



International Journal of Biomedical Engineering and Technology

ISSN online: 1752-6426 - ISSN print: 1752-6418
<https://www.inderscience.com/ijbet>

Speech signals as biomarkers: using glottal features for non-invasive COVID-19 testing

Girish Gidaye, Abhay Barage, Nirmayee Dighe, Kadria Ezzine, Varsha Turkar, Gajanan Nagare

DOI: [10.1504/IJBET.2024.10067292](https://doi.org/10.1504/IJBET.2024.10067292)

Article History:

Received:	11 May 2024
Last revised:	05 July 2024
Accepted:	22 July 2024
Published online:	06 January 2025

Speech signals as biomarkers: using glottal features for non-invasive COVID-19 testing

Girish Gidaye*

Department of Electronics and Computer Science,
Vidyalankar Institute of Technology,
Mumbai, India

Email: girish.gidaye@vit.edu.in

*Corresponding author

Abhay Barage

Department of Population Health Sciences,
Weill Cornell Medicine,
Cornell University,
1300 York Avenue, New York, NY 10065, USA
Email: agb4002@med.cornell.edu

Nirmayee Dighe

Department of Information Systems,
Northeastern University,
360 Huntington Ave, Boston, MA 02115, USA
Email: nirmayee.dighe1998@gmail.com

Kadria Ezzine

University of Lorraine,
Unnamed Road, 54000 Nancy, France
Email: kadria.ezzine@gmail.com

Varsha Turkar

Electronics and Telecommunication Engineering Department,
Vidyalankar Institute of Technology,
Mumbai, India
Email: varsha.turkar@vit.edu.in

Gajanan Nagare

Department of Biomedical Engineering,
Vidyalankar Institute of Technology,
Mumbai, India
Email: gajanan.nagare@vit.edu.in

Abstract: The COVID-19 pandemic was the most significant global health crisis in recent history, with lasting impacts on societies worldwide. Current screening methods are invasive, slow, frequently inaccurate, and limited in capacity. To overcome these limitations researchers used conventional features extracted from speech signals. In the proposed methodology, the change in vibratory pattern due to COVID-19 is captured by extracting glottal features from glottal signal acquired by inverse filtering. Various machine learning models like naïve Bayes, ADABOOST, gradient boost, decision tree, support vector machine (SVM), stochastic gradient descent (SGD), K-nearest neighbours, and random forest (RF) are used in this study. It is observed that the gradient boost gives the maximum classification accuracy (99.97%) for COVID-19 detection and SGD gives the maximum classification accuracy (99.76%) for severity grading. It is also observed that the time instants glottal features outperform frequency domain and model-based features. The Coswara database is used for this study.

Keywords: COVID-19; glottal signal; glottal features; speech signal; machine learning; quasi-closed phase; QCP; support vector machine; SVM; stochastic gradient descent; SGD.

Reference to this paper should be made as follows: Gidaye, G., Barage, A., Dighe, N., Ezzine, K., Turkar, V. and Nagare, G. (2025) ‘Speech signals as biomarkers: using glottal features for non-invasive COVID-19 testing’, *Int. J. Biomedical Engineering and Technology*, Vol. 47, No. 1, pp.65–85.

Biographical notes: Girish Gidaye is a Professor in the Department of Electronics and Computer Science at Vidyalankar Institute of Technology, Mumbai, India. He earned both his BE and ME degrees from the Shivaji University in Kolhapur, India. In December 2021, he was awarded a PhD from the K.J. Somaiya College of Engineering in Mumbai, India. His research interests encompass speech and image processing.

Abhay Barage completed his Bachelor’s in Biomedical Engineering in June 2024 from the Vidyalankar Institute of Technology, Mumbai, India. He has recently been accepted to Cornell University in New York to pursue his MS. His research interest focuses on developing applications for the biomedical field based on speech processing.

Nirmayee Dighe is a graduate research assistant at the Northeastern University. She has a Master’s in Information Systems from the Northeastern University and a Bachelor of Engineering in Electronics and Telecommunications from the University of Mumbai. Her research interests include statistical data analysis and processing.

Kadria Ezzine is a Post-doctoral researcher at the Lorraine University, having completed her PhD in Computer Science at the Carthage University, Tunisia. Her career encompasses diverse projects in signal and image processing, biomedical data analysis, and software development. Her extensive technical skills enable her to design and implement advanced IT solutions tailored to specific needs.

Varsha Turkar serves as a Professor and the Head at the Electronics and Telecommunication Engineering Department of Vidyalankar Institute of Technology, Mumbai, India. Her doctorate was conferred by the Indian Institute of Technology Bombay, India in 2013, following her completion of both a Bachelor’s and a Master’s degree from university of Mumbai. Her research interests include speech and image processing.

Gajanan Nagare is the Head and a Professor in the Department of Biomedical Engineering at Vidyalkar Institute of Technology, Mumbai, India. He is an active researcher in the area of Biomedical Engineering and Multidisciplinary field. His contribution majorly involves application of microdevices, embedded systems, and artificial intelligence for solving the issues in healthcare and environmental science. He received his PhD (Biomedical Engineering) from the Indian Institute of Technology, Bombay, India. He has done his Master's and Bachelor's in Instrumentation and Control from the Government College of Engineering, Pune and North Maharashtra University, Jalgaon, respectively.

1 Introduction

The COVID-19 pandemic has profoundly affected the world, caused significant human casualties, and triggered an economic crisis. The existing diagnostic methods for COVID-19 face multiple challenges, including invasiveness, diagnostic delays due to manual testing processes, false-negative results, reliance on COVID-19 testing kits, and risks of exposure. Consequently, there is a pressing need for new diagnostic approaches that are both more efficient and effective.

The process of phonation in speech production involves the vibration of the vocal folds and is a critical aspect of human vocalisation. Many COVID-19 patients experience respiratory dysfunction and have asynchronous, asymmetrical, and restrained vocal fold vibrations during phonation (Al Ismail et al., 2020). Along with voice alterations, many people often describe a general inability to maintain a normal voice. The researchers used speech signals to diagnose various diseases like vocal fold cyst, vocal paralysis, Laryngitis, Reink Edema, Nodules, Sulcus, Polyp, Dysphonia and COVID-19 (Shrivast et al., 2022; Zhu et al., 2023; Chetupalli et al., 2023; Sindhu and Sainin, 2024).

The proposed technique overcomes the limitations of existing kit-based screening methods. It allows individuals to conduct tests independently and in a remote environment. This paper uses machine learning algorithms to analyse the glottal vibration for the early detection and assessment of COVID-19 infection. The research utilises vowel /a/ voice recordings from the Coswara corpus (Sharma et al., 2020). The individuals in the dataset are classified into three classes – COVID negative, mild COVID positive or moderate COVID positive. Several machine learning algorithms like support vector machine (SVM), stochastic gradient descent (SGD), and random forest (RF), etc. are applied and a comparative analysis is conducted to identify the most suitable classifier.

This paper is structured as follows: Section 2 provides a comprehensive review of existing research on using machine learning techniques to detect COVID-19-related voice disorders. Section 3 delves into the specifics of this study, detailing the chosen dataset, extracted features, and evaluated algorithms. Section 4 discusses the results. Finally, Section 5 concludes the paper by summarising the key findings and potential future directions.

2 Related work

Researchers actively explored the potential of chest imaging for automatic COVID-19 screening, with studies analysing both chest radiographs (Wang et al., 2020; Afshar et al., 2020; Farooq and Hafeez, 2020; Nguyen et al., 2023; Fan and Gong, 2024) and computed tomography (CT) scans (Ng et al., 2020; Xu, 2020; Ni et al., 2020; Lin et al., 2024) to identify characteristic COVID-19 related lung abnormalities. These are clinical methods and must be performed under medical supervision.

Alternatively, several studies investigated the analysis of cough sounds (Laguarta et al., 2020; Imran et al., 2020; Andreu-Perez et al., 2020; Subirana et al., 2020; Pahar et al., 2020; Benmalek et al., 2023) and respiratory signals (Trivedy et al., 2020; Hassan et al., 2020; Hossain and Shorfuzzaman, 2023; Bhaskar et al., 2023) as potential tools to detect COVID-19 induced changes.

Several researches suggest using voice samples to detect COVID-19. Han et al. (2020) investigated the possibility of using voice analysis to assess the severity of COVID-19. Performed analysis on 52 patients from two hospitals in Wuhan and collected recordings of five sentences spoken by everyone. They then employed two sets of acoustic features, along with a SVM algorithm, to analyse the voice samples and categorise the illness severity. This approach achieved an accuracy of 69% in differentiating between different levels of illness. Alternately Venkata et al. (2020) used Mel filter bank features as input for SVM. The researchers employed a small dataset (ten pathological and nine healthy subjects) collected from YouTube videos and achieved an accuracy of 70.5% and F1-score of 77%. Deshmukh et al. (2020) proposed a convolutional neural network (CNN) model to detect anomalies in the voice source. The model analysed the glottal flow waveform (GFW) of 19 voice samples (ten healthy and nine with suspected illness) uttering the vowels /a/, /i/, and /u/. The samples were obtained from a private database collected in Chile. The model's performance, evaluated using the area under the receiver operating characteristic curve, achieved a score of 0.900 with a standard deviation of 0.062, demonstrating its effectiveness in analysing the vowels /i/ and /u/. Chetupalli et al. (2023) developed and assessed a COVID-19 diagnosis technique based on multi-modal acoustic and symptom data. They acquired data from the open-access Coswara repository. They investigated the use of various modalities, specifically breathing, coughing, speaking, and symptoms for COVID-19 prediction. Ponomarchuk et al. (2022) created and launched a mobile application that detects COVID-19 infection by combining a symptom checker with speech, breath, and cough signals. Three publicly available databases namely, EPFL COUGHVID, Coswara, and COVID-19 cough were used to test the proposed model. Recently, Zhu et al. (2023), proposed a new modulation spectral feature and linear prediction analysis to detect COVID-19 using two-stage classification framework. They evaluated on three diverse datasets and cross dataset experiments were performed to develop a generalised model.

Fine fluctuations in airflow dynamics within the upper respiratory tract or minor impairments in laryngeal muscles can disrupt the vocal folds' oscillation patterns, symmetry of motion, and frequency synchronisation. These disturbances may lead to asymmetrical, asynchronous vocal fold vibrations and unstable eigen-modes. COVID-19 has been clinically observed to adversely affect the lower and mid-respiratory tract, including vital structures like the lungs, airways, and associated musculature (Al Ismail et al., 2020). Symptomatic patients with confirmed COVID-19 infections frequently report vocal changes and a general difficulty in normal voice production. This suggests

potential anomalies in vocal fold oscillation patterns during phonation, potentially offering a detectable signature of COVID-19 infection (Deshmukh et al., 2020). Inspired by this, the proposed technique in this paper aims to utilise the glottal excitation for pathological voice detection. The signals are extracted coming out of glottis by glottal inverse filtering to capture the altered dynamics of the vocal folds to detect COVID-19. The glottal signals are then quantified by extracting various features and classified as COVID positive and negative along with severity grading. The results are compared with the existing techniques which use speech signal instead of glottal signals. It is observed that the accuracy of the proposed methodology is higher than the existing techniques.

3 Materials and methods

Voice samples from the Coswara dataset are used in this work which is publicly available collection of coughing, breathing, and voice recordings from both healthy individuals and those with medical conditions (Sharma et al., 2020). The Coswara dataset is generated via crowd sourcing; it contains a lot of noisy samples. Hence, voice samples with minimal noise contamination are chosen. Notably, only vowel /a/ is analysed in this investigation, aligning with established medical practices where vowel characteristics are assessed to evaluate potential voice disorders. The vowel /a/ recordings of 152 individuals are included and divided equally into healthy and COVID-19 positive (mild and moderate) groups. The samples, extracted from the dataset comprises of 41 females and 111 males with an average age of 31 years. Additionally, research suggests that sustained vowels offer greater analytical value compared to continuous speech (Verde et al., 2021a, 2021b). Further details regarding the sample demographics are presented in Table 1.

Table 1 Details about the individuals involved in this study

<i>Category of the recorded subjects</i>					
<i>Healthy</i>		<i>COVID positive</i>			
<i>Male</i>	<i>Female</i>	<i>Mild</i>		<i>Moderate</i>	
		<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
57	19	22	8	35	11

The sounds of vowel /a/ are processed for all the samples to extract the features which are the inputs for machine learning algorithms. Compared to the speech signal, the glottal source is an elementary time-domain signal due to the absence of vocal tract resonances (Juvela et al., 2019). Building on recent research (Narendra and Alku, 2020), which suggests that the quasi-closed phase (QCP) analysis method offers more precise estimates of the voice source. Prior work (Airaksinen et al., 2014) has demonstrated the superiority of QCP analysis in evaluating glottal flow. This evaluation is conducted using two types of synthetic vowels Liljencrants-Fant (LF) model and a physical human voice production model. Inspired by these findings, the proposed methodology hypothesises that incorporating QCP analysis within the traditional pipeline system could lead to an improved classification accuracy for COVID-19 detection.

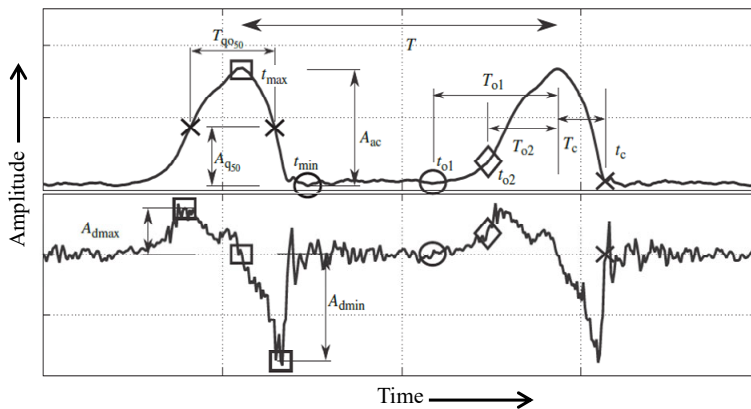
Various types of glottal wave parameterisation techniques are already developed (Airas, 2008). The direct method estimates time-based and spectral based measurements from a glottal source. Whereas the indirect method measurements are derived by fitting a

parametric model on glottal pulse. This paper utilises a total of 30 time-domain, spectral and model-based features, which are computed using the QCP algorithm (Airaksinen et al., 2014) as the glottal inverse filtering method. These 30 features are explained in detail in the following section.

3.1 Time-domain features

The classical time-based features are open quotient (OQ), speed quotient (SQ) and closing quotient (CIQ). OQ is the ratio of duration for which vocal fold remains open to the length of fundamental period and it quantifies the vibrations of vocal cords. SQ measures the ratio between the opening phases to the closing phase time. The highest excitation is applied to vocal tract during the closing phase of the vocal folds. The CIQ feature is the relation between the vocal fold closing time and length of fundamental period. It is also important in quantifying glottal excitation (Holmberg et al., 1988). The commonly used amplitude parameters are amplitude quotient (AQ) and its normalised form, the normalised amplitude quotient (NAQ) (Alku et al., 2002). These both measures are robust compared to time-based parameters. The AQ gives the relation between ac flow and the maximum air flow declination rate. The various time instants like the maximum flow ($t_{\max}$), the minimum flow ($t_{\min}$), the time instant of derivative maximum flow ($t_{d\max}$), the minimum flow derivative ($t_{d\min}$), the closure time instant (t_c), the primary opening instant (t_{p0}), the secondary opening instant (t_{s0}), the quasi-opening (t_{q0}) and quasi-closing time instants (t_{qc}) are estimated from the glottal volume velocity waveform (glottal flow). These time instants are shown in Figure 1. After finding the time and amplitude instants from glottal flow and its derivative wave, various time domain parameters are derived. The definitions of calculated parameters are as follows (Airas, 2008).

Figure 1 Various amplitude and time moments marked on glottal wave and its derivative shown in upper and lower pane respectively



Source: Airas (2008)

$$OQ_1 = \frac{t_c - t_{o1}}{T} \quad (1)$$

$$OQ_2 = \frac{t_c - t_{t02}}{T} \quad (2)$$

$$OQ_a = A_{ac} \left(\frac{\pi}{2A_{d \max}} + \frac{1}{A_{d \min}} \right) f_0 \quad (3)$$

$$QOQ = \frac{t_{qc} - t_{qo}}{T} \quad (4)$$

$$SQ_1 = \frac{t_{\max} - t_{01}}{t_c - t_{\max}} \quad (5)$$

$$SQ_2 = \frac{t_{\max} - t_{02}}{t_c - t_{\max}} \quad (6)$$

$$CIQ = \frac{t_c - t_{\max}}{T} \quad (7)$$

$$AQ = \frac{A_{ac}}{A_{d \min}} \quad (8)$$

$$NAQ = \frac{AQ}{T} \quad (9)$$

3.2 Frequency-domain features

Three different frequency domain parameters are used for quantisation of glottal source are harmonic richness factor (HRF), harmonic level difference (DH_{12}), parabolic spectral parameter (PSP). The HRF measures relation between sum of higher amplitude and the first harmonic. Among these three measures DH_{12} is straight forward to calculate and widely used in quantification of glottal source (Alku, 2011). This harmonic level difference is calculated by simply subtracting the amplitude of second harmonic from the first one ($H_1 - H_2$) as shown in Figure 2.

$$DH_{12} = (H_1 - H_2) \quad (10)$$

$$HRF = \frac{\sum_{k \geq 2} H_k}{H_1} \quad (11)$$

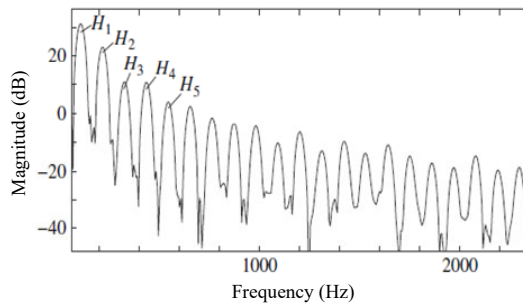
Various temporal and spectral domain measures fall within the category of direct measures estimated from the glottal pulse and its derivative.

3.3 Model based features

An alternative method to extract the relevant information embedded in the glottal pulse is by fitting the synthetic pulse derived from the LF model to each single glottal pulse. This is referred to as model-based parameterisation and extracted metrics are called model-based measurements. The LF model is very common in voice flow analysis and

hence considered by Rieglesberger and Krishnamurthy (1993) to model the glottal pulse. They have used time domain fitting algorithms: Prony and gradient descent on natural as well as on synthetic speech. The Prony fitting algorithm performs better than gradient descent for synthetic speech, however, it shows worse fitting on natural speech. The gradient descent technique needs the manual tuning of parameters to get perfect fit, whereas Prony algorithm is automatic. Iseli and Alwan (2000) proposed a technique based on minimum square error minimisation to fine-tune the extracted glottal wave to the LF model. The spectral tilt, open quotient and perceptually important R-parameter R_d are estimated from the fitted model and are shown to follow the same trend as in the existing work.

Figure 2 Spectrum of glottal source with the harmonic levels represented as H_1 to H_5



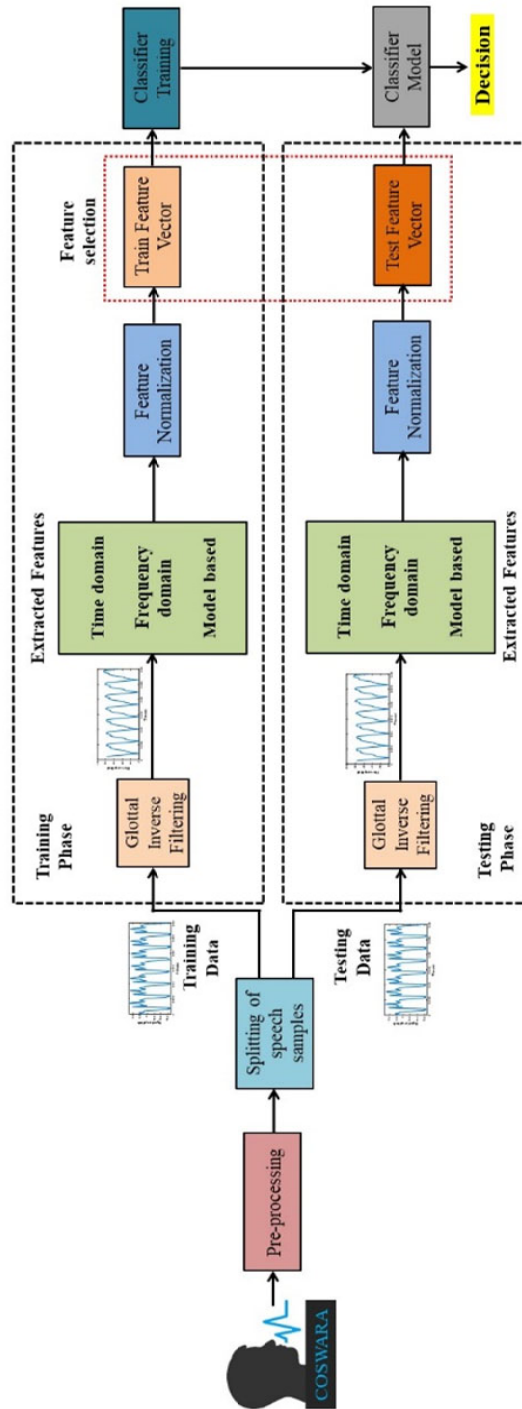
Source: Airas (2008)

In this work a TKK tool (Airas, 2008) is employed to extract the LF model-based measures. This tool uses a modified version of standard model fitting algorithm proposed by Strik and Boves (1994). They used two steps to refine the parameters. This algorithm requires finding a preliminary set of measures centred on the LF model. First, low-pass filtering process is performed on differentiated glottal pulse to prevent the undesirable effect of high-frequency (HF) components on the fitting. The fitting of synthetic glottal pulse on each extracted glottal pulse initiates from the glottal closing instant (GCI). Then the pulse is searched for the maximum negative amplitude at the vicinity of GCI location and that time instant is marked as t_e and the corresponding amplitude noted as EE . After that, a part of the pulse on the left-hand side of the t_e is screened to locate the very first zero-crossing and this point is marked as t_p . The search on the left side continues until the pulse drops below a small threshold of 0.1 times the peak amplitude. Then t_a value is estimated by assuming $RA = 0.02$. These initial estimated measures are refined in two steps using root mean square error measure. The first step involves the use of simplex technique, and second step uses steepest descent process (Timcke et al., 1958).

The synthetic pulse derived from the LF model can be further described using a set of three dimensions: RG , RA and Rk . These parameters are commonly referred to as R parameters and are computed using the relative time instants (T parameters) of the glottal pulse. The parameter RG quantifies the rising speed of the pulse and is estimated by (Timcke et al., 1958):

$$RG = \frac{T_0}{2T_p} \quad (12)$$

Figure 3 Glottal signal-based framework for a COVID-19 assessment (see online version for colours)



The second parameter RA is the normalised version of the return phase time and is computed as:

$$RA = \frac{T_a}{T_0} \quad (13)$$

The third parameter Rk characterises the glottal pulse symmetry and is calculated as:

$$Rk = \frac{t_e - t_p}{t_p} \quad (14)$$

Other than R parameters, one more dimension termed as open quotient (OQ) and its normalised version OQ' can be estimated from LF model (Timcke et al., 1958). This may be suitable for discerning breathy and tense voice (Bernardoni et al., 2004). It quantifies the proportion of the glottal cycle for which the glottis is open and is defined as:

$$OQ = \frac{T_e}{T_0} = 1 + \frac{Rk}{2RG} \quad (15)$$

3.4 Methodology

This section presents a framework for screening COVID-19 individuals based on glottal features. The systematic workflow of the voice evaluation scheme is visually depicted in Figure 3. As shown in the figure before extracting the features the speech signal is pre-processed. Pre-processing involves three key steps: normalising the speech signal amplitude, framing, and windowing. The proposed framework has a two phased approach: training and testing. Speech samples of both healthy and pathological speakers are divided into training set (70%) and testing set (30%). In the training stage, a classification model, validated through a 10-fold stratified cross-validation, is constructed using both healthy and pathogenic speech training recordings. In the evaluation phase, speech recordings which are not used during training are tested on the cross-validated model.

Careful selection of optimum features to get good classification accuracy is very important. Applying the information gain (IG) centred feature selection method; the most effective subset of features is carefully selected while excluding noisy, redundant, and irrelevant ones. A min/max normalisation process is applied to standardise all features, confining to the $[0, 1]$ range before applying to the classifier. Following training data normalisation, the identified min and max values for each feature are leveraged to standardise the test data, allowing for comparable analysis. In this work, several machine learning algorithms are used to distinguish between a pathological and a healthy voice. The Orange data mining (Demsar et al., 2013) tool, version 3.36.2, is used to perform the analyses. All the experiments are performed on a machine with 8GB RAM and Intel ® Pentium ® quad core processor N3540 with 2.66 GHz.

4 Results and discussion

Machine learning algorithms are used in the proposed technique for COVID-19 detection through acoustic analysis of speech signals. A two-part experimental design is implemented. The first experiment adopted a binary classification approach, differentiating between individuals confirmed positive for COVID-19 and those testing negative. The second experiment utilised a multi-class classification strategy, categorising individuals as negative (healthy), positive with mild severity, or positive with moderate severity. The utilisation of a multi-class approach in the second experiment shows potential for evaluating COVID-19 severity levels, which could assist in customising rehabilitation programs and tracking patient recovery.

Table 2 T-test results for glottal features: binary classification

<i>Rank</i>	<i>Feature</i>	<i>t</i>	<i>p</i>	<i>Rank</i>	<i>Feature</i>	<i>t</i>	<i>p</i>	<i>Rank</i>	<i>Feature</i>	<i>t</i>	<i>p</i>
1	<i>t_{po}</i>	513.8	0.000	11	<i>OQ</i>	199.9	0.000	21	<i>PSP</i>	014.3	0.000
2	<i>t_{so}</i>	516.1	0.000	12	<i>QO</i>	202.1	0.000	22	<i>CIQ</i>	013.2	0.000
3	<i>t_c</i>	655.4	0.000	13	<i>RK</i>	053.5	0.000	23	<i>SQ1</i>	007.9	0.000
4	<i>t_{max}</i>	594.5	0.000	14	<i>Ee</i>	010.8	0.000	24	<i>t0</i>	001.0	0.250
5	<i>t_{min}</i>	652.5	0.000	15	<i>tp</i>	033.9	0.000	25	<i>SQ2</i>	006.8	0.000
6	<i>t_{dmin}</i>	631.8	0.000	16	<i>RA</i>	023.3	0.000	26	<i>DH12</i>	007.9	0.000
7	<i>t_{dmax}</i>	631.8	0.000	17	<i>NAQ</i>	030.5	0.000	27	<i>AQ</i>	008.3	0.000
8	<i>t_{qo}</i>	543.9	0.000	18	<i>te</i>	018.0	0.000	28	<i>HRF</i>	004.8	0.000
9	<i>t_{qc}</i>	628.1	0.000	19	<i>OQ1</i>	001.8	0.000	29	<i>QOQ</i>	005.9	0.000
10	<i>RG</i>	229.7	0.000	20	<i>OQ2</i>	001.8	0.000	30	<i>OQA</i>	001.2	0.220

Note: Features marked in italic are having $p \leq 0.05$.

Table 3 ANOVA-test results for glottal features: multiclass classification

<i>Rank</i>	<i>Feature</i>	<i>F-ratio</i>	<i>p</i>	<i>Rank</i>	<i>Feature</i>	<i>F-ratio</i>	<i>p</i>	<i>Rank</i>	<i>Feature</i>	<i>F-ratio</i>	<i>p</i>
1	<i>t_{qc}</i>	8,109.5	0.000	11	<i>t0</i>	12,750	0.000	21	<i>SQ2</i>	0101.2	0.000
2	<i>t_{max}</i>	7,530.6	0.000	12	<i>RG</i>	0959.1	0.000	22	<i>SQ1</i>	0116.9	0.000
3	<i>t_{so}</i>	6,388.4	0.000	13	<i>QO</i>	0440.2	0.000	23	<i>QOQ</i>	0012.5	0.000
4	<i>t_{min}</i>	6,704.0	0.000	14	<i>OQ</i>	0436.0	0.000	24	<i>te</i>	0073.9	0.000
5	<i>t_c</i>	6,715.8	0.000	15	<i>Ee</i>	0111.4	0.000	25	<i>AQ</i>	0053.1	0.000
6	<i>t_{qo}</i>	5,827.9	0.000	16	<i>OQ1</i>	0018.8	0.000	26	<i>PSP</i>	0038.3	0.000
7	<i>t_{dmin}</i>	5,503.0	0.000	17	<i>OQ2</i>	0018.8	0.000	27	<i>tp</i>	0005.5	0.064
8	<i>t_{dmax}</i>	5,503.0	0.000	18	<i>DH12</i>	0480.9	0.000	28	<i>CIQ</i>	0023.3	0.000
9	<i>t_{po}</i>	4,847.5	0.000	19	<i>NAQ</i>	0196.2	0.000	29	<i>HRF</i>	0010.5	0.000
10	<i>RK</i>	3,010.8	0.000	20	<i>RA</i>	0058.6	0.000	30	<i>OQA</i>	0002.6	0.069

Note: Features marked in italic are having $p \leq 0.05$.

Figure 4 Kernel density distribution plots: healthy vs. COVID positive (see online version for colours)

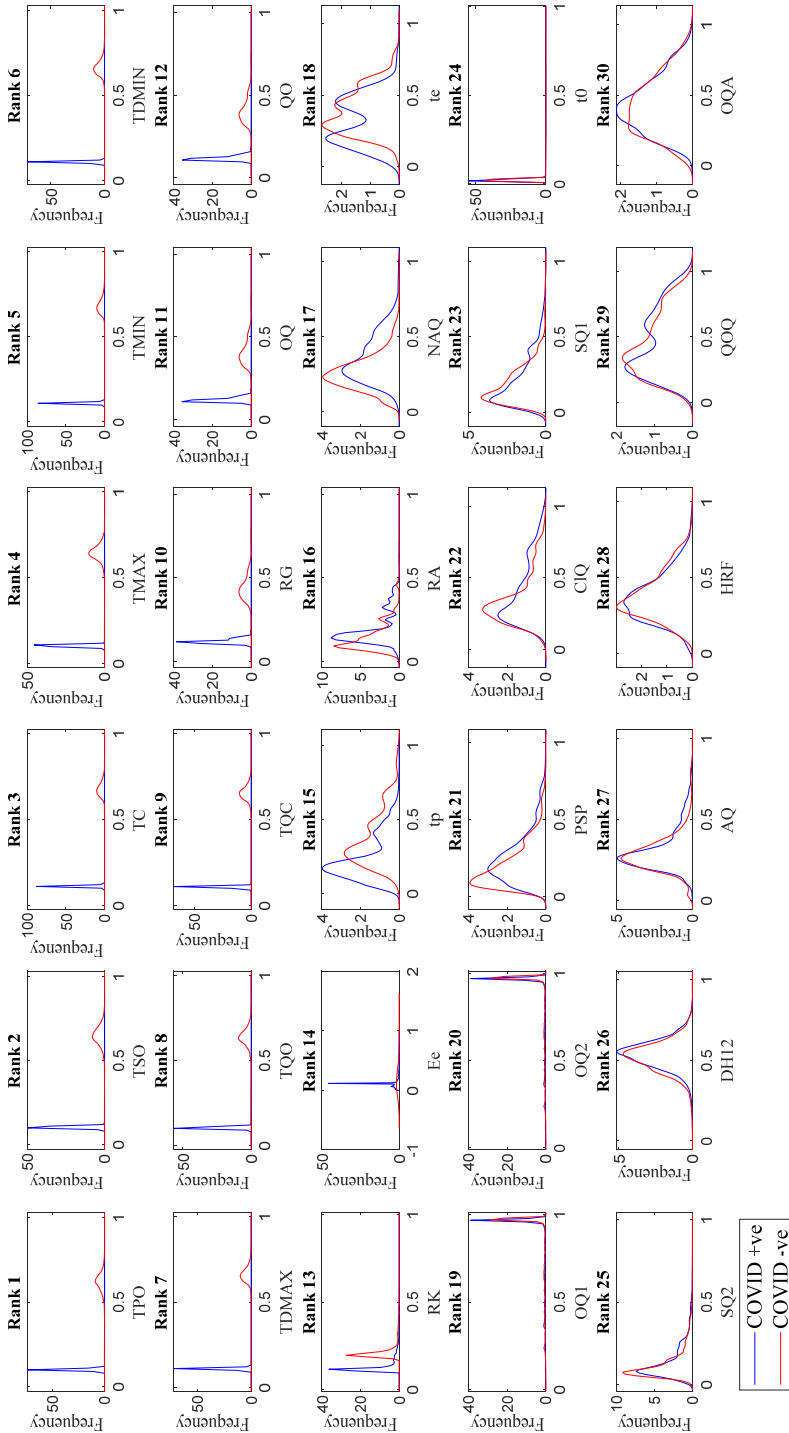


Figure 5 Kernel density distribution plots: healthy vs. COVID positive moderate vs. COVID positive mild (see online version for colours)

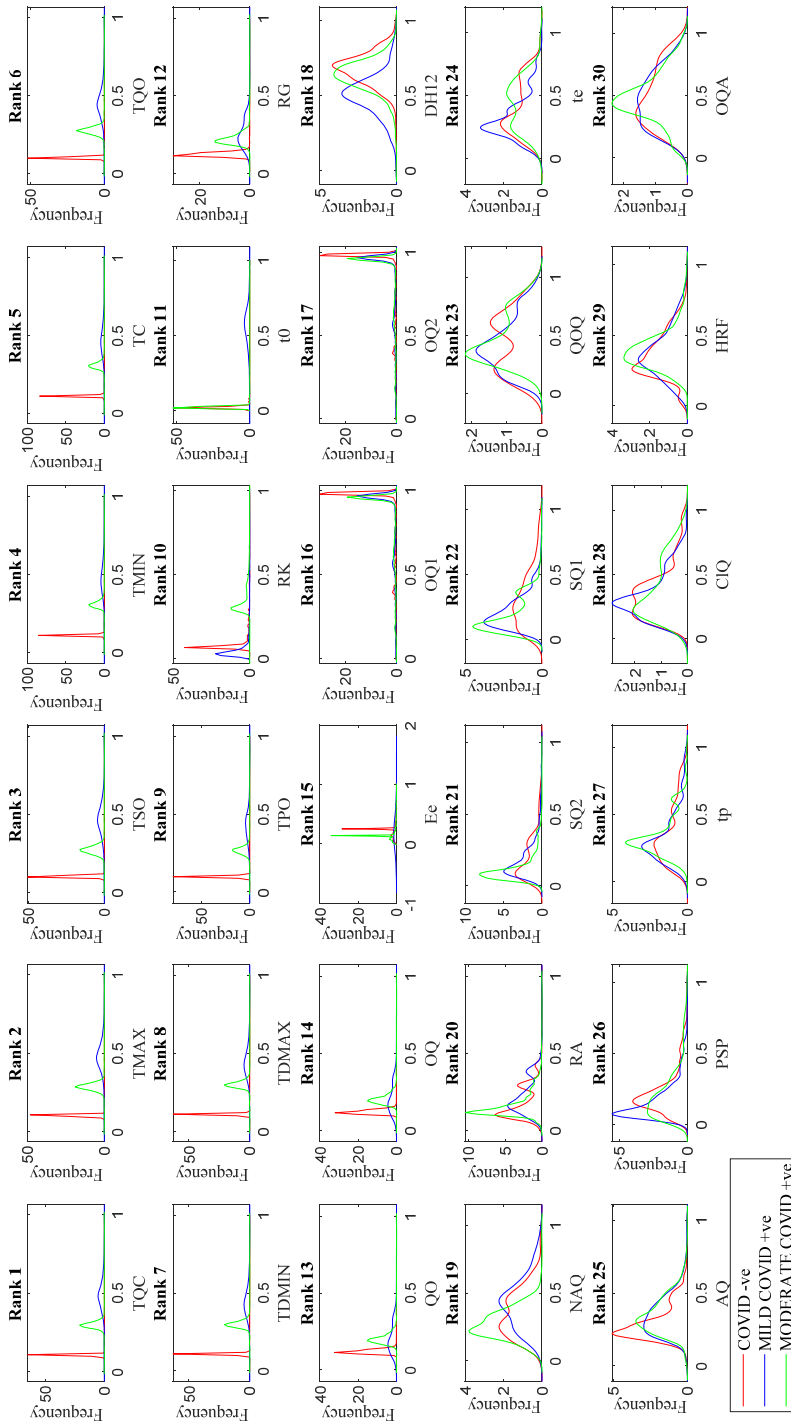


Table 4 Binary classifier performance: healthy vs. COVID +ve

Model	AUC		CA		F1		PPV		Recall	
	Top 3	All 30	Top 3	All 30	Top 3	All 30	Top 3	All 30	Top 3	All 30
kNN	0.9959	0.9991	0.9847	0.9931	0.9844	0.9931	0.9843	0.9935	0.9844	0.9831
Decision tree	0.9888	0.9971	0.9870	0.9970	0.9869	0.9969	0.9873	0.9973	0.9870	0.9977
SVM	0.9997	1.0000	0.9970	0.9981	0.9969	0.9981	0.9973	0.9981	0.9970	0.9981
SGD	0.9961	1.0000	0.9981	0.9981	0.9971	0.9981	0.9914	0.9981	0.9981	0.9981
Random forest	1.0000	1.0000	0.9969	0.9989	0.9969	0.9987	0.9969	0.9982	0.9969	0.9989
Naïve Bayes	0.9924	0.9990	0.9922	0.9940	0.9926	0.9740	0.9924	0.9954	0.9926	0.9940
AdaBoost	1.0000	0.9999	0.9969	0.9994	0.9969	0.9990	0.9969	0.9994	0.9969	0.9994
Gradient boost	0.9891	1.0000	0.9817	0.9997	0.9817	0.9997	0.9817	0.9991	0.9817	0.9991

The postulated glottal attributes are tested for significance using a t-test between the two groups (positive and negative) of individuals. The Welch's t-test is a parametric test that compares means from two distinct groups without supposing equal population variances (Welch, 1947). Table 2 presents the statistical test results for the 30 features derived from the glottal wave segments in the context of binary classification. For multiclass classification, an ANOVA test was performed, and Table 3 displays the corresponding statistical test results for the same 30 features. A significance level of 95% ($p \leq 0.05$) was employed. In both binary and multiclass classification scenarios, 28 out of the 30 features have p-values ≤ 0.05 , and these significant features are highlighted in italic.

To better illustrate the discriminatory capability of attributes, kernel density distribution plots are generated and displayed in Figures 4 and 5 for binary and multiclass classification, respectively. Also to arrange the attributes in descending order as per their discrimination ability, IG attributes ranking technique is harnessed. It is observed that the vocal fold opening and closing instances have achieved top ranking.

Our objective is to utilise the optimum features necessary to differentiate between non-COVID and COVID individuals and to grade the severity of COVID cases. Thus, a series of experiments are conducted by considering the top-ranked 3, 5, 10, 15, 20, 25, and all 30 features. It is observed that the performance metrics are nearly identical for most machine learning models for the top three ranked features and with all 30 features. Consequently, Tables 4 and 5 present the machine learning model metrics using only the top three ranked features alongside all 30 features. The top three ranked features performed exceptionally well in discernment and severity assessment, as evidenced by the 3D scatter plots presented in Figures 6 and 7. Using the top three ranked features, the SGD model achieved the highest (0.9981) classification rate for binary classification. However, for multiclass classification, the kNN model attained the highest (0.9831) classification rate. With all 30 features in use, the AdaBoost model achieved the highest (0.9994) classification rate for binary classification. Conversely, for multiclass classification, the SGD model attained the highest (0.9976) classification rate.

Figure 6 3D scatter plot with top three features: healthy vs. COVID positive (see online version for colours)

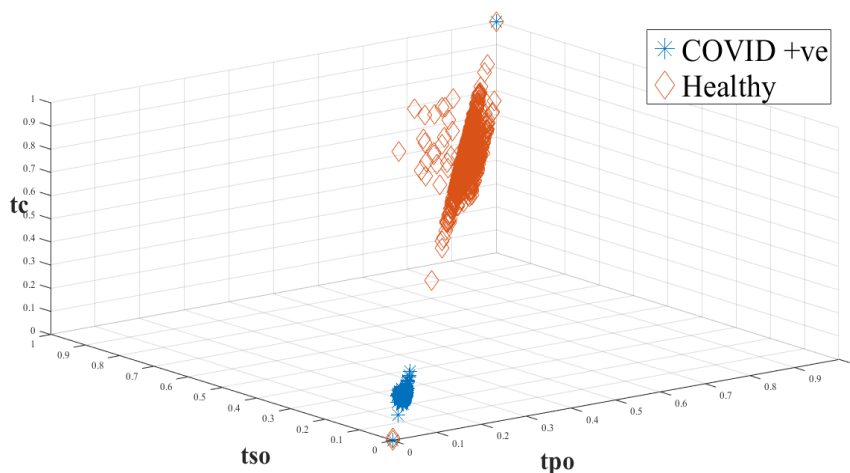


Table 5 Multiclass classifier performance

Model	AUC		CA		F1		PPV		Recall	
	Top 3	All 30	Top 3	All 30	Top 3	All 30	Top 3	All 30	Top 3	All 30
kNN	0.9951	1.0000	0.9831	0.9923	0.9831	0.9923	0.9835	0.9925	0.9831	0.9923
Decision tree	0.9870	0.9969	0.9770	0.9954	0.9769	0.9954	0.9773	0.9954	0.9770	0.9954
SVM	0.9933	1.0000	0.9770	0.9969	0.9769	0.9960	0.9773	0.9970	0.9770	0.9969
SGD	0.9461	0.9999	0.9281	0.9976	0.9271	0.9976	0.9314	0.9976	0.9281	0.9976
Random forest	0.9947	1.0000	0.9847	0.9919	0.9846	0.9917	0.9847	0.9922	0.9847	0.9919
Naïve Bayes	0.9424	0.9990	0.8226	0.9740	0.8261	0.9740	0.8810	0.9754	0.8226	0.9740
AdaBoost	0.9863	0.9995	0.9541	0.9954	0.9541	0.9950	0.9541	0.9954	0.9541	0.9954
Gradient boost	0.9891	1.0000	0.9817	0.9917	0.9817	0.9917	0.9817	0.9921	0.9817	0.9921

Figure 7 3D scatter plot with top three ranked features: healthy vs. moderate COVID +ve vs. mild COVID +ve (see online version for colours)

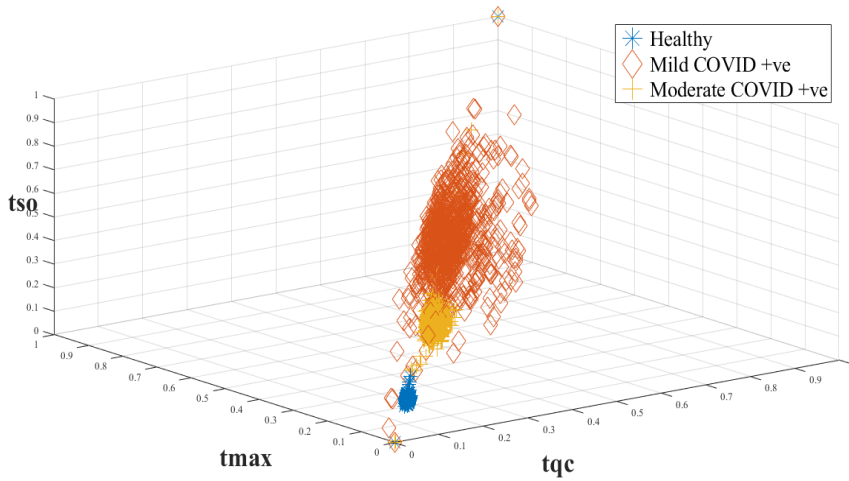


Table 6 Comparative analysis of existing and proposed method

<i>Ref. no.</i>	<i>Descriptors used</i>	<i>Database</i>	<i>Classifier</i>	<i>Best conveyed metric</i>
Zhu et al. (2023)	Modulation spectrum and linear prediction speech features	DiCOVA2	SVM	UAR = 0.7700
		CSS		UAR = 0.6800
		Cambridge		UAR = 0.6200
Verde et al. (2021a)	Fundamental frequency, shimmer, jitter, HNR, MFCC, Δ , $\Delta\Delta$, SC and SR	Coswara	SVM	Accuracy = 0.9700
Dash et al. (2022)	Spectral features, cepstral features, spectral descriptors, and periodicity features	Coswara	LGM	Accuracy = 0.9900
			SVM	Accuracy = 0.5000
			RF	Accuracy = 0.9500
			KNN	Accuracy = 0.5500
Verde et al. (2021b)	Fundamental frequency, shimmer, jitter, HNR, MFCC, Δ , $\Delta\Delta$, SC and SR	Coswara	SVM	Accuracy = 0.9900
			SGD	Accuracy = 0.5000
			Bagging	Accuracy = 0.9500
Proposed approach	Glottal features: time domain, frequency domain and model-based features	Coswara	SVM	Accuracy = 0.9981
			SGD	Accuracy = 0.9981
			Random forest	Accuracy = 0.9989
			AdaBoost	Accuracy = 0.9994
			Gradient Boost	Accuracy = 0.9997

As mentioned in the earlier section, only 152 samples were selected to train and test the ML model. Future efforts will be taken to collect a greater number of noise free speech samples for males and females to validate the claims made by the proposed method. Currently, all available databases rely on crowdsourcing, with volunteers recording

samples independently. To validate our approach for early COVID-19 detection, a controlled clinical trial is necessary. This ensures samples are labelled by medical professionals.

Table 6 presents a comparison between the accuracies obtained by the proposed and existing work. Various classifiers like SVM, SGD, RF, kNN and AdaBoost are applied on the extracted glottal features, and it is observed that all the classifiers give accuracy more than 0.99 indicating that use of glottal features are more effective than the features extracted from the speech signals (traditional methods).

5 Conclusions

This paper examines how glottal feature sets can identify changes in vocal fold vibratory patterns caused by both symptomatic and asymptomatic COVID-19 infections. Changes in the timing of vocal fold opening and closure, as well as time domain, frequency domain, and model-based characteristics, are explored. It is observed that the accuracy of the performance matrix is almost alike for the top three-ranked and with all 30 features. Hence the computational complexity can be reduced by using top three ranked features instead of all features.

The proposed glottal features-based system outperforms traditional benchmarks such as short time, long time, spectral, and cepstral features. The results indicate that time instances and time domain glottal features perform better than frequency domain and model-based features. The performance is assessed on various metrics such as accuracy, sensitivity, specificity, F1-score, and area under the curve. The SVM, SGD, RF, AdaBoost, and kNN approach demonstrated high reliability in detecting COVID-19 illnesses, with an accuracy of more than 99%. The outcome indicates the high reliability and efficiency of the proposed method, which could potentially offer a non-invasive, low cost, in-house and rapid diagnostic solution for COVID-19. The results are encouraging as the extraction of voice source information is found to be most effective in detection and to grade the severity of COVID-19. It is observed that the gradient boost gives the maximum classification accuracy (99.97%) for COVID-19 detection and SGD gives the maximum classification accuracy (99.76%) for severity grading.

The suggested technique is effective for the current dataset which has only two main classes COVID-19 positive and negative. In future, the robustness of the proposed technique could be evaluated by testing it on a dataset containing samples for COVID-19 positive cases, negative cases, and other respiratory disorders to check whether glottal features are sufficient to separate COVID-19 positive cases from other respiratory disorders. It is crucial to develop a generalised machine learning model which will be independent of disease, gender, age, regional accent, and regional language.

Acknowledgements

We express our gratitude to Indian Institute of Science (IISc), Bengaluru for providing the data access (COSWARA).

References

- Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K. and Mohammadi, A. (2020) *COVID-CAPS: A Capsule Network-based Framework for Identification of COVID-19 Cases from X-ray Images* [online] <http://arxiv.org/abs/2004.02696> (accessed 4 June 2023).
- Airaksinen, M., Raitio, T., Story, B. and Alku, P. (2014) 'Quasi closed phase glottal inverse filtering analysis with weighted linear prediction', *IEEE/ACM Transactions on Audio Speech and Language Processing*, Vol. 22, No. 3, pp.596–607.
- Airas, M. (2008) 'TKK Aparat: an environment for voice inverse filtering and parameterization', *Logopedics Phoniatrics Vocology*, Vol. 33, No. 1, pp.49–64.
- Al Ismail, M., Deshmukh, S. and Singh, R. (2020) *Detection of COVID-19 through the Analysis of Vocal Fold Oscillations*, [online] <http://arxiv.org/abs/2010.10707> (accessed 4 June 2023).
- Alku, P. (2011) 'Glottal inverse filtering analysis of human voice production – a review of estimation and parameterization methods of the glottal excitation and their applications', *Sadhana*, Vol. 36, No. 5, pp.623–650.
- Alku, P., Backstrom, T. and Vilkman, E. (2002) 'Normalized amplitude quotient for parametrization of the glottal flow', *Journal of Acoustical Society of America*, Vol. 112, No. 2, pp.701–710.
- Andreu-Perez, J., Pérez-Espinos, H., Timone, E., Girón-Pérez, M., Kiani, M., Benitez-Trinidad, A., Jarchi, D., Rosales-Pérez, A., Ali, Z. and Gatzoulis, N. (2020) *A Novel Deep Learning-Based Recognition Method and Web-App for COVID-19 Infection Test from Cough Sounds with a Clinically Validated Dataset*, School of Computer Science and Electronic Engineering, Univ. Essex, Colchester, U.K., Tech. Rep.
- Benmalek, E., Elmhadi, J., Jilbab, A. and Jbari, A. (2023) 'Automatic COVID-19 detection using machine learning and voice recording', *Research on Biomedical Engineering*, Vol. 39, pp.597–612.
- Bernardoni, N., d'Alessandro, C., Doval, B. and Castellengo, M. (2004) 'On the use of the derivative of electroglottographic signals for characterization of non-pathological phonation', *Journal of the Acoustical Society of America*, Vol. 115, No. 3, pp.1321–1332.
- Bhaskar, N., Bairagi, V., Munot, M., Gaikwad, K. and Jadhav, S. (2023) 'Automated COVID-19 detection from exhaled human breath using CNN-CatBoost ensemble model', *IEEE Sensors Letters*, Vol. 7, No. 10, pp.1–4.
- Chetupalli, S., Krishnan, P., Sharma, N., Muguli, A., Kumar, R., Nanda, V., Pinto, L., Ghosh, P. and Ganapathy, S. (2023) 'Multi-modal point-of-care diagnostics for COVID-19 based on acoustics and symptoms', *IEEE Journal of Translational Engineering in Health and Medicine*, Vol. 11, pp.199–210.
- Dash, T., Chakraborty, C., Mahapatra, S. and Panda, G. (2022) 'Gradient boosting machine and efficient combination of features for speech-based detection of COVID-19', *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, No. 11, pp.5364–5371.
- Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M. and Zupan, B. (2013) 'Orange: data mining toolbox in Python', *Journal of Machine Learning Research*, Vol. 14, pp.2349–2353.
- Deshmukh, S., Al Ismail, M. and Singh, R. (2020) *Interpreting Glottal Flow Dynamics for Detecting COVID-19 from Voice* [online] <http://arxiv.org/abs/2010.16318> (accessed 4 June 2023).
- Fan, Y and Gong, H. (2024) 'An improved COVID-19 classification model on chest radiography by dual-ended multiple attention learning', *IEEE Journal of Biomedical and Health Informatics*, Vol. 28, No. 1, pp.145–156.
- Farooq, M and Hafeez, A. (2020) *COVID-ResNet: A Deep Learning Framework for Screening of COVID-19 from Radiographs* [online] <http://arxiv.org/abs/2003.14395> (accessed 6 June 2023).

- Han, J., Qian, K., Song, M., Yang, Z., Ren, Z., Liu, S., Liu, J., Zheng, H., Ji, W., Koike, T., Li, X., Zhang, Z., Yamamoto, Y. and Schuller, B. (2020) *An Early Study on Intelligent Analysis of Speech under COVID-19: Severity, Sleep Quality, Fatigue and Anxiety* [online] <http://arxiv.org/abs/2005.00096>.
- Hassan, A., Shahin, I. and Alsabek, M. (2020) 'COVID-19 detection system using recurrent neural networks', in *Proceedings of the International Communications, Computing, Cybersecurity and Informatics (CCCI)*, p.15.
- Holmberg, E., Hillman, R. and Perkell, J. (1988) 'Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice', *Journal of Acoustical Society of America*, Vol. 84, No. 4, pp.511–529.
- Hossain, M. and Shorfuazzaman, M. (2023) 'Non-invasive COVID-19 screening using deep-learning-based multilevel fusion model with an attention mechanism', *IEEE Open Journal of Instrumentation and Measurement*, Vol. 2, pp.1–12, Article Sequence Number: 2500512.
- Imran, A., Posokhova, I., Qureshi, H., Masood, U., Riaz, M., Ali, K., John, C., Hussain, M. and Nabeel, M. (2020) *AI4COVID-19: AI Enabled Preliminary Diagnosis for COVID-19 from Cough Samples Via An App* [online] <http://arxiv.org/abs/2004.01275> (accessed 6 June 2023).
- Iseli, M. and Alwan, A. (2000) 'Inter- and intra-speaker variability of glottal flow derivative using the LF model', in *Proceedings of ICSLP 2000*, pp.477–480.
- Juvela, L., Bollepalli, B., Tsiaras, V. and Alku, P. (2019) 'Glotnet – a raw waveform model for the glottal excitation in statistical parametric speech synthesis', *IEEE/ACM Transactions on Audio Speech and Language Processing*, Vol. 27, No. 6, pp.1019–1030.
- Laguarta, J., Hueto, F. and Subirana, B. (2020) 'COVID-19 artificial intelligence diagnosis using only cough recordings', *IEEE Open Journal of Engineering in Medicine and Biology*, Vol. 1, pp.275–281.
- Lin, Z., He, Z., Yao, R., Wang, X., Liu, T., Deng, Y. and Xie, S. (2024) 'Deep dual attention network for precise diagnosis of COVID-19 from chest CT images', *IEEE Transactions on Artificial Intelligence*, Vol. 5, No. 1, pp.104–114.
- Narendra, N. and Alku, P. (2020) 'Automatic intelligibility assessment of dysarthric speech using glottal parameters', *Speech Communication*, Vol. 123, No. 1, pp.1–9.
- Ng, M., Lee, E., Yang, J., Yang, F., Li, X., Wang, H., Lui, M., Lo, C., Leung, B. and Khong, P. (2020) 'Imaging prole of the COVID-19 infection: radiologic findings and literature review', *Radiology Cardiothoracic Imaging*, Art. No. e200034, Vol. 2, No. 1.
- Nguyen, T., Nguyen, V. and Tran, M. (2023) 'Collaborative consultation doctors model: unifying CNN and ViT for COVID-19 diagnostic', *IEEE Access*, Vol. 11, pp.95346–95357.
- Ni, Q., Sun, Z., Qi, L., Chen, W., Yang, Y., Wang, L., Zhang, X., Yang, L., Fang, Y., Xing, Z., Zhou, Z., Yu, Y., Lu, G. and Zhang, L. (2020) 'A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images', *European Radiology*, Vol. 30, No. 12, pp.6517–6527.
- Pahar, M., Kloppe, M., Warren, R. and Niesler, T. (2020) *COVID-19 Cough Classification using Machine Learning and Global Smartphone Recordings* [online] <http://arxiv.org/abs/2012.01926> (accessed 6 June 2023).
- Ponomarchuk, A., Burenko, I., Malkin, E., Nazarov, I., Kokh, V., Avetisian, M. and Zhukov, L. (2022) 'Project Aahoo: a practical model and application for COVID-19 detection from recordings of breath, voice and cough', *IEEE Journal of Selected Topics in Signal Processing*, Vol. 16, No. 2, pp.175–187.
- Rieglesberger, E. and Krishnamurthy, A. (1993) 'Glottal source estimation methods of applying the LF model to inverse filtering', in *Proceedings of ICASSP 1993*, Vol. 2, pp.542–545.
- Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S., Nirmala, R., Ghosh, P. and Ganapathy, S. (2020) *Coswara – A Database of Breathing, Cough and Voice Sounds for COVID-19 Diagnosis*, arXiv.

- Shrivastava, A., Deshpande, S., Gidaye, G., Nirmal, J., Ezzine, K., Frikha, M., Desai, K., Shinde, S., Oza, A.D. and Burduhos-Nergis, D.D. (2022) 'Employing energy and statistical features for automatic diagnosis of voice disorders', *Diagnostics*, Vol. 12, No. 11, p.2758.
- Sindhu, I. and Sainin, M. (2024) 'Automatic speech and voice disorder detection using deep learning – a systematic literature review', *IEEE Access*, Vol. 12, pp.49667–49681.
- Strik, H. and Boves, L. (1994) 'Automatic estimation of voice source parameters', in *3rd International Conference on Spoken Language Processing (ICSLP 94)*, pp.18–22.
- Subirana, B., Hueto, F., Rajasekaran, P., Laguarda, J., Puig, S., Malvehy, J., Mitja, O., Trilla, A., Iván Moreno, C., Francisco Muñoz Valle, J., Esther Mercado González, A., Vizmanos, B. and Sarma, S. (2020) *Hi Sigma, Do I Have the Coronavirus?: Call for a New Artificial Intelligence Approach to Support Health Care Professionals Dealing with the COVID-19 Pandemic* [online] <http://arxiv.org/abs/2004.06510> (accessed 12 June 2023).
- Timcke, R., von Leden, H. and Moore, P. (1958) 'Laryngeal vibrations: measurements of the glottic wave: part I. the normal vibratory cycle', *AMA Archives of Otolaryngology*, Vol. 68, No. 1, pp.1–19.
- Trivedy, S., Goyal, M., Mohapatra, P. and Mukherjee, A. (2020) 'Design and development of smartphone-enabled spirometer with a disease classification system using convolutional neural network', *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 9, pp.7125–7135.
- Verde, L., De Pietro, G. and Sannino, G. (2021a) 'Artificial intelligence techniques for the non-invasive detection of COVID-19 through the analysis of voice signals', *Arabian Journal for Science and Engineering*, Vol. 48, pp.11143–11153.
- Verde, L., De Pietro, G., Ghoneim, A.M., Al-Mutib, K.N. and Sannino, G. (2021b) 'Exploring the use of artificial intelligence techniques to detect the presence of coronavirus COVID-19 through speech and voice analysis', *IEEE Access*, Vol. 9, pp.65750–65757.
- Wang, L., Lin, Z. and Wong, A. (2020) 'COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images', *Scientific Reports*, Vol. 10, No. 1, p.12.
- Welch, B. (1947) 'The generalization of 'STUDENT'S' problem when several different population variances are involved', *Biometrika*, Vol. 34, Nos. 1/2, pp.28–35.
- Xu, X. (2020) 'A deep learning system to screen novel coronavirus disease 2019 pneumonia', *Engineering*, Vol. 6, No. 10, pp.1122–1129.
- Zhu, Y., Tiwari, A., Monteiro, J., Kshirsagar, S. and Falk, T. (2023) 'COVID-19 detection via fusion of modulation spectrum and linear prediction speech features', *IEEE/ACM Transactions on Audio, Speech and Language Processing*, Vol. 31, pp.1536–1549.