2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-0358-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BHI50953.2021.9508482

Rapid and Scalable COVID-19 Screening using Speech, Breath, and Cough Recordings

Drew Grant, Ian McLane, and James West*
Electrical and Computer Engineering
Johns Hopkins University
Baltimore, Maryland 21218
Email: jimwest@jhu.edu

***Abstract*—Over the course of the COVID-19 pandemic, efforts have been made to rapidly scale diagnostic tests to increase access and throughput. Though the primary mechanism for testing has been wet tests, several recent studies have shown that acoustic signatures of COVID-19 can be used to accurately discriminate between positive and negative subjects. These methods show promise of wide scale access and more regular and rapid testing, but are faced with several questions involving the robustness of the methods and the sanitary nature of forced cough recordings. Here we propose an alternative method to triage patients using acoustic signatures in speech and breathing sounds. Using a crowd-sourced database with sound recordings from self-identified COVID-19 positive and negative subjects, we develop a simple method that can be applied to analyze sounds that can be deployed in a system to unobtrusively detect COVID-19. Mel-frequency cepstral coefficients (MFCCs) and relAtive specTrA perceptual linear prediction (RASTA-PLP) features are evaluated independently and conjointly with two different classification techniques, random forests (RF) and deep neural networks (DNN). The optimal results are achieved for speech and breathing sounds using a combination of MFCC and RASTA-PLP, with an area-under-the-curve (AUC) of 0.7938 for detecting COVID-19 via speech sound analysis, and 0.7575 for detecting COVID-19 via breathing sound analysis. This is compared to an AUC of 0.6836 for cough sounds using MFCCs alone. These results show promise in future deployment of a rapid screening tool using speech recordings as the world moves to contain future outbreaks and accelerate vaccination efforts.**

I. INTRODUCTION

The emergence of the coronavirus SARS-CoV-2 and its associated disease (COVID-19) has led to unprecedented global disruptions. As of May 2021, there have been over 150 million confirmed cases of coronavirus globally and over 3 million confirmed deaths [1]. It is well recognized that in order to limit outbreaks, testing is needed to identify as many individuals that are infected as quickly as possible so they and their contacts can be isolated. Many of the key factors behind the rapid spread of COVID-19 across countries and continents stem from the speed, scarcity, supply chain, and costs of clinical tests such as antigen and polymerase chain reaction (PCR) tests [2]. Even with mass vaccinations being administered at record rates in developed countries, developing countries continue to be impacted by several compounding issues: the spread of COVID-19, the challenges associated with testing, the challenges associated with mass vaccinations, and medical supply scarcity. It becomes imperative that issues of testing improve and become more accessible and responsive.

Previous literature has shown promising results identifying COVID-19 positive patients via cough sound analysis [3] [4] [5]. However, they have failed to address issues related to mass deployment of the system. Cough is a well known symptom of COVID-19, but coughs are also a symptom of over 100 non-COVID-19 related medical conditions [6]. Questions remain regarding the differentiation between coughing for patients with chronic diseases, other types of infections, or asymptomatic patients. Furthermore, systems for COVID detection that require forced coughing from users provide significant sanitary concerns as a significant vector for transmitting respiratory diseases that render them unfit for deployment in public settings [7].

Speech and other paralinguistic sounds such as exhalation have been shown to be affected by the same mechanisms of acoustic production of cough in other diseases [8], but they provide significantly fewer sanitary concerns and are much more natural sounds to produce and to monitor in public settings, actively and passively. There is a gap in analysis of COVID-19 using breathing and speech sounds however. Here, we propose a versatile approach that shows promise in detecting COVID-19 through speech and breathing sound analysis as compared to cough analysis. The methods are trained and validated on a publicly available, crowd sourced data set comprising all three sounds from the same set of patients. The use of a publicly available dataset allows for direct comparisons by future work, a major limitation of existing cough sound analysis work. The proposed method offers an unobtrusive COVID testing alternative that provides diagnosis within minutes and can be deployed in smart phones, making this solution highly scalable for rapid screening.

This paper is structured as follows: Section II includes an overview of the data set and data preprocessing used in our analysis. In Section III, we discuss the feature extraction and classification techniques. Results are presented in Section IV and compared against other methods from literature. Finally, in Section V, we discuss considerations that need to be made with this analysis and future work needed to further validate.

978-1-6654-0358-0/21/$31.00 ©2021 IEEE

II. DATA

The Coswara dataset is a crowd-sourced dataset of sound recordings from COVID-19 positive and non-COVID-19 individuals [9]. Launched in April 2020, individuals can contribute to this web-based data collection by simply volunteering audio recordings electronically using their smartphones or computers. The database houses an audio collection of fast and slow breathing, deep and shallow coughing, phonation of sustained vowels, and spoken digits at two speeds, normal and fast. Information such as age, gender, geographic location, current health status, and preexisting medical conditions are also queried to supplement the recordings.

The authors used a subset of the Coswara dataset with two groups. The first group was composed of cough sound recordings, while the second group was composed of deep breathing and number counting speech recordings. The first group was composed of 1040 total subjects, of which only 75 were COVID positive subjects. The second group was composed of 1199 total subjects, of which only 80 were COVID positive subjects. A more detailed breakdown of the data with respect to the groups, sound event, COVID status, gender, and nationality is presented in Table 1. All the audio files were sampled at 44.1kHz. Both groups were segmented into 5-fold stratified cross validation splits for classification tasks.

A. Preprocessing

Various operations were applied to prepare signals for analysis and subsequent classification. Normalization is applied to constrain all amplitudes to the range of [-1,1]. Due to the quasi-stationary nature of speech, breathing, and cough recordings, short time window analysis is required [10]. Short time window analysis consists of analyzing uniformly spaced time frames of short durations (usually 20 to 40ms). Short time windows 25ms in duration were used in this study. Speech activity detection was applied by thresholding the short term energy E_{STE} given by [11]:

$$E_{STE} = \sum_{t=-\infty}^{\infty} |x(t)w(t-\tau)|^2 dt, \tag{1}$$

where $x(t)$ is the voice recording signal and $w(t-\tau)$ is a limited time window sequence of window length 25ms. An empirically derived thresholding value was used to remove frames with low energy, which were deemed to be silence, and preserve frames with sufficient energy. Data augmentation was employed by considering each frame of an audio signal as a separate observation for training and testing. The number of 25ms segments extracted from each subject was contingent upon the length of each subjects' recordings.

III. METHODS

The aim of feature extraction is to capture deeper characteristics and information from recordings. Hand-crafted acoustic features were extracted from the preprocessed recording segments for classification. We have considered Mel-Frequency Cepstral Coefficients (MFCCs), MFCC deltas, and RelAtive SpecTrA Perceptual Linear Prediction (RASTA-PLP) because of their widescale adoption in speech processing, and more specifically because of the success in the use of MFCCs in prior cough sound analysis work for COVID-19 detection. For this study the authors employed a Python adaptation of the RASTA-MAT toolbox called "Rasta_py" with default parameters for MFCC and RASTA-PLP extraction.

A. Mel-Frequency Cepstral Coefficients (MFCCs)

Mel-Frequency Cepstral Coefficient (MFCC) is a time-frequency representation commonly used in speech recognition that effectively emulates human auditory perception by logarithmically warping sound in Mel filter banks [12]. The algorithm consists of mapping the short-term power spectra to the Mel scale with a filter bank of triangular filters that mimic human's nonlinear perception to audio frequencies [12].

$$f_{mel}(f) = 2595 * (1 + \frac{f}{700}), \tag{2}$$

Next, the resulting logarithmic spectra signal m_j is transformed to the cepstral domain by applying the discrete cosine transform for compression and decorrelation [13]:

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j \cdot cos\left(\frac{\pi i}{N}(j-0.5)\right), \tag{3}$$

where N is the number of filterbank channels. Finally, the desired number of MFCC coefficients are extracted [14]. The first coefficient represents the average power in the spectrum and the lower-order coefficients describe the overall spectral shape of the signal [15]. The higher-order coefficients represent finer spectral details such as pitch and tonal information [15]. In practice, the first 13 MFCC coefficients are commonly used in speech processing applications. However, some applications require more higher-order coefficients. For example, Wang et al. [16] observed that the use of higher-order cepstral coefficients improved their system's recognition of Chinese speech performance by 30%. Given the effects respiratory diseases have on pitch and tone, the authors extracted 20 MFCC coefficients.

Time feature derivatives have been proven to provide acoustic information regarding the temporal trajectory and greatly enhance the performance speech recognition systems [13]. Delta (differential) and delta-delta (acceleration) coefficients provide insights into the dynamics of the of MFCCs over time. The delta coeffients are computed by:

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2}, \tag{4}$$

where d_t is a delta coefficient from frame t computed in terms of the static coefficients c_{t-n} to c_{t+n}, and $n = 2$.

B. RelAtive SpecTrA Perceptual Linear Prediction (RASTA-PLP)

RelAtive SpecTrA Perceptual Linear Prediction (RASTA-PLP) is used to eliminate channel distortions and improve

TABLE I. Demographic breakdown of the subjects present in the dataset used for analysis in this study.

Dataset Group	Sound Event	COVID Status (n)		Gender (n)		Nationality (n)	
		Positive	Negative	Male	Female	India	Other
Group 1	Cough	75	965	791	249	923	11
Group 2	Speech & Breathing	81	1118	911	288	1074	125

automatic speech recognition systems' speaker independence [10]. Perceptual Linear Prediction (PLP) combines three components from the psychophysics of human hearing to improve the estimation of the auditory spectrum [17].

1) Critical Band Spectral Resolution: Critical Band Spectral Resolution is computed by warping the short-term power spectrum into Bark frequency by using the following equation:

$$\Omega(f) = 6sinh^{-1}\left(\frac{f}{600}\right), \tag{5}$$

where f is frequency in Hz. Next the resulting warped spectrum is convolved with the power spectrum of the simulated critical band masking curve $\Psi(\Omega)$ approximated by:

$$\Psi(\Omega) = \begin{cases} 0 & for\,\Omega < -1.3 \\ 10^{2.5(\Omega-0.5)} & for -1.3 \leq \Omega \geq -0.5 \\ 1 & for -0.5 < \Omega < 0.5 \\ 10^{-(\Omega-0.5)} & for\,0.5 \leq \Omega \geq 2.5 \\ 0 & for\,\Omega > 2.5 \end{cases}, \tag{6}$$

2) Equal-loudness Preemphasis: Equal-loudness preemphasis consists of preamphasizing the samples using the simulated equal-loudness curve $E(f)$, which approximates the unequal sensitivity of human hearing at different frequencies as: [18]

$$E(f) = \left[\frac{f^2}{f^2 + 1.6 \times 10^5}\right] \times \left[\frac{f^2 + 1.44 \times 10^6}{f^2 + 9.6 \times 10^6}\right], \tag{7}$$

3) Intensity-loudness Power Law: Intensity-loudness power law approximates the power law of human hearing and the nonlinear relationship between the intensity of sound and it's perceived loudness by scaling the amplitude of the resulting spectra to the cubic-root.

Finally, an all-pole spectral model is applied using the autocorrelation method referred to as the Linear Prediction technique [19].

The RelAtive SpecTrAl (RASTA) technique employs a band-pass filter to the energy in each frequency subband to remove short-term noise variations and make PLP more robust to linear spectral distortions. While the PLP algorithm suppresses speaker-dependent information, the RASTA filter emulates the auditory critical band resolution, frequency resolution, and equal loudness perception to the short-term power spectra [20]. In our work we extracted 19th order RASTA-PLP features.

C. Classification

Two popular classification techniques were used to discriminate COVID subject voice recordings from non-COVID subjects.

1) Random Forest (RF): A random forest (RF) classifier is an ensemble of tree-structured classifiers $h_1(x)$ $h_2(x)$, ... , $h_K(x)$ that cast a unit vote for the most popular class to generate classification predictions. RFs are widely used due to their robustness to overfitting and high prediction accuracy [21]. RF have been employed in various medical applications and have demonstrated to have excellent performance in comparison to other machine learning algorithms [22] [23] [24] [25]. For this study the authors employed the Scikit Learn (version 0.24.2) Ensemble module's RF classifier with 100 trees and an information entropy criterion.

2) Deep Neural Networks (DNN): A Deep Neural Network (DNN) is a deep learning architecture that is widely used for a range of classification problems due to its ability to effectively model complex and nonlinear relationships. DNNs are feedforward networks comprised of neurons for the input features in the input layer, hidden layers, and an output layer with 2 neurons for binary classification. Each hidden layer m typically uses an activation function y_m determine whether a neuron should be activated in a layer x_m. Here we use ReLu activation functions because of their computational efficiency and their superior performance on smaller datasets [26] [27].

$$y_m = \text{ReLu}(x_m) = max(0, x_m), \tag{8}$$

$$x_m = b_m + \sum_n y_n w_{nm}, \tag{9}$$

where b_m is the bias of unit m, n is an index over units in the layer below, and w_{nm} is the weight on a connection to unit m from unit n in the layer below. For this study the authors employed a TensorFlow Core v2.5.0 model with 8 hidden layers, Adam optimization, a 0.0001 learning rate, and categorical cross-entropy loss.

IV. Results

In order to measure the success of our classifiers in discriminating COVID subjects from non-COVID subjects, we quantified the model performance of our classifiers by computing the receiver operating characteristic (ROC) curve and the area under the curve (AUC). While the ROC is useful in determining the decision threshold that minimizes the difference between the true positive rate (TPR) and false positive rate (FPR), AUC measures the entire two-dimensional area underneath the entire ROC curve and is classification-threshold-invariant. AUC measures the quality of the model's

TABLE II. ROC-AUC obtained when testing different feature sets and classifying schemes on different sound events.

Sound Event	Features	Total # Features	RF	DNN
Cough	MFCC	20	.6521	.6287
	MFCC+Deltas	60	.6687	**.6836**
	RASTA-PLP	20	.6614	.6569
	All	80	.6732	.6732
Speech	MFCC	20	.7506	.7123
	MFCC+Deltas	60	.7740	.7542
	RASTA-PLP	20	.6925	.6991
	All	80	**.7938**	.7564
Breath	MFCC	20	.7370	.6704
	MFCC+Deltas	60	.7384	.7306
	RASTA-PLP	20	.7368	.7393
	All	80	.7492	**.7575**

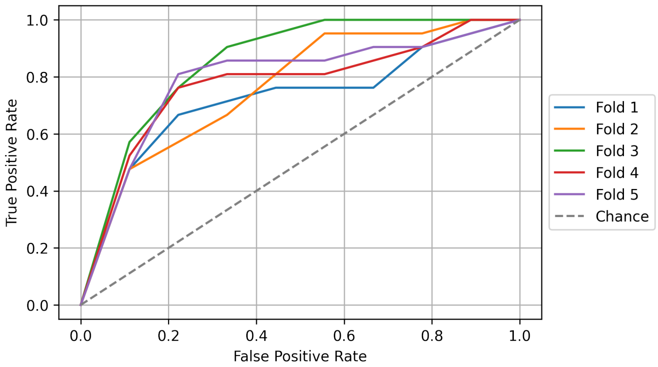

Fig. 1. Receiver Operating Curves of the best performing model for speech, computed over the five validation folds for the dataset.

predictions irrespective of what classification threshold is chosen and is therefore a good metric for understanding how well a classifier has performs more generally [28] [29].

Table II presents the final AUC results for each sound event, combination of features, and classifier scheme. Here the average AUC across the 5 folds is presented as the overall performance of each classification method. The best performing feature set for speech and breath sounds included MFCC, MFCC deltas, and RASTA-PLP features. The improvement of the combined features over the singular features is consistent regardless of classifier type. Overall, speech performed the best of the three acoustic groups, over both breathing and cough. A more detailed view of the highest performing model is presented in Fig. 1, which illustrates the ROC curves for each of the five folds as compared to absolute chance. It is clear from this figure that the classifier operates similarly across the subset of data.

V. DISCUSSION

In order to better contextualize these results within the broader context of existing literature, we compare the best performing results from speech to other methods. Since there are significantly fewer models presented for speech than for coughing and many studies show results from private datasets, several methods were re-implemented on speech recordings from this dataset to provide a more direct comparison. Table III provides a summary of these results and the methods are very briefly summarized below.

Imran et al. [4] used a combination of three deep learning and classical machine learning classifiers and combines their results using a mediator. The first and second classifiers are two convolutional neural network run on the mel-spectrogram of the signal. For the third classifier, instead of using a spectrogram like the first two classifiers, it uses MFCC and PCA based feature extraction to train a multi-class support vector machine (SVM). The classifiers' outcomes are consolidated by an automated mediator in which all three classifiers must agree in order to achieve a positive result. This model performed with an average AUC of 0.6502 across the 5-folds.

Bagad et al. [29] use the popular ResNet-18 architecture on an input spectrogram followed by adaptive pooling layer in both the time and frequency dimensions. Finally, the output is passed through 2 linear layers and then a final predictive layer with 2 neurons and a softmax activation function, which is used to predict whether the input cough sample has COVID-19. This method performed with an AUC of 0.7215, similar to the performance in their work on cough sound analysis but underperforming when compared to the proposed methodology.

Brown et al. [30] use a combination of handcrafted and deep learning audio features, and a logistic regression classifier. Handcrafted features include duration, onset, tempo, period, RMS energy, spectral centroid, roll-off frequency, zero-crossing rate, MFCC, Δ-MFCC, Δ^2-MFCC. For the features that generate time series, statistical features (mean, median, root-mean square, maximum, minimum, 1st and 3rd quartile, interquartile range,standard deviation, skewness, and kurtosis) are extracted to generate a 477-element feature vector for each recording. VGGish is also employed to extract features: VGGish is a convolutional neural network commonly used as a feature extractor for audio data. The mean and standard deviation across the VGGish features generate a feature vector with dimension 256 (128 × 2). The final feature vector (733 elements) is used to train a logistic regression classifier, which performed with an AUC of 0.7506.

Pinkas et al. [31] proposed a method for detecting COVID from speech using an attention-based transformed and recurrent neural networks. The details of this model were not able to be recreated, but the authors report an AUC of 0.81. Their performance is included here for reference, but direct comparison of these results to the proposed results cannot be made however because of the differences in the datasets and number of patients (88 patients in their work versus 1199 in this work).

There are several considerations that need to be made before being able to deploy this system at scale. The data used in this analysis was collected via crowd-sourcing, which has the possibility to introduce several issues. First, crowd-sourced

TABLE III. ROC-AUC comparing the proposed method to other methods in literature.

Reference	Features	Classifier	AUC
Proposed	MFCC+Deltas, RASTA-PLP	Random Forest	0.7938
Brown et al.	Duration, onset, tempo, period, RMS energy, spectral centroid, roll-off frequency, zero-crossing rate, MFCC+Deltas, VGGish	Logistic Regression	0.7506
Bagad et al.	Linear Spectrogram	CNN	0.7215
Imran et al.	Mel Spectrograms, MFCC, MFCC-PCA	CNN, SVM	0.6502
Pinkas et al.	Mel Spectrograms	GRU	*0.81*

data means the quality of the audio files has high variability. The presence of noise was not considered in this analysis and would need to be taken into account to translate to audio data collected in the wild. Second, the labels here are relying of self-assessment of the users who are volunteering their information. We cannot be entirely sure about the status of the patient, the mechanism for their positive test, or their state at the time of recording. Comparative studies would need to be conducted using PCR confirmation of diagnoses of patients.

The second set of considerations that need to be made relate to the data and models that currently exist in relation to cough sounds. It is clear that the results for each of the studies is highly depended on the kind of data that it is trained on. Without access to the data used for training in comparative systems, it is challenging to do a direct comparison between the models. A further complication is the significant data imbalance in this dataset. However, the data imbalance reflects the average positivity rate in the US (roughly 7%) and the metric employed in this study, AUC, is not biased to the majority or minority class. Other work tries to correct for data imbalance using data augmentation through synthetic minority oversampling technique (SMOTE) or variational auto-encoders (VAEs), but these introduce concerns about overfitting, leading to misleadingly high accuracies. Finally, replicating machine learning models is especially challenging because the use of initial states and hyperparameters of these models can dramatically change their end performance.

These results are intended to be a benchmark for the use of speech sounds in COVID-19 detection and illustrate the feasibility of this work. Further analysis will be conducted to understand any dependence on demographic factors in the screening decisions as well as the ability of the system to discriminate between COVID-19 and other respiratory infections. Finally, a prospective analysis would need to be done to ensure safety and efficacy of use.

VI. CONCLUSION

In this work, we propose the use of speech sounds for use in rapid and scalable COVID-19 screening. Previous work has shown the promise of cough sound analysis, but do not address the sanitary and applicability concerns associated with forced cough recordings. Using a publicly available dataset, we validate the use of MFCCs, MFCC deltas, and RASTA-PLP to classify recordings as COVID positive or negative, which results in a performance of 0.7938. A recent study by Cochrane [32], an international healthcare not-for-profit, showed that antigen tests correctly identified COVID-19 infection in an average of 72% of people with symptoms, compared to 58% of people without symptoms. The use of speech sounds to identify COVID-19 cases shows the possibility of comparable performance to antigen tests at a fraction of the cost and deployable to the nearly 3.8 billion smartphones globally. The deployment of this system would be highly beneficial in rapid and repetitive screening of patients. The use of acoustic testing may be most useful to identify outbreaks, or to select people with symptoms for further testing with PCR, allowing self-isolation or contact tracing and reducing the burden on laboratory services.

ACKNOWLEDGMENTS

First and foremost, the authors would like to express our deepest condolences for the lives lost due to COVID-19 and express our gratitude and respect for the clinicians and scientists on the front lines. We would also like to thank Drs. Mounya Elhilali and Hynek Hermansky for their insights during development. Finally, the authors would like to acknowledge and thank the team behind the Coswara dataset for making their data available to the public to accelerate development for the good of humanity.

REFERENCES

[1] WHO Coronavirus (COVID-19) Dashboard. Available: https://covid19.who.int.

[2] Y. C. Manabe, J. S. Sharfstein and K. Armstrong, ”The Need for More and Better Testing for COVID-19,” Jama, vol. 324, (21), pp. 2153-2154, 2020.

[3] J. Laguarta, 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. . DOI: 10.1109/OJEMB.2020.3026928.

[4] A. Imran et al, ”AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app,” Informatics in Medicine Unlocked, vol. 20, pp. 100378, 2020. Available: https://www.sciencedirect.com/science/article/pii/S2352914820303026. DOI: https://doi.org/10.1016/j.imu.2020.100378.

[5] M. Pahar et al, ”COVID-19 Cough Classification using Machine Learning and Global Smartphone Recordings,” arXiv Preprint arXiv:2012.01926, 2020.

[6] T. Higenbottam, ”Chronic Cough and the Cough Reflex in Common Lung Diseases,” Pulm. Pharmacol. Ther., vol. 15, (3), pp. 241-247, 2002. Available: https://www.sciencedirect.com/science/article/pii/S109455390290341X. DOI: https://doi.org/10.1006/pupt.2002.0341.

[7] J. Wei and Y. Li, ”Enhanced spread of expiratory droplets by turbulence in a cough jet,” Build. Environ., vol. 93, pp. 86-96, 2015. Available: https://www.sciencedirect.com/science/article/pii/S0360132315300329. DOI: https://doi.org/10.1016/j.buildenv.2015.06.018.

[8] J. E. Huber and M. Darling-White, ”Longitudinal Changes in Speech Breathing in Older Adults with and without Parkinson's Disease,” Semin Speech Lang, vol. 38, (3), pp. 200-209, 2017. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7193992/. DOI: 10.1055/s-0037-1602839.

[9] N. Sharma et al, ”Coswara–A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis,” arXiv Preprint arXiv:2005.10548, 2020.

[10] K. K. Paliwal, J. G. Lyons and K. K. Wójcicki, ”Preference for 20-40 ms window duration in speech analysis,” in - 2010 4th International Conference on Signal Processing and Communication Systems, 2010, . DOI: 10.1109/ICSPCS.2010.5709770.

[11] M. Jalil, F. A. Butt and A. Malik, ”Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals,” in May 2013, . DOI: 10.1109/TAEECE.2013.6557272.

[12] H. Hermansky, J. R. Cohen and R. M. Stern, ”Perceptual Properties of Current Speech Recognition Technology,” Proceedings of the IEEE, vol. 101, (9), pp. 1968-1985, 2013. . DOI: 10.1109/JPROC.2013.2252316.

[13] S. Young et al, ”The HTK book,” Cambridge University Engineering Department, vol. 3, (175), pp. 12, 2002.

[14] P. Taylor, Text-to-Speech Synthesis. 2009Available: https://www.cambridge.org/core/books/texttospeech-synthesis/D2C567CEF939C7D15B2F1232992C7836.

[15] D. Mitrović, M. Zeppelzauer and C. Breiteneder, ”Chapter 3 - Features for Content-Based Audio Retrieval,” Advances in Computers, vol. 78, pp. 71-150, 2010. Available: https://www.sciencedirect.com/science/article/pii/S0065245810780037. DOI: https://doi.org/10.1016/S0065-2458(10)78003-7 ”.

[16] Xia Wang et al, ”Noise robust chinese speech recognition using feature vector normalization and higher-order cepstral coefficients,” in - WCC 2000 - ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000, 2000, . DOI: 10.1109/ICOSP.2000.891617.

[17] A. Benba, A. Jilbab and A. Hammouch, ”Voice assessments for detecting patients with Parkinson's diseases using PCA and NPCA,” Int. J. Speech Technol., vol. 19, (4), pp. 743-754, 2016.

[18] D. W. Robinson and R. S. Dadson, ”A re-determination of the equal-loudness relations for pure tones,” British Journal of Applied Physics, vol. 7, (5), pp. 166, 1956.

[19] H. Hermansky and N. Morgan, ”RASTA processing of speech,” IEEE Transactions on Speech and Audio Processing, vol. 2, (4), pp. 578-589, 1994.

[20] H. Hermansky et al, ”RASTA-PLP speech analysis technique,” in - [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992, . DOI: 10.1109/ICASSP.1992.225957.

[21] A. Kumar et al, ”Event detection in short duration audio using gaussian mixture model and random forest classifier,” in - 21st European Signal Processing Conference (EUSIPCO 2013), 2013.

[22] M. M. Jaber et al, ”A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms,” Measurement, vol. 162, pp. 107883, 2020. Available: https://www.sciencedirect.com/science/article/pii/S0263224120304218. DOI: https://doi.org/10.1016/j.measurement.2020.107883.

[23] L. Breiman, ”Random forests,” Mach. Learning, vol. 45, (1), pp. 5-32, 2001.

[24] V. Svetnik et al, ”Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” J. Chem. Inf. Comput. Sci., vol. 43, (6), pp. 1947-1958, 2003. Available: https://doi.org/10.1021/ci034160g. DOI: 10.1021/ci034160g.

[25] K. J. Archer and R. V. Kimes, ”Empirical characterization of random forest variable importance measures,” Comput. Stat. Data Anal., vol. 52, (4), pp. 2249-2260, 2008. Available: https://www.sciencedirect.com/science/article/pii/S0167947307003076. DOI: https://doi.org/10.1016/j.csda.2007.08.015.

[26] H. K. Vydana and A. K. Vuppala, ”Investigative study of various activation functions for speech recognition,” in - 2017 Twenty-Third National Conference on Communications (NCC), 2017, . DOI: 10.1109/NCC.2017.8077043.

[27] G. E. Dahl, T. N. Sainath and G. E. Hinton, ”Improving deep neural networks for LVCSR using rectified linear units and dropout,” in - 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, . DOI: 10.1109/ICASSP.2013.6639346.

[28] T. Fawcett, ”An introduction to ROC analysis,” Pattern Recog. Lett., vol. 27, (8), pp. 861-874, 2006. Available: https://www.sciencedirect.com/science/article/pii/S016786550500303X. DOI: https://doi.org/10.1016/j.patrec.2005.10.010.

[29] P. Bagad et al, ”Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds,” 2020. Available: https://arxiv.org/abs/2009.08790v2.

[30] C. Brown 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, 2020.

[31] G. Pinkas et al, ”SARS-CoV-2 Detection From Voice,” IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 268-274, 2020.

[32] D. Dinnes. How accurate are rapid tests for diagnosing COVID-19?. Available: /CD013705/INFECTN_how-accurate-are-rapid-tests-diagnosing-covid-19. DOI: 10.1002/14651858.CD013705.pub2.