wolkenhauer, M. Meyer-Hermann et al., "The quest for system-theoretical medicine in the COVID-19 era," *Frontiers of Medicine*, vol. 8, 2021.
- [52] W. Naudé, "Artificial intelligence vs COVID-19: limitations, constraints and pitfalls," *AI & Society*, vol. 35, no. 3, pp. 761–765, 2020.
- [53] M. W. Tobias, "AI and Medical Diagnostics: Can A Smartphone App Detect Covid-19 from Speech or A Cough? Forbes," 2020.
- [54] J. Chu, *Artificial Intelligence Model Detects Asymptomatic Covid-19 Infections through Cellphone-Recorded Coughs*, MIT News, Cambridge, MA, USA, 2020.
- [55] M. Scudellari, *AI Recognizes COVID-19 in the Sound of a Cough*, IEEE Spectrum - The Institute, USA, 2020.
- [56] H. Coppock, A. Gaskell, P. Tzirakis, A. Baird, L. Jones, and B. Schuller, "End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study," *BMJ Innovations*, vol. 7, no. 2, pp. 356–362, 2021.
- [57] M. MacGill, "Heart rate: What is a normal heart rate? [Internet]," 2017, <https://www.medicalnewstoday.com/articles/235710.php>.
- [58] M. Usman, M. Zubair, M. Shiblee, P. Rodrigues, and S. Jaffar, "Probabilistic modeling of speech in spectral domain using maximum likelihood estimation," *Symmetry*, vol. 10, no. 12, p. 750, 2018.
- [59] P. Partila, M. Voznak, M. Mikulec, and J. Zdralek, "Fundamental frequency extraction method using central clipping and its importance for the classification of emotional state," *Advances in Electrical and Electronic Engineering*, vol. 10, no. 4, pp. 270–275, 2012.
- [60] Z.-H. Tan and B. Lindberg, "Low-complexity variable frame rate analysis for speech recognition and voice activity detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 5, pp. 798–807, 2010.
- [61] J. J. Wolf, "Speech signal processing and feature extraction," in *Spoken Language Generation and Understanding* Dordrecht: Springer, Netherlands, Europe, 1980.
- [62] K. J. Reilly and C. A. Moore, "Respiratory sinus arrhythmia during speech production," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 1, pp. 164–177, 2003.
- [63] C. Von Euler, "Speech motor control," in *Proceedings of the International Symposium on Speech Motor Control, Wenner-Gren Center International Symposium Series*, pp. 95–103, Stockholm, Sweden, May 1981, <https://www.sciencedirect.com/science/article/pii/B978008028892550013X>.
- [64] F. Yasuma and J.-i. Hayano, "Respiratory sinus arrhythmia," *Chest*, vol. 125, no. 2, pp. 683–690, 2004.
- [65] S. B. Magre, R. R. Deshmukh, and P. P. Shrishrimal, "A comparative study on feature extraction techniques in speech recognition," in *Proceedings of the International Conference on Recent Advances in Statistics and Their Applications*, Aurangabad, Maharashtra, India, December 2013.
- [66] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, Prentice Hall PTR, Hoboken, NJ, USA, 2001.
- [67] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.
- [68] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, 1990.
- [69] Z. Tufekci and J. N. Gowdy, "Feature extraction using discrete wavelet transform for speech recognition," in *Proceedings of the IEEE SoutheastCon 2000 "Preparing for the New Millennium" (Cat No00CH37105)*, pp. 116–123, IEEE, Nashville, TN, USA, April 2000.
- [70] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 578–589, 1994.
- [71] J. Laguarda, F. Hueto, and B. Subirana, "COVID-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 2020.
- [72] G. Deshpande and B. W. Schuller, "Audio, Speech, Language, & Signal Processing for COVID-19: A Comprehensive Overview," *Pattern Recognit*, vol. 122, Article ID 108289, 2020.
- [73] C. Brown, J. Chauhan, A. Grammenos et al., "Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3474–3484, ACM, New York, NY, USA, August 2020.
- [74] I. Trancoso, "Project to Detect COVID-19 from Coughs and Speech," 2021, <https://www.inesc-id.pt/project-to-detect-covid-19-from-coughs-and-speech/>.
- [75] E. Anthes, "Alexa, do I have COVID-19?" *Nature*, vol. 586, no. 7827, pp. 22–25, 2020.
- [76] A. Imran, I. Posokhova, H. N. Qureshi et al., "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, Article ID 100378, 2020.
- [77] 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," *Digit. Health*, vol. 3, 2021.
- [78] J. Han, C. Brown, J. Chauhan et al., "Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data," in *Proceedings of the IEEE ICASSP 2021*, Toronto, Canada, June 2021.
- [79] P. Bühlmann and B. Yu, "Boosting with the L2Loss," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 324–339, 2003.
- [80] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Foundations and Trends® in Computer Graphics and Vision*, vol. 7, no. 2-3, pp. 81–227, 2011.
- [81] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.
- [82] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5–6, pp. 352–359, 2002.
- [83] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of Electronic Imaging*, vol. 16, no. 4, Article ID 049901, 2007.
- [84] S. Roychowdhury and M. Bihis, "AG-MIC: azure-based generalized flow for medical image classification," *IEEE Access*, vol. 4, pp. 5243–5257, 2016.