

COVID-19 activity screening by a smart-data-driven multi-band voice analysis

*Gabriel Silva, *†Patricia Batista, and *Pedro Miguel Rodrigues, *†Porto, Portugal

Summary: COVID-19 is a disease caused by the new coronavirus SARS-COV-2 which can lead to severe respiratory infections. Since its first detection it caused more than six million worldwide deaths. COVID-19 diagnosis non-invasive and low-cost methods with faster and accurate results are still needed for a fast disease control. In this research, 3 different signal analyses have been applied (per broadband, per sub-bands and per broadband & sub-bands) to Cough, Breathing & Speech signals of Coswara dataset to extract non-linear patterns (Energy, Entropies, Correlation Dimension, Detrended Fluctuation Analysis, Lyapunov Exponent & Fractal Dimensions) for feeding a XGBoost classifier to discriminate COVID-19 activity on its different stages. Classification accuracies ranged between 83.33% and 98.46% have been achieved, surpassing the state-of-art methods in some comparisons. It should be emphasized the 98.46% of accuracy reached on pair Healthy Controls vs all COVID-19 stages. The results shows that the method may be adequate for COVID-19 diagnosis screening assistance.

Key Words: COVID-19–Cough–Breathing–Speech signals–Non-linear patterns–Classification.

INTRODUCTION

The COVID-19 pandemic provoked a vast negative impact on health, social and economic systems. Therefore, the development of methodologies to detect the virus has grown exponentially for controlling the pandemic. New diagnostic methods have been developed, such as reverse transcription-polymerase chain reaction (RT-PCR), serologic testing for immunoglobulins, and rapid diagnostic kits.^{1–5} The RT-PCR is the most widely used method, but these tests are time-consuming, expensive, and invasive.^{1,6} So, new diagnostic methods with higher sensibility, non-invasive and low-cost are needed.^{1,5}

Diagnostic methods through voice analysis seems like a solution. Nowadays, voice is considered an important digital biomarker for detection and monitoring the disease progress such as Parkinson's,⁷ Vocal Nodule,⁷ Alzheimer's,⁸ Multiple Sclerosis,⁹ Asthma,¹⁰ Heart disease,¹¹ and recent studies have reported that human audio analysis is found to be useful for COVID-19 detection.^{4,10,12} It is known that Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) affects various parts of the body, with an emphasis on the respiratory system.^{13–15} COVID-19 disease is characterized by infections of the respiratory tract, that affects not only the respiratory system but also the structure that is used for voice production.¹³ So, symptoms such as coughing, altered breathing, voice, and speech have been reported.

Voice is a parameter related with the vocal cords characteristics and speech includes the speech speed and hesitation analyses. The voice production is conditioned by the respiratory process, being the sound generated when air is expelled from the lungs. Then the air passes through the larynx, where the vocal folds are in a vibrating position, occurring the phonation and resonating process.¹³ Although voice modifications have been a less reported parameter, computed tomography scan data and human biopsies have shown that the virus affects the lungs, nasal tract, and vocal folds & tract.¹⁶ So, voice changes in COVID-19 patients can be one of the sensitive symptoms detected at the beginning or in asymptomatic patients' case.¹⁶ Other recent studies based on clinical trials, reported that voice perturbation can be considered as a COVID-19 manifestation.^{14,15,17} Lechien and colleagues (2020, 2022) observed dysphonia and aphonia in COVID-19 patients.^{14,17} However, sometimes, the voices changes are very difficult to be accurately/efficiently distinguished by humans. So, computer technology has shown a potential to bridge this gap by allowing voice quantification.¹⁸ In this way, the use of artificial intelligence through voice analysis has allowed the development of tools for virus diagnosis, prediction, and monitoring (see Table 1).

Other technologies such as the Internet of Things, Big Data, and Blockchain also have been used to predict, detect, and control the virus.²⁵ So, Artificial Intelligence can play a significant role in the development of more effective and reliable, simple, non-invasive, faster, and less expensive diagnostic methods.^{6,20,26} Thus, machine learning trained models and algorithms can be used for voice analysis in prognosis and scanning of SARS-COV-2 infection. So, new real-time algorithms based on voice analysis must be constructed in sense of being more reliably and sensitively to turn possible the distinguish between Healthy and pathological voice.^{20,26} To contribute for this development, the present research was designed to build a new COVID-19 screening activity machine learning tool based on Cough, Breathing & Speech analysis.

Accepted for publication November 7, 2022.

From the *CBQF-Centro de Biotecnologia e Química Fina-Laboratório Associado, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Rua de Diogo Botelho, 1327, 4169-005 Porto, Portugal; and the †HNL/CEDH-Human Neurobehavioural Laboratory/Research Centre for Human Development, Universidade Católica Portuguesa, Rua de Diogo Botelho, 1327, 4169-005 Porto, Portugal.

Address correspondence and reprint requests to Pedro Miguel Rodrigues, Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Office EB112, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal. E-mail: pmrodrigues@ucp.pt

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■–■■■
0892-1997

© 2022 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

<https://doi.org/10.1016/j.jvoice.2022.11.008>

TABLE 1.
Sound-based COVID-19 diagnosis state-of-art methods

Reference	Dataset	Participants	Sound type	Features	Classifier	Accuracy
¹	Coswara	N=1376 participants (127 COVID-patients (CP) and 1249 Healthy Controls (HC))	Cough, Breathing & Speech	Spectral Centroid, Spectral Bandwidth, Spectral Roll-off, Zero Crossing, Mel Frequency Cepstral coefficients (MFCC) & RMS Energy, Skewness, Kurtosis, Coefficient of variation, Standard Error of Mean	Deep Model Shallow classifiers	96.4% CP vs HC
⁴	Coswara	N=1103 participants (84 CP & 1019 HC)	Cough, Breathing & Speech	Geneva Minimalistic Acoustic Parameter Set (GeMaps), extended Geneva Minimalistic Acoustic Parameter Set (eGeMaps), ComParE feature set & Wavelet scattering transform	Multi-layer Perceptron (MLP)	88.52% CP vs HC
⁵	Coswara	N=210 participants (70 CP, 70 recovered negative COVID-19 patients (RP group) & 70 HC)	Cough & Speech	OpenSMILE features & INTER-SPEECH2016 challenge features	Machine learning-based voice assessment (MLVA)	90.07% CP and HC & 92.81% CP vs RP & 92.81% RP vs HC
¹⁵	Not Coswara	N=116 participants (76 Pos-COVID-19 (PosC) patients & 40 HC)	Cough & Speech	Log-mel spectrograms	VGG19 Convolution neural network (CNN)	85% PosC vs HC
¹⁹	Not Coswara	N=54 participants (40 CP and 14 HC)	Cough & Speech	Computational Paralinguistics Challenge features and PRAAT and LIBROSA acoustic features	Deep Convolution Neural Network (dCNN)	83% CP vs HC
²⁰	Coswara	N=1027 participants (77 CP & 950 HC)	Speech	Fundamental Frequency (F0), jitter, shimmer and Harmonic to Noise Ratio (HNR), MFCC, Spectral Centroid and Roll-off	VGG19 CNN	97% CP vs HC
²¹	Not Coswara	N=88 participants (29 positive COVID-19 & 59 HC)	Cough & Speech	80 Mel-scaled frequencies & 80 first derivatives	Support-vector Machines (SVM)	78% CP vs HC
²²	Not Coswara	N=355 participants (62 positive COVID-19 & 293 HC)	Cough & Breathing	Mel spectrograms	ResNet CNN	84.6% CP vs HC
²³	Coswara	N=2883 participants (539 positive COVID-19 & 2344 HC)	Cough	Mel spectrograms, MFCC & clinical features	Ensemble Deep Learning Model	77.1% CP vs HC
²⁴	Coswara	N=3621 participants (2001 positive COVID-19 & 1620 HC)	Cough	Mel spectrograms, MFCC & RMS Energy & clinical features	ResNet-18 & Shallow classifiers	72% CP vs HC

In terms of structure, this paper is organized in six main sections. In [Section 2](#), the database is described. Thereafter, the methodology concerning the signal processing and the classification procedure is explained in [section 3](#). The obtained results are detailed on [Section 4](#) and its inherent discussion were covered in [Section 5](#). Lastly, [Section 6](#) makes remarks about conclusions.

MATERIALS

The Coswara public database has been used on present work (released until 11 April 2022).²⁷ The sound samples are collected via worldwide crowd-sourcing using the internet and also public available on Web.²⁷ Samples are labeled with 6 COVID-19 statuses (No respiratory illness exposed (nRE), Positive - asymptomatic (PA), Positive - moderate (PMo), Positive - mild (PM), Recovered full (RF), and Respiratory illness not identified (RIn)) and 1 group of control (Healthy Controls - HC).

Every participant in the dataset has 9 different sound types with a sampling frequency (FS) of 44100Hz (Breathing-deep, Breathing-shallow, Cough-heavy, Cough-shallow, Counting-fast, Counting-normal, Vowel /a/, Vowel /e/, Vowel /o/), the data is described in detail here.²⁷ After removing items with either missing audio files or demographic data (therefore labels could not be retrieved), a total of 2092 participants were identified in the study, with an age range between 1 and 87 years, and more male participation, 70.4% (see [Table 2](#)).

The participants reported various locations, with the majority being from India (92.3%) and the United States (3.1%), however participants from Germany, Portugal, Finland, Thailand, South Africa, Australia, Oman, Spain, Canada, Switzerland, Sweden, France, United Kingdom,

United Arab Emirates, Netherlands, China, Japan, Hungary, Turkey, Singapore, Saudi Arabia, Bahrain, Philippines, Argentina, Indonesia, Italy, Romania, Bangladesh, Greece, Malaysia, Russia, Brazil, Pakistan, Sri Lanka, Belgium, Korea South, and Syria were also involved in the study.

The data obtained showed that the majority of the participants were Healthy Controls (55.9%), however 24.5% of the participants were infected with the virus, either with different symptoms or without symptoms (3.2%) (check [Table 2](#) for more details).

METHODS

The proposed methodology is divided into three main steps: (1) Preprocessing, (2) Signal Processing and Feature Extraction and (3) Classification. [Figure 1](#) summarizes the methodology implementation steps.

Preprocessing

The duration of each phoneme is always longer than 2 s, so, as recommended in,²⁸ a stable and artifact-free with at least 2.2 s-long signal was guaranteed from each participant 9 different sound recording types.

All signals amplitude was normalized in the range $[-1, 1]$ to prevent the speaker-microphone distance from affecting the model.

At the end, the signals were then split into non-overlapping 20 ms-long segments using the Hamming window.

Signal Processing and feature extraction

This section describes the set of features extracted from each voice type sound recording.

TABLE 2.
Socio-demographic and clinical characteristics

	Continuous Measure	Min	Max	Mean	SD
Age		1	87	35.60	13.87
	Categorical Measure				%
Gender					
Female					29.6
Male					70.4
Country					
India					92.3
USA					3.1
Germany					0.5
Other					4.1
Clinical information					
Healthy Controls					55.9
No respiratory illness exposed					8.8
Positive - asymptomatic					3.2
Positive - mild					15.4
Positive - moderate					5.9
Recovered full					5.0
Respiratory illness not identified					5.8

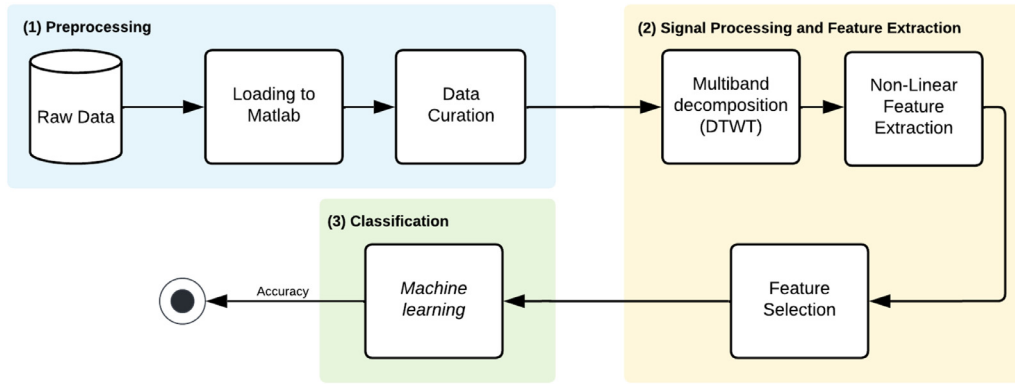


FIGURE 1. Methodology steps.

Multiband Decomposition via Wavelet Transform

The discrete-time wavelet transform (DTWT) of a discrete-time finite-energy signal is its breakdown into a set of basis functions made from a finite number of prototype sequences and their time-shifted versions.²⁹ It is an optimal tool for time-frequency signal analysis because it not only allows us to change the signal domain from time to frequency, and vice versa, but it also allows us to localise the origin of frequency compounds in time.^{30,31}

This structured expansion and its corresponding reconstruction are implemented by means of an octave-band critically decimated filter bank.^{29,32} Considering only the positive frequencies, the m th sub-band is confined to

$$W_k = \begin{cases} [0, \pi/2^S], & m = 0, \\ [\pi/2^{S-m+1}, \pi/2^{S-m}], & m = 1, 2, \dots, S, \end{cases} \quad (1)$$

where S is the number of decomposition stages or levels, $S + 1$ is the number of sub-bands and π is the normalized angular frequency which is equivalent to half the sampling rate.

The DTWT uses an analysis scale function $\sim \phi_1(n)$ and an analysis wavelet function $\sim \psi_1(n)$ defined as

$$\sim \phi_1(n) = h_{LP}(n) \quad (2)$$

and

$$\sim \psi_1(n) = h_{HP}(n), \quad (3)$$

where $h_{LP}(n)$ and $h_{HP}(n)$ are the impulse responses of the half-band low-pass and high-pass analysis filters, respectively.

Defining the following recursion formulas

$$\sim \phi_{i+1}(n) = \sim \phi_i(n/2) * \sim \phi_1(n), \quad (4)$$

$$\sim \psi_{i+1}(n) = \sim \phi_i(n) * \sim \psi_1(n/2^i), \quad (5)$$

where the symbol $*$ denotes the convolution operator, the equivalent analysis filter of the m th sub-band is given by

$$h_m(n) = \begin{cases} \sim \phi_S(n), & m = 0, \\ \sim \psi_{S+1-m}(n), & m = 1, 2, \dots, S. \end{cases} \quad (6)$$

The m th sub-band signal is given by

$$x_m(n) = \begin{cases} \sum_{k=-\infty}^{\infty} x(k)h_m(2^S n - k), & m = 0, \\ \sum_{k=-\infty}^{\infty} x(k)h_m(2^{S-m+1} n - k), & m = 1, 2, \dots, S. \end{cases} \quad (7)$$

In this study, the DTWT was applied to each voice segment for sub-band decomposition up to the third level, i.e., $S=3$. The sub-band signals $x_m(n)$, $m = \{0, 1, 2, 3\}$, were resampled to the original sampling rate using the wavelet interpolation method.³³ The Wavelet decomposition tree is shown on Figure 2.

Non-linear Analysis

Non-linear dynamic features characterize a complex system, according to non-linear dynamic theory. Non-linearity features are generally associated with entropy, exponents, and fractal dimensions parameters.

Energy. Energy is one of the most informative concepts in information theory, that has the ability to exploit multiple meaningful features from a non-stationary signal and it can be seen as the system's ability to accomplish work. In the case of a speech signal, such as the one used in this work, the energy describes the effort done by the speaker's lungs and vocal tract to make sound as a function of time.^{34,35} The energy of $x(n)$ is defined as

$$EN = \sum_{n=1}^N |x(n)|^2 \quad (8)$$

Entropy. Entropy is a measure that considers the amount of energy present in a complex system. Entropy features may also be used to quantify the information content that has been masked in a signal, modelling the unpredictability and irregularities of a pathological speech signal, as accurate as possible, within a certain signal's band.³⁵ The Shannon (*SE*), Logarithmic (*LE*) and Approximate (*AE*)

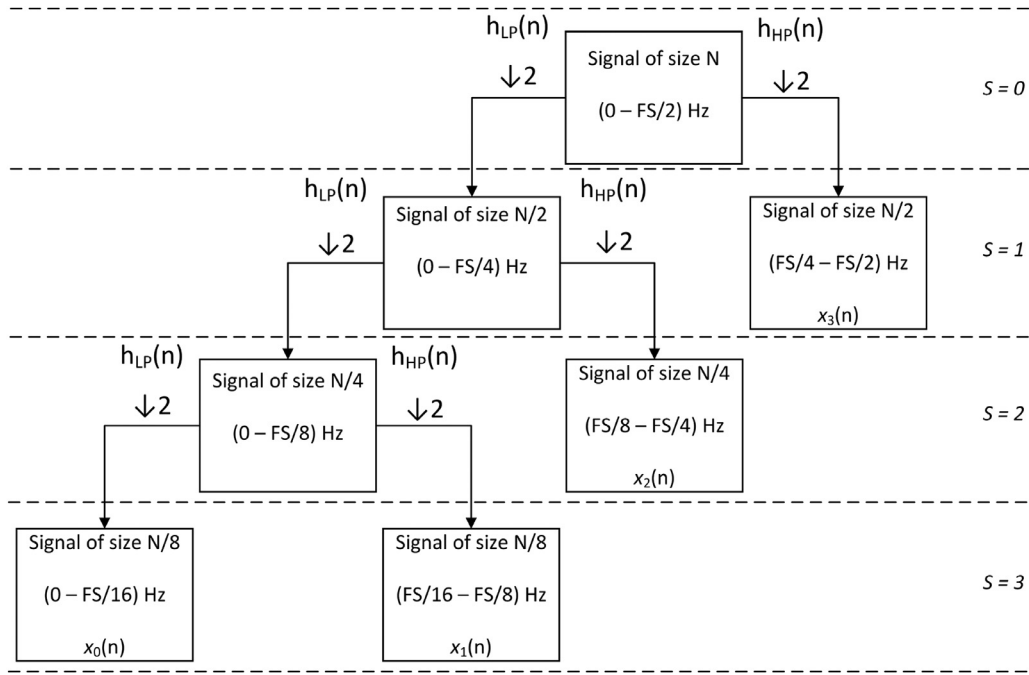


FIGURE 2. The discrete-time wavelet transform decomposition tree

entropies can be estimated as^{36–38}

$$SE = - \sum_{n=1}^N |x(n)|^2 \log[|x(n)|^2], \quad (9)$$

$$LE = \sum_{n=1}^N \log[|x(n)|^2], \quad (10)$$

and

$$AE(m, r, N) = H^m(r) - H^{m+1}(r) \quad (11)$$

where N is the data length (suggested to be 1000 of the signal standard deviation), r is the similar tolerance (between 0.1 and 0.25) and m represents the embedding dimension (between 2 and 3). H is the Heaviside function that results from intermediate calculations.³⁹

Chaos theory. Chaos theory is a concept that is closely related to dynamic systems. Since a dynamic system lacks the properties of an equilibrium system for sure unpredictable disturbances can influence its behavior. As a result of these perturbations, the system travels from one state to another. The concept of phase space refers to the collection of all possible states that a dynamic system can experience over time.

There are two main exponents which provide a comprehensive framework of chaos.^{37,40} At each time instant, the state of a dynamic system characterized by m variables may be represented as a point in m -dimensional space.

The succession of states over time produces arcs called trajectories in this space, which is known as state or phase

space. When these trajectories are tracked over extended periods of time, they can converge to a certain geometric shape, known as an attractor, regardless of the system's original settings.

- Correlation Dimension (D_2): It characterizes the distribution of the attractor points, reflecting the complexity of a dynamic system, and is estimated as

$$D_2 = \lim_{r \rightarrow 0} \frac{\log(C(r, M))}{\log(r)}, \quad (12)$$

where

$$C(r, M) = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (13)$$

is the probability of the pair of points $\{\mathbf{x}_i, \mathbf{x}_j\}$ on the attractor is separated by a distance less than or equal to r and Θ is the Heaviside function.^{39,41}

- Lyapunov Exponent (L): It provides information of trajectories evolution over time,³⁹ reflecting the stability of dynamic systems.

$$LE(x_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \ln |f'(x_k - 1)| \quad (14)$$

where f' is the iterator function f derivative.³⁷

- Detrended Fluctuation Analysis: It is a method that provides a feature for quantifying long-range correlations (self-similarity) of an apparently non-stationary

time series.⁴² From $x(n)$, the cumulative deviation series is calculated as follows

$$y(k) = \sum_{i=1}^k [x(i) - \bar{x}]. \quad (15)$$

Then, for each m -long segment of $y(k)$, a linear approximation denoted by $y_m(k)$ is estimated. The average fluctuation of the signal as a function of m is defined as

$$F(m) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_m(k)]^2}. \quad (16)$$

The slope of the best linear approximation of $\log[F(m)]$ as a function of $\log(m)$ is the scale exponent α that represents the correlation properties of the signal $x(n)$.⁴³

Fractal Dimension. Self-similarity features of fractal structures allow for the description of both irregular processes and structures. Fractal geometry has been shown to be a beneficial strategy for identifying irregularities produced by various disorders in biomedical signals such as the voice.⁴⁴ In fact, monitoring the self-similarity of the speech enables for the clinical status of the patient to be assessed. Fractal dimension (FD) feature extraction such as the Higuchi fractal dimension is considered to be helpful in voice signal classification.⁴⁰

- Higuchi Exponent (D):

$$L(k)\alpha k^{-D} \quad (17)$$

where k indicates the time interval, $L(k)$ is the length of the curve in the k time interval and D is the Higuchi Exponent.⁴⁵

- Katz Algorithm: According to Katz (1988),⁴⁶ the FD of a waveform $x(n)$ can be defined as

$$FD_k = \frac{\log(L/a)}{\log(d/a)} \quad (18)$$

where L is the sum of the distances between the successive points of $x(n)$, a is the average distance between the successive points, and d is the greatest distance between $x(1)$ and the remaining points of $x(n)$.

Feature extraction process

For each study participant voice recording, 9 non-linear features (EN , SE , LE , AE , CD , L , α , D , FD_k) have been extracted from each 20 ms-long segments and their respective sub-bands computed by DTWT decomposition until level 3 using biorthogonal 3.1. As we are trying to find pathological patterns, most of times characterized as being time domain small signal oscillations imperceptible to human

eye, the small size of Biorthogonal 3.1 filter ($N = 4$) helps to keep the quality of signal's information in time domain for each signal's sub-band, avoiding the lose of those pathological patterns along the DTWT analysis.⁴⁷ This Wavelet also proves to be a good choice for speech signal analysis in previous works.^{7,48} Treating the 9 extracted features of all segments as time series distributions, mean statistic was used to compress them along time per sub-band and per broadband, reducing the problem dimensionality and, at the same time, ensuring that the number of metrics per subjects was equal for the classification task.

Feature selection and classification procedure

The study goal is to infer about the capacity of a XGBoost ensemble of machine learning model for evaluating the COVID-19 activity evolution along its different stages by voice analysis. The XGBoost classifier,⁴⁹ that is a tree-based ensemble machine learning algorithm, has been used for the propose with the following optimized parameters: boosted trees to fit = 150, learning rate = 0.1, max. depth of the tree = 6, L2 regularization term = 1, and L1 regularization term = 0.

3 different modalities of binary comparison analyses have been studied: (1) per broadband (1 signal analysis resulting in a feature vector of 81 features per subject – 9 features \times 9 voice recordings); (2) per sub-bands (4 signal analyses resulting in a feature vector with 324 features per subject – 9 features \times 4 signal analysis \times 9 voice recordings); and (3) per broadband and sub-bands (5 signal analyses resulting in a feature vector with 405 features per subject – 9 features \times 5 signal analysis \times 9 voice recordings). In each 22 binary comparisons, it was ensured that all datasets have been equally balanced for discrimination, for that, subgroups have been randomly selected from the main database previously defined on Table 2 taking in to account the lower maximum number of samples found individually in each group that form the study group pair. Thus, the group that has less number of samples in pair indicate the number of samples in both groups, respectively. Per analysis modality and binary comparison, the data have been normalized by z-score algorithm⁵⁰ and 5 different sets of features were selected by f-score algorithm,⁵¹ representing 5%, 10%, 20%, 50% and 100% of total of features, have been presented to the entries of the XGBoost ensemble machine learning model for discrimination. In all cases, to verify the generalization capacity of the classifier, a leave-one-out cross-validation procedure is used.

For a better understanding of the whole methodology Figure 3 presents a summary of the procedure.

RESULTS

Classification accuracies for each pair of study groups, according to the procedure described in subsection 3.3, are presented in Table 3.

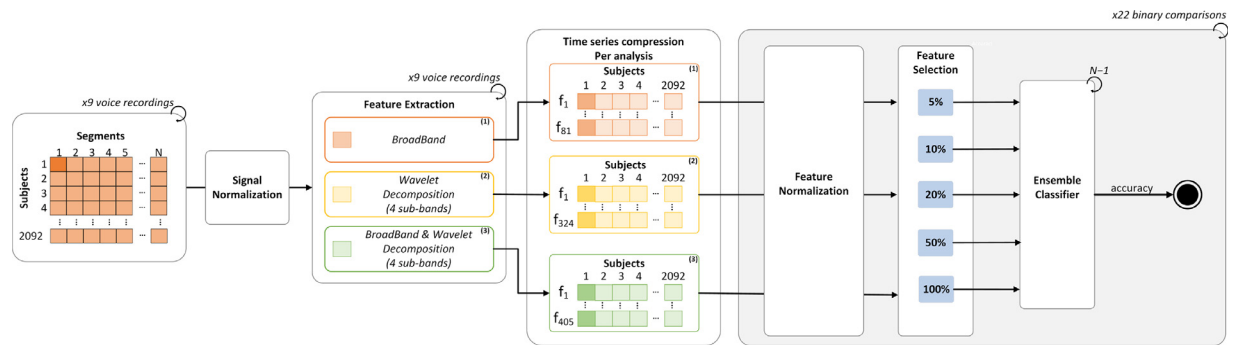


FIGURE 3. Methodology workflow.

DISCUSSION

By analysing Table 3, some observations can be drawn about classification results between pairs of study groups. Accuracies higher than 83% were obtained for all pairs. Analysis 1 provided the best features set in 2 of 22 (9%) study pairs classifications, features from Analysis 2 have been used in 8 of 22 (36.5%) study pairs classifications and Analysis 3 brings the best features for 12 of 22 (54.5%) study pairs classifications. All comparisons that involved the Recovery full group reached accuracies lower than 90%, meaning that even when the patients no longer carry the virus, it still has some impact on patients' health, the so called in the literature Long COVID-19 effect.⁵² Besides

that, all other comparisons reached higher accuracy rates than 90% except the pair Positive - mild vs Positive - moderate. The similarity between their symptoms and effects difficult the correct identification of those stages⁵³ and that reason can explain the obtained results.

Table 3 analysis also demonstrates the significance of Wavelet Transform decomposition, as 91% of the classifications with greater accuracy came either from analyses 2 or 3, in which wavelet information is employed entirely or in combination with metrics extracted from broadband, respectively. This finding helps us to argue the relevance of utilizing the Wavelet Transform for a signal multi-band analysis to extract metrics that can discriminate

TABLE 3.
Classification accuracy results for each pair of study groups.

Study pair comparison	# samples per group	Best analysis	% of features	Accuracy
Healthy Controls vs No respiratory illness exposed	181	2	10	96.69%
Healthy Controls vs Positive - asymptomatic	65	3	20	96.92%
Healthy Controls vs Positive - mild	318	2	10	98.11%
Healthy Controls vs Positive - moderate	120	3	20	97.92%
Healthy Controls vs Recovered full	105	2	20	88.57%
Healthy Controls vs Respiratory illness not identified	120	2	20	97.92%
No respiratory illness exposed vs Positive - asymptomatic	65	3	20	97.69%
No respiratory illness exposed vs Positive - mild	181	1	10	98.90%
No respiratory illness exposed vs Positive - moderate	120	3	10	97.50%
No respiratory illness exposed vs Recovered full	105	2	20	88.10%
No respiratory illness exposed vs Respiratory illness not identified	120	3	10	96.67%
Positive - asymptomatic vs Positive - mild	65	3	20	98.46%
Positive - asymptomatic vs Positive - moderate	65	3	20	97.69%
Positive - asymptomatic vs Recovered full	65	3	20	86.15%
Positive - asymptomatic vs Respiratory illness not identified	65	3	20	96.92%
Positive - mild vs Positive - moderate	120	3	10	86.67%
Positive - mild vs Recovered full	105	2	20	87.62%
Positive - mild vs Respiratory illness not identified	120	2	10	96.67%
Positive - moderate vs Recovered full	105	1	20	83.33%
Positive - moderate vs Respiratory illness not identified	120	3	10	97.92%
Recovered full vs Respiratory illness not identified	105	2	20	97.62%
Healthy Controls vs COVID-19 all stages	909	3	50	98.46%

between several binary groups with higher efficiency and accuracy.

Regarding state-of-art methods showed on Table 1, it can be observed that most of the authors have focused mainly their research on Healthy Controls vs COVID-19 all stages. Particularly on that comparison, our algorithm outperforms the best reported accuracy on state-of-art methods that used the same dataset by 1%,²⁰ showing in this way to be a good candidate for this kind of discrimination. As for the studies that didn't use the Coswara dataset, although comparisons must be done with some caution as different databases have been employed, our method outperformed the best result from those studies by 13%.¹⁵ Regarding Positive - mild vs Recovered full our algorithm provided a slightly higher accuracy than the achieved in Suppakitjanusan *et. al.* study,¹⁵ 2% higher. About the other remaining 20 pairs, it can not be done a direct comparison as far as we know those discrimination's have been not reported on state-of-art works.

CONCLUSIONS

This study investigated the detection of COVID-19 by using multiband non-linear parameters of Cough, Breathing & Speech. The detection was performed between pairs of study groups using 3 different modalities of signal analyses: (1) per broadband; (2) per sub-bands; and (3) per broadband and sub-bands. For each pair of study groups, a feature selection was carried out by f-score algorithm with different combinations, 5%, 10%, 20%, 50%, and 100% of the total features, and they were used as input for the XGBoost ensemble machine learning model for discrimination within a leave-one-out cross-validation procedure. The following classification accuracies have been obtained: 96.69% HC vs nRE, 98.11% HC vs PM, 97.92% HC vs RIn, 88.57% HC vs RF, 96.92% HC vs PA, 97.92% HC vs PMo, 97.50% nRE vs PM, 96.67% nRE vs RIn, 88.10% nRE vs RF, 97.69% nRE vs PA, 97.50% nRE vs PMo, 97.92% PM vs RIn, 83.33% PM vs RF, 98.46% PM vs PA, 86.67% PM vs PMo, 97.62% RIn vs RF, 96.62% RIn vs PA, 97.92% RIn vs PMo, 86.15% RF vs PA, 97.69% RF vs PMo, 97.69% PA vs PMo and 98.46% COVID-19 patients vs HC.

The results demonstrated that the combination of cough, breathing & speech signal multi-band analysis is adequate to detect COVID-19 activity. As voice recording systems are relatively inexpensive, non-invasive, mobile, and fast, this kind of solution can be widely spread and help in COVID-19 diagnosis screening at clinics and hospitals. As far as we could check this is the first attempt to distinguish between different COVID-19 stages by voice analysis. In future works, the results must be updated with a larger population to ensure generalization, an analysis of the most suitable wavelet family for each study group pairs may be performed, and it should be consider new methods for feature selection based on paraconsistency and paraconsistent

feature engineering that have shown to be good candidates for the propose in other works.^{54,55}

Acknowledgments

This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, through the project UIDB/50016/2020.

REFERENCES

1. Aly M, Rahouma KH, Ramzy SM. Pay attention to the speech: COVID-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings. *Alexandria Engineering Journal*. 2022;61(5):3487–3500. <https://doi.org/10.1016/j.aej.2021.08.070>.
2. Bastos ML, Tavaziva G, Abidi SK, Campbell JR, Haraoui L-P, Johnston JC, et al. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *BMJ*. 2020;m2516. <https://doi.org/10.1136/bmj.m2516>.
3. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*. 2020;25(3). <https://doi.org/10.2807/1560-7917.es.2020.25.3.2000045>.
4. Despotovic V, Ismael M, Cornil M, Call RM, Fagherazzi G. Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results. *Computers in Biology and Medicine*. 2021;138:104944. <https://doi.org/10.1016/j.compbiomed.2021.104944>.
5. Robotti C, Costantini G, Saggio G, Cesarini V, Calastri A, Maiorano E, et al. Machine learning-based voice assessment for the detection of positive and recovered COVID-19 patients. *Journal of Voice*. 2021. <https://doi.org/10.1016/j.jvoice.2021.11.004>.
6. Heidari A, Navimipour NJ, Unal M, Toumaj S. The COVID-19 epidemic analysis and diagnosis using deep learning: A systematic literature review and future directions. *Computers in Biology and Medicine*. 2022;141:105141. <https://doi.org/10.1016/j.compbiomed.2021.105141>.
7. Alves M, Silva G, Bispo BC, Dajer ME, Rodrigues PM. Voice disorders detection through multiband cepstral features of sustained vowel. *Journal of Voice*. 2021. <https://doi.org/10.1016/j.jvoice.2021.01.018>.
8. Toth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z, et al. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research*. 2018;15(2):130–138. <https://doi.org/10.2174/1567205014666171121114930>.
9. Rusz J, Benova B, Ruzickova H, Novotny M, Tykalova T, Hlavnicka J, et al. Characteristics of motor speech phenotypes in multiple sclerosis. *Multiple Sclerosis and Related Disorders*. 2018;19:62–69. <https://doi.org/10.1016/j.msard.2017.11.007>.
10. Popadina AO, Salah A-M, Jalal K. Voice analysis framework for asthma-COVID-19 early diagnosis and prediction: AI-based mobile cloud computing application. *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*. IEEE; 2021:1803–1807. <https://doi.org/10.1109/elconrus51938.2021.9396367>.
11. Usman M, Zubair M, Ahmad Z, Zaidi M, Ijyas T, Parayangat M, et al. Heart rate detection and classification from speech spectral features using machine learning. *Archives of Acoustics*. 2021;46:41–53. <https://doi.org/10.24425/AOA.2021.136559>.
12. Lella KK, Pja A. Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath. *Alexandria Engineering Journal*. 2022;61(2):1319–1334. <https://doi.org/10.1016/j.aej.2021.06.024>.
13. Asiaee M, Vahedian-azimi A, Atashi SS, Keramatfar A, Nourbakhsh M. Voice quality evaluation in patients with COVID-19: An acoustic analysis. *Journal of Voice*. 2020. <https://doi.org/10.1016/j.jvoice.2020.09.024>.

14. Lechien JR, Chiesa-Estomba CM, Cabaraux P, Mat Q, Huet K, Harnegnies B, et al. Features of mild-to-moderate COVID-19 patients with dysphonia. *Journal of Voice*. 2022;36(2):249–255. <https://doi.org/10.1016/j.jvoice.2020.05.012>.
15. Suppakitjanusant P, Sungkanuparph S, Wongsinin T, Virapongsiri S, Kasemkosin N, Chailurkit L, et al. Identifying individuals with recent COVID-19 through voice classification using deep learning. *Scientific Reports*. 2021;11(1). <https://doi.org/10.1038/s41598-021-98742-x>.
16. Sondhi S, Salhan A, Santoso CA, Doucoure M, Dharmawan DM, Sureka A, et al. Voice processing for COVID-19 scanning and prognostic indicator. *Heliyon*. 2021;7(10):e08134. <https://doi.org/10.1016/j.heliyon.2021.e08134>.
17. Lechien JR, Chiesa-Estomba CM, Place S, Laethem YV, Cabaraux P, Mat Q, et al. Clinical and epidemiological characteristics of 1420 european patients with mild-to-moderate coronavirus disease 2019. *Journal of Internal Medicine*. 2020;288(3):335–344. <https://doi.org/10.1111/joim.13089>.
18. Shin D, Cho WI, Park CHK, Rhee SJ, Kim MJ, Lee H, et al. Detection of minor and major depression through voice as a biomarker using machine learning. *Journal of Clinical Medicine*. 2021;10(14):3046. <https://doi.org/10.3390/jcm10143046>.
19. Shimon C, Shafat G, Dangoor I, Ben-Shitrit A. Artificial intelligence enabled preliminary diagnosis for COVID-19 from voice cues and questionnaires. *The Journal of the Acoustical Society of America*. 2021;149(2):1120–1124. <https://doi.org/10.1121/10.0003434>.
20. Verde L, De Pietro G, Ghoneim A, Alrashoud M, Al-Mutib KN, Sannino G. Exploring the use of artificial intelligence techniques to detect the presence of coronavirus covid-19 through speech and voice analysis. *IEEE Access*. 2021;9:65750–65757. <https://doi.org/10.1109/access.2021.3075571>.
21. Pinkas G, Karny Y, Malachi A, Barkai G, Bachar G, Aharonson V. SARS-CoV-2 detection from voice. *IEEE Open Journal of Engineering in Medicine and Biology*. 2020;1:268–274. <https://doi.org/10.1109/ojemb.2020.3026468>.
22. Coppock H, Gaskell A, Tzirakis P, Baird A, Jones L, Schuller BW. End-2-end COVID-19 detection from breath & cough audio. 2021. <https://doi.org/10.48550/ARXIV.2102.08359>.
23. Chaudhari G, Jiang X, Fakhry A, Han A, Xiao J, Shen S, et al. Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. 2020. <https://doi.org/10.48550/ARXIV.2011.13320>.
24. Bagad P., Dalmia A., Doshi J., Nagrani A., Bhamare P., Mahale A., et al. Cough against COVID: Evidence of COVID-19 signature in cough sounds. 2020. <https://doi.org/10.48550/ARXIV.2009.08790>.
25. Shahid O, Nasajpour M, Pouriye S, Parizi RM, Han M, Valero M, et al. Machine learning research towards combating COVID-19: Virus detection, spread prevention, and medical assistance. *Journal of Biomedical Informatics*. 2021;117:103751. <https://doi.org/10.1016/j.jbi.2021.103751>.
26. Anand S, Sharma V, Pourush R, Jaiswal S. A comprehensive survey on the biomedical signal processing methods for the detection of COVID-19. *Annals of Medicine and Surgery*. 2022;76:103519. <https://doi.org/10.1016/j.amsu.2022.103519>.
27. Sharma N, Krishnan P, Kumar R, Ramoji S, Chetupalli SR, Nirmala R, et al. Coswara — a database of breathing, cough, and voice sounds for COVID-19 diagnosis. *Interspeech 2020*. ISCA; 2020:4811–4815. <https://doi.org/10.21437/interspeech.2020-2768>.
28. Coleman RF. Sources of variation in phonetograms. *Journal of Voice*. 1993;7(1):1–14. [https://doi.org/10.1016/S0892-1997\(05\)80107-9](https://doi.org/10.1016/S0892-1997(05)80107-9).
29. Kovačević MV, Jelena. *Wavelets and Subband Coding*. Englewood Cliffs, New Jersey: Prentice Hall; 1995.
30. Guido RC, Pedroso F, Furlan A, Contreras RC, Caobianco LG, Neto JS. CWT × DWT × DTWT × SDTWT: Clarifying terminologies and roles of different types of wavelet transforms. *International Journal of Wavelets, Multiresolution and Information Processing*. 2020;18(06):2030001. <https://doi.org/10.1142/s0219691320300017>.
31. Guido RC. Wavelets behind the scenes: Practical aspects, insights, and perspectives. *Physics Reports*. 2022;985:1–23. <https://doi.org/10.1016/j.physrep.2022.08.001>.
32. Malvar HS. *Signal Processing with Lapped Transforms*. Norwood, Massachusetts: Artech House; 1992.
33. Rioul O, Vetterli M. Wavelets and signal processing. *IEEE Signal Processing Magazine*. 1991;8(4):14–38. <https://doi.org/10.1109/79.91217>.
34. Guido RC. A tutorial on signal energy and its applications. *Neurocomputing*. 2016;179:264–282. <https://doi.org/10.1016/j.neucom.2015.12.012>.
35. Akbari A, Arjmandi MK. An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. *Biomedical Signal Processing and Control*. 2014;10:209–223. <https://doi.org/10.1016/j.bspc.2013.11.002>.
36. Hu L, Zhang Z. *EEG Signal Processing and Feature Extraction*. Springer Singapore; 2019. <https://doi.org/10.1007/978-981-13-9113-2>.
37. Sanei S, Chambers JA. *EEG Signal Processing*. John Wiley & Sons Ltd.; 2007. <https://doi.org/10.1002/9780470511923>.
38. Sairamy NJ, Subathra MSP, Suvisheshamuthu ES, George ST. A new approach for automatic detection of focal EEG signals using wavelet packet decomposition and quad binary pattern method. *Biomedical Signal Processing and Control*. 2021;63:102096. <https://doi.org/10.1016/j.bspc.2020.102096>.
39. Faust O, Bairy MG. Nonlinear analysis of physiological signals: A review. *Journal of Mechanics in Medicine and Biology*. 2012;12(4). <https://doi.org/10.1142/S0219519412400155>.
40. Albertovich TD, Aleksandrovna RI. The fractal analysis of the images and signals in medical diagnostics. *Fractal Analysis - Applications in Health Sciences and Social Sciences*. InTech; 2017:57–87. <https://doi.org/10.5772/intechopen.68167>.
41. Bermúdez GR, Laencina PJG. Analysis of EEG signals using nonlinear dynamics and chaos - a review. *Applied Mathematics & Information Sciences*. 2015;9:1–13. <https://doi.org/10.12785/amis/090512>.
42. Gifani P, Rabiee HR, Hashemi MH, Taslimi P, Ghanbari M. Optimal fractal-scaling analysis of human EEG dynamic for depth of anesthesia quantification. *Journal of the Franklin Institute*. 2007;344(3-4):212–229.
43. Lee J-M, Kim D-J, Kim I-Y, Park K-S, Kim SI. Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data. *Computers in biology and medicine*. 2002;32(1):37–47.
44. Ali Z, Elamvazuthi I, Alsulaiman M, Muhammad G. Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals. *Journal of Medical Systems*. 2015;40(1). <https://doi.org/10.1007/s10916-015-0392-2>.
45. Wijayanto I, Hartanto R, Nugroho HA. Higuchi and katz fractal dimension for detecting interictal and ictal state in electroencephalogram signal. *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE; 2019:1–6. <https://doi.org/10.1109/icitee.2019.8929940>.
46. Katz MJ. Fractals and the analysis of waveforms. *Computers in Biology and Medicine*. 1988;18(3):145–156. [https://doi.org/10.1016/0010-4825\(88\)90041-8](https://doi.org/10.1016/0010-4825(88)90041-8).
47. Guido RC. Effectively interpreting discrete wavelet transformed signals [lecture notes]. *IEEE Signal Processing Magazine*. 2017;34(3):89–100. <https://doi.org/10.1109/msp.2017.2672759>.
48. Alves M, Silva G, Bispo B, Dajer M, Rodrigues P. Detecção de distúrbios vocais através de parâmetros cepstrais multibanda de vogal sustentada. *Anais de XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. Sociedade Brasileira de Telecomunicações; 2020:1–5. <https://doi.org/10.14209/sbrt.2020.1570641723>.
49. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery; 2016:9781450342322785–794. <https://doi.org/10.1145/2939672.2939785>.

50. Rodrigues PM, Bispo BC, Garrett C, Alves D, Teixeira JP, Freitas D. Lacsogram: A new EEG tool to diagnose alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*. 2021;25(9):3384–3395. <https://doi.org/10.1109/jbhi.2021.3069789>.
51. Sevani N, Hermawan I, Jatmiko W. Feature selection based on f-score for enhancing CTG data classification. *2019 IEEE International Conference on Cybernetics and Computational Intelligence (Cybernetics-Com)*. IEEE; 2019:18–22. <https://doi.org/10.1109/cyberneticscom.2019.8875656>.
52. Subramanian A, Nirantharakumar K, Hughes S, Myles P, Williams T, Gokhale KM, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nature Medicine*. 2022. <https://doi.org/10.1038/s41591-022-01909-w>.
53. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*. 2020;19(3):141–154. <https://doi.org/10.1038/s41579-020-00459-7>.
54. Guido RC. Paraconsistent feature engineering [lecture notes]. *IEEE Signal Processing Magazine*. 2019;36(1):154–158. <https://doi.org/10.1109/msp.2018.2874549>.
55. Marcondes FS, Durães D, Santos F, Almeida JJa, Novais P. Neural network explainable AI based on paraconsistent analysis: An extension. *Electronics*. 2021;10(21):2660. <https://doi.org/10.3390/electronics10212660>.