

## Article

# Acoustic and Clinical Data Analysis of Vocal Recordings: Pandemic Insights and Lessons

Pedro Carreiro-Martins <sup>1,2,\*</sup>, Paulo Paixão <sup>1</sup>, Iolanda Caires <sup>1</sup>, Pedro Matias <sup>3</sup>, Hugo Gamboa <sup>3,4</sup>, Filipe Soares <sup>3</sup>, Pedro Gomez <sup>5</sup>, Joana Sousa <sup>6</sup> and Nuno Neuparth <sup>1,2</sup>

- <sup>1</sup> Comprehensive Health Research Center (CHRC), LA-REAL, NOVA Medical School, Campo Mártires da Pátria, 130, 1169-056 Lisboa, Portugal; nneuparth@gmail.com (N.N.)
- <sup>2</sup> Serviço de Imunoalergologia, Hospital de Dona Estefânia, ULS São José, Rua Jacinta Marto, 1169-045 Lisbon, Portugal
- <sup>3</sup> Fraunhofer Portugal AICOS, Rua Alfredo Allen 455/461, 4200-135 Porto, Portugal; filipe.soares@aicos.fraunhofer.pt (F.S.)
- <sup>4</sup> Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys), Faculdade de Ciências e Tecnologia, NOVA University of Lisbon, Caparica, 2820-001 Lisbon, Portugal
- <sup>5</sup> NeuSpeLab, CTB, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28223 Madrid, Spain; pedrogvilda@telefonica.net
- <sup>6</sup> NOS Inovação, Rua Actor António Silva, 9–6° Piso, Campo Grande, 1600-404 Lisboa, Portugal
- \* Correspondence: pmartinsalergo@gmail.com

**Abstract: Background/Objectives:** The interest in processing human speech and other human-generated audio signals as a diagnostic tool has increased due to the COVID-19 pandemic. The project OSCAR (vOice Screening of Corona viRus) aimed to develop an algorithm to screen for COVID-19 using a dataset of Portuguese participants with voice recordings and clinical data. **Methods:** This cross-sectional study aimed to characterise the pattern of sounds produced by the vocal apparatus in patients with SARS-CoV-2 infection documented by a positive RT-PCR test, and to develop and validate a screening algorithm. In Phase II, the algorithm developed in Phase I was tested in a real-world setting. **Results:** In Phase I, after filtering, the training group consisted of 166 subjects who were effectively available to train the classification model (34.3% SARS-CoV-2 positive/65.7% SARS-CoV-2 negative). Phase II enrolled 58 participants (69.0% SARS-CoV-2 positive/31.0% SARS-CoV-2 negative). The final model achieved a sensitivity of 85%, a specificity of 88.9%, and an F1-score of 84.7%, suggesting voice screening algorithms as an attractive strategy for COVID-19 diagnosis. **Conclusions:** Our findings highlight the potential of a voice-based detection strategy as an alternative method for respiratory tract screening.

**Keywords:** diagnostic tests; machine learning; SARS-CoV-2; speech; voice



**Citation:** Carreiro-Martins, P.; Paixão, P.; Caires, I.; Matias, P.; Gamboa, H.; Soares, F.; Gomez, P.; Sousa, J.; Neuparth, N. Acoustic and Clinical Data Analysis of Vocal Recordings: Pandemic Insights and Lessons.

*Diagnostics* **2024**, *14*, 2273. <https://doi.org/10.3390/diagnostics14202273>

Academic Editor: Juan Rafael Orozco-Arroyave

Received: 16 August 2024

Revised: 3 October 2024

Accepted: 9 October 2024

Published: 12 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Diseases of the respiratory tract frequently affect the vocal apparatus, which leads to changes in the timbre of the voice.

Analysing the voice and other audio signals is an attractive tool for screening diseases, as sound recordings are easy to obtain and non-invasive [1–3]. However, due to the subtle changes in voice and cough characteristics, artificial intelligence techniques are required to recognise specific disease patterns [4]. Especially for biomarkers derived from audio recordings, these patterns may present geographical, linguistic, gender, and age-related differences, which makes this problem more challenging than it could be at first glance.

Interest in this topic has also increased as a result of the COVID-19 pandemic [5–7], during which, some research groups have tried to develop more convenient and accessible diagnostic methods by using machine learning (ML) and deep learning (DL) to analyse voice recordings and other audio signals such as coughing, breathing sounds, and respiratory rate [5,6].

However, most algorithms have used crowdsourced databases, which do not guarantee the quality of the recordings and only contain a limited number of positive samples. The most commonly used and well-known respiratory audio-based datasets include the Coswara dataset [8], created as part of the Coswara Project developed by the Indian Institute of Science Bangalore, and the COVID-19 Sounds database [9], the largest multimodal dataset of respiratory sounds. These two datasets were made available (the latter upon request) to the research community.

Research groups have continuously explored ML and DL methodologies for COVID-19 detection using cough, breathing, and speech recordings to capture changes in the vocal tract of infected subjects. Brown et al. [10] proposed a machine learning (ML) pipeline that extracts features from audio recordings of coughing and breathing (from the COVID-19 Sounds database) and feeds them into a statistical model to distinguish between cases with COVID-19 and without COVID-19. Their best results achieved a specificity of 80% and a sensitivity of 72%. Anupam et al. [11] implemented their COVID-19 screening framework by analysing cough recordings from a subset of the Coswara database. The processing steps included audio cleaning, feature extraction, and classification. The best result was obtained with a Support Vector Machine (SVM) model, which had a sensitivity of 97% and a specificity of 98%.

Coppock et al. [12] developed a custom convolutional neural network (CNN) to analyse spectrograms of cough and breathing audios from COVID-19 Sounds, achieving 90.9% area under the receiver operating characteristic curve (AUROC) and 77.4% sensitivity in discriminating COVID-19 positive cases from those with asthma. A similar study was conducted by Pahar et al. [13] that compared classical ML and deep neural network (DNN) models using raw cough audio features from Coswara and Sarcos databases, with a ResNet50 model achieving 98% and 57% specificity and 93% sensitivity, respectively. Melek et al. [14] studied the detection of COVID-19 from single cough sounds using Virufy and NoCoCoDa datasets, finding a K-nearest neighbors model yielding around 94% accuracy. Meister et al. [15] conducted an extensive feature ranking across Coswara and COVID-19 Sounds audio data, highlighting the importance of spectral and cepstral features, with a Random Forest model achieving around 87% AUROC.

Nevertheless, the validity of these previous results is compromised by the unreliable data sources containing either unconfirmed COVID-19 infection labels or uncontrolled recording frameworks. Therefore, some researchers have taken it upon themselves to create a recording protocol that ensures the reliability of their results [16]. For instance, using an RBF-SVM model, Robotti et al. [17] employed multiple vocal data sources (including vowel phonation, speech, and cough) to execute a series of feature extraction, selection, and classification steps. The authors achieved an average accuracy of 80% using data recorded under controlled settings. Similarly, Costantini et al. [18] applied two custom ML (AdaBoost classifier) and DL (CNNs) algorithms to identify COVID-19 infection using a pool of 310 recruited individuals. The authors tried to distinguish between COVID-19 infected/recovered patients and healthy controls in two separate classification tasks, having obtained an accuracy of 100% and 86.1%, respectively. An identical data collection study by Hassan et al. [19] used a DNN model to analyse cough, breathing, and voice samples, achieving 98% precision and 100% sensitivity only using breathing recordings. Modi et al. [20] also carried out a data collection procedure, recording vocal samples from 117 COVID-19 infected patients and healthy controls. Short-time Fourier representations were used as features to feed a logistic regression algorithm, resulting in an average file-level accuracy of 74.80%. This helps support the idea that high-quality datasets (obtained with validated labels and in controlled settings) lead to better outcomes, either in terms of performance results or even the meaningfulness of insights captured from data. Nonetheless, most studies still lack a robust assessment of the validity and quality of the audios recorded before undergoing further processing stages, in order to increase confidence in the obtained outcomes.

Unlike the studies described in this section, our proposed framework relies on a more robust data curation pipeline executed before extracting features and training ML models. This involves: (i) recording high-quality uncompressed audio, (ii) discarding silent portions of each audio, and (iii) validating the quality of the audios and their relevant segments using pre-trained DL models.

This paper presents the dataset and the results of the OSCAR (vOice Screening of CoronA viRus) project. This study aimed to develop an algorithm for early screening of COVID-19 and test it in a real-world subject scenario. The OSCAR dataset is the first audio dataset collected in Portugal, combining (a) audio recording exercises, (b) self-reported data, and (c) validated labels of SARS-CoV-2 infection, providing a robust resource for testing screening algorithms of COVID-19 or other respiratory diseases. By combining a carefully designed audio processing pipeline, meaningful features, and a reliable source of data containing validated SARS-CoV-2 infection labels, we believe the presented outcomes may support future validation trials and foster the integration of digital screening tools into clinical practice.

## 2. Methods

### 2.1. Study Design, Setting, and Participants

The OSCAR project was a cross-sectional observational study in Portugal divided into two phases.

Phase I aimed to characterise the pattern of sounds produced by the vocal apparatus in patients with SARS-CoV-2 infection documented by a positive RT-PCR test to develop and validate an algorithm. Participant recruitment took place from 1 June 2021 to 8 April 2022 at four centres: Centro de Medicina Laboratorial Germano de Sousa, Hospital da Luz—Lisbon, Hospital Beatriz Ângelo, and NOVA Medical School. A team member from each recruitment centre extended the invitation to participate. For this purpose, a leaflet with a Q.R. code and a unique participant code was distributed.

To be eligible, participants had to be  $\geq 18$  years of age, be able to read, understand, and speak Portuguese, have a nasopharyngeal swab taken to identify SARS-CoV-2 using RT-PCR (real-time protein chain reaction), have access to a computer/smartphone or tablet, and have an internet connection to answer the online questionnaire and complete the requested audio recording. Exclusion criteria were the inability to understand the consent form, produce vocal sounds, or articulate small sentences in Portuguese.

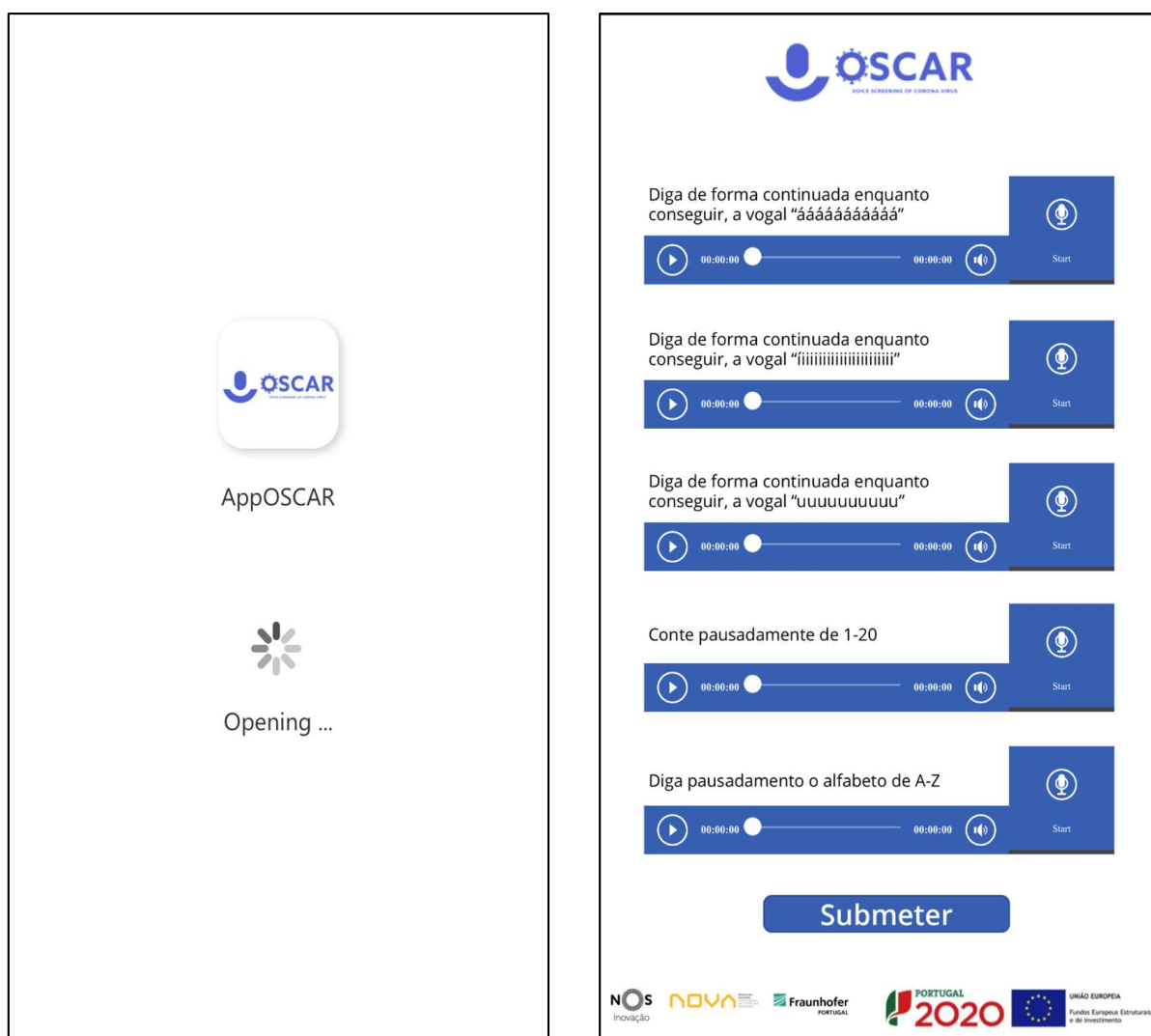
Phase II (from 19 January to 31 May 2022) tested the algorithm developed in Phase I in a real-world setting. In this Phase, the general population could participate, and invitations were made through the project partners' social networks, allowing for a more diverse and broader sample. The inclusion and exclusion criteria were similar to Phase I, but RT-PCR and TRAg (rapid antigen test) were eligible for participation. Convenience sampling was used in Phases I and II.

This study was approved by the Ethics Committee of NOVA Medical School/Faculdade de Ciências Médicas, Universidade Nova de Lisboa (78/2021/CEFCM, 31 May 2021) and by the Ethics Committees of the different recruitment centres. Participants were informed about the study, and their digitally signed consent was obtained according to Portuguese laws and regulations.

### 2.2. Data Collection

Data were collected through an online form created by the study team (see Figure 1). Participants independently accessed the study link and answered the questionnaire after signing the digital informed consent form (DocuSign, Inc., San Francisco, CA, USA). The data collection comprised two stages: (1) self-reported data, such as sociodemographic information (gender, age, country, and region), clinical data (symptoms and prior illnesses) and the RT-PCR/TRAg test result for SARS-CoV-2 (collected previously), and (2) vocal data, obtained through exercises of (a) breathing, (b) coughing, and (c) vocalisation of vowels and phrases (previously defined in the form). For the development of the algorithm, only

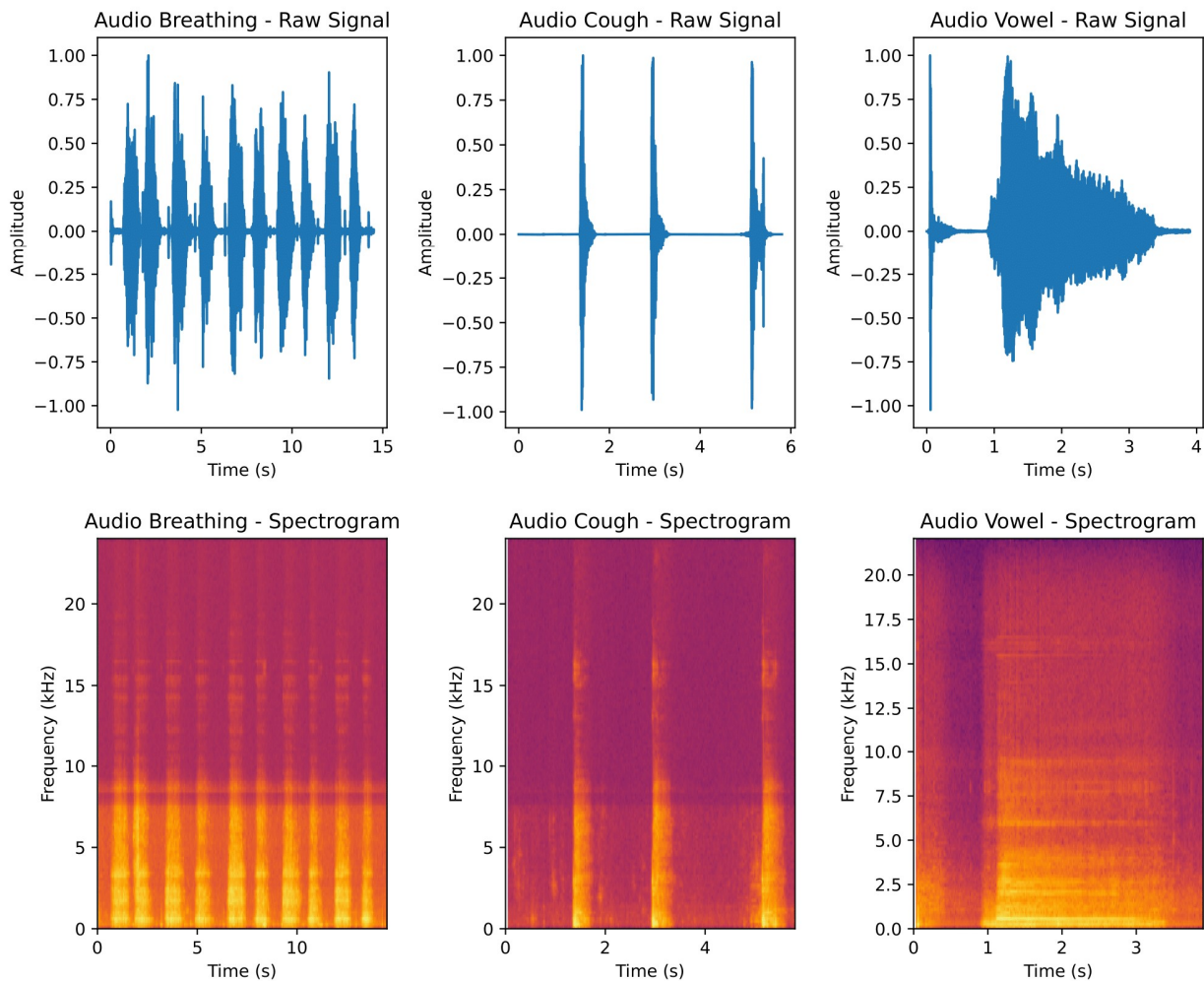
the responses and vocal recordings given up to five days after performing the SARS-CoV-2 test were considered. European Union General Data Protection and Regulation (GDPR) privacy laws were respected.



**Figure 1.** Data collection user-interface. The (left) screen outlines the application initialisation. The (right) screen shows up upon the collection of audio samples.

### 2.3. Audio Quality Assessment

Audios were recorded into a raw, uncompressed format (WAV, 16-bit), sampled at 16 kHz, ensuring a good quality standardisation of this procedure across different devices. After, recordings underwent a bit-depth normalisation to make them comparable under the same amplitude scale. To ensure their quality to feed further stages of the algorithm development (feature extraction, training, evaluation), a pre-trained audio classification model called YAMNet was used [21]. YAMNet is a DL architecture derived from MobileNetV1 and pre-trained on 16 kHz audios from a large audio dataset (AudioSet). It can classify audio segments into 521 audio/speech events. Specific events (breathing, cough) were used as reference points to validate our collected audio recordings by passing them through this model. Participants were only accepted if at least 50% of their audio recordings met the quality assessment criteria (i.e., fell into one of the pre-selected acceptable sound classes). Some examples of audio recordings collected from different vocalisation exercises are shown in Figure 2. In further steps, the features were only calculated for cough and breath recordings, as these were generally of better quality.



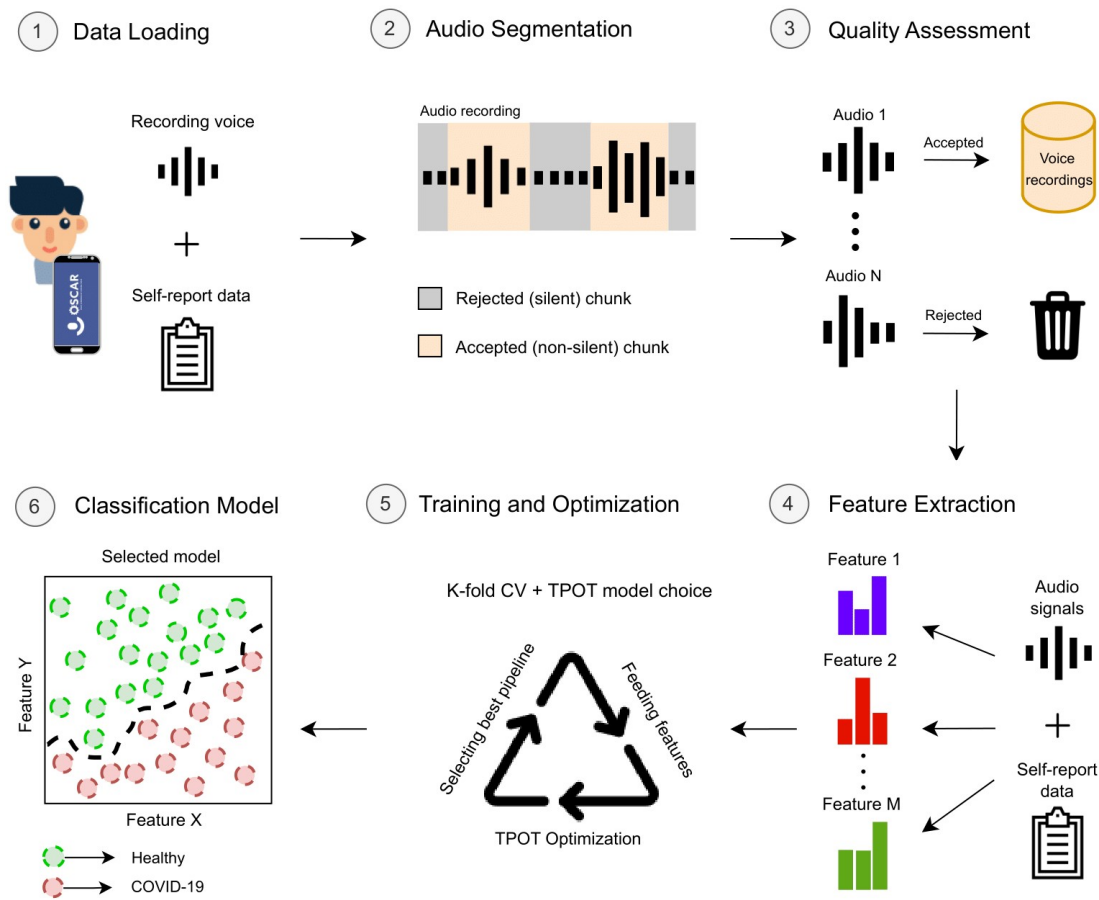
**Figure 2.** Representation of multiple audio recordings in the temporal domain (**top row**) and time-spectral domain (**bottom row**). The spectrogram representation indicates the prevalence of frequencies over specific time intervals, with lighter colours representing higher frequency preponderance.

#### 2.4. Algorithm Development

The algorithm designed was intended to solve a binary classification problem in which participants were classified as infected or not with the SARS-CoV-2 virus. Thus, the positive label was assigned to infected participants and the negative label to non-infected participants.

Machine Learning (ML) aims to learn from observational data, which, in this case, comprises the participants' voice recordings and questionnaire answers. Using signal processing expertise, all these data sources and extracted relevant characteristics from the audio signals were gathered and fed into decision-based (ML) algorithms that adjust their parameters according to the data observed during the training process. The whole data processing flow included multiple stages, namely, (a) data acquisition, (b) audio segmentation, (c) quality assessment, (d) feature extraction, (e) model training and optimisation, and (f) evaluation. Figure 3 illustrates the processing steps, after which the model is ready to evaluate data from new incoming participants. The algorithms were developed in Python language, and the pipeline (including the extracted features) was equivalent to that applied in [4], optimised using the TPOT framework [22]. Every pipeline was structured to be composed of a (i) feature scaler, (ii) feature selector, and a (iii) classification model. The TPOT's optimisation search is conducted using genetic algorithms, involving selection, crossover, and mutation of the parameters across generations. The parameter search of scalers, selectors, and classifiers for our experiments was already defined by default in

TPOT’s API [22]. Table 1 provides a detailed description of the classification parameters over which optimisation was performed during the training stage.



**Figure 3.** Illustration of the full data processing pipeline. It goes from the collection of audio recordings up to model training, optimisation, and evaluation.

**Table 1.** Description of the parameter search space for classification models considered by TPOT. Parameters were based on TPOT’s default ranges and used to select the best-performing classification model for the proposed binary classification task.

Model	Parameters	Search Range
NB	alpha	$\{1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}, 1, 10, 100\}$
	fit_prior	{True, False}
Tree-based(RF, ET, GB, XGB, DT)	criterion	{“gini”, “entropy”}
	max_depth	[1, 11]
	max_features	[0.05, 1.0]
	min_samples_split	[2, 21]
	min_samples_leaf	[1, 21]
	min_child_weight	[1, 21]
	learning_rate	$\{1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}, 0.5, 1\}$
	subsample	[0.05, 1]
	bootstrap	[True, False]

Table 1. Cont.

Model	Parameters	Search Range
KNN	n_neighbors	[1, 101]
	weights	{"uniform", "distance"}
	p	{1, 2}
SVM	penalty	{"l1", "l2"}
	loss	{"hinge", "squared_hinge"}
	dual	{True, False}
	tol	$[1 \times 10^{-5}, 1 \times 10^{-1}]$
	c	$[1 \times 10^{-4}, 25]$
LR	penalty	{"l1", "l2"}
	c	$[1 \times 10^{-4}, 25]$
	dual	{True, False}
SGD	loss	{"log", "hinge", "modified_huber", "squared_hinge", "perceptron"}
	penalty	{"elasticnet"}
	alpha	$\{0, 1 \times 10^{-1}, 1 \times 10^{-2}\}$
	learning_rate	{"invscaling", "constant"}
MLP	alpha	$\{1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}\}$
	learning_rate_init	$\{1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}, 0.5, 1\}$

NB—Naïve Bayes; RF—Random Forest; ET—Extra Trees, GB—Gradient Boosting, XGB—eXtreme Gradient Boosting; DT—Decision Tree; KNN—K-Nearest Neighbors; SVM—Support Vector Machine; LR—Logistic Regression; SGD—Stochastic Gradient Descent; MLP—Multi-Layer Perceptron.

### 2.5. Audio Segmentation and Event Detection

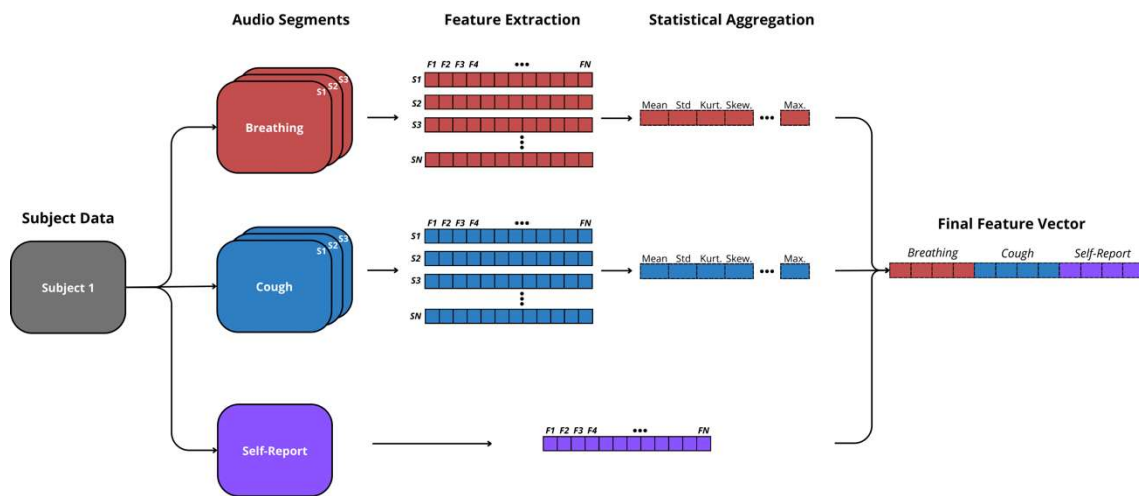
Audio segmentation and segment quality assessment are essential steps in the proposed pipeline, as errors in this stage will directly propagate to the subsequent stages, particularly feature extraction and model training (e.g., speech portion found during a breathing recording will affect feature's meaningfulness). In our data curation scenario, two quality verification filters are applied before audio segments undergo further stages, namely, (a) a non-silent segment extractor, which retrieves non-silent audio portions of a particular recording, and (b) a speech event detector that accepts each segment retrieved if it belongs to the recording it has been extracted from (e.g., breathing segments are only accepted in breathing audio files). The non-silent segment extractor is an energy-based algorithm from Librosa Python library [23], whereas the event detector is a pre-trained CNN that assigns an audio signal portion into one of (i) breathing, (ii) cough, (iii) voice, (iv) vowel-a, (v) vowel-e, or (vi) vowel-o categories. The event detector has been pre-trained in a curated subset of the Coswara dataset, achieving a macro F1-score of 84.6% [4].

### 2.6. Feature Extraction

Feature extraction is a crucial step in the processing pipeline, as it involves transforming raw audio data into a set of measurable characteristics that can be used for ML model training. After audio recordings underwent segmentation and quality assessment, features were extracted from various domains to capture the diverse aspects of the audio signals.

Each audio segment was divided into equally sized time windows, from which features were computed and averaged into a single value. The window size varied depending on the recording type, specifically, 32 ms + 50% overlap for breathing, and 16 ms + 75% overlap for cough, whose values were chosen after preliminary tests performed in [4]. Features were extracted from the (a) spectral, (b) cepstral, (c) temporal, (d) amplitude, and (e) phonetic domains of the audio segments retrieved, some of them clinically relevant from a physiological perspective (all features computed are detailed in [4]). This segment-wise approach resulted in a distinct set of features for each segment, so compatible features were aggregated using statistical descriptors (e.g., mean, std dev, kurtosis, skewness, maximum, minimum). Features from self-reported questions were simply encoded in this process.

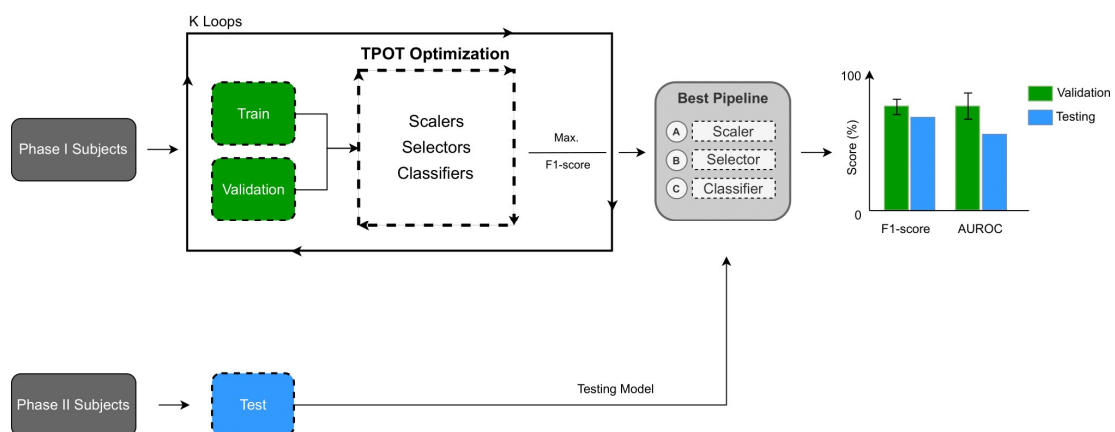
Finally, all feature domains were merged into a single vector that subsequently fed the prediction model training stage. The procedure behind the feature extraction stage is illustrated in Figure 4.



**Figure 4.** Illustration of the feature extraction pipeline. It starts with audio segments extracted by the segmenter, specifically focusing on breathing and cough sounds. The feature extraction procedure computes spectral, cepstral, temporal, amplitude, and phonetic features. These are aggregated using statistical descriptors. The final vector combines all aggregated features for breathing, cough, and self-reported data.

### 2.7. Training and Evaluation

Model training and validation procedures were carried out using a K-fold cross-validation strategy [24]. This process involved an iterative training method in which the whole training set was divided into K equally sized folds of subjects. In each iteration, the model is trained with K-1 subject folds and tested on the remaining fold. The process is repeated K times to ensure the entire data set has been evaluated in distinct iterations. In our experiments, K = 5, which means that, at each iteration, 80% of the subjects were kept for training and 20% were left for validation. Finally, the results were averaged to capture the global performance scores and decide which pipeline best suits the use case. The optimisation process is illustrated in Figure 5.



**Figure 5.** Illustration of the training, validation, and optimisation strategies. Data from Phase I subjects undergo the training loop fine-tuned by a TPOT optimisation strategy. Each pipeline (scaler, selector, and classifier) is trained and tested using K distinct combinations of patients (K = 5, 80% to train and 20% for validation). The best pipeline (maximise F1-score, macro) is selected and tested against Phase II subjects, on whom final performance is reported.

The chosen pipeline was tested on Phase II participants (test set) after training and validation of the algorithm on Phase I subjects. Standard metrics were used to evaluate the model performance and translate the algorithm capabilities into the clinical context. A confusion matrix was used to represent a comprehensive overview of the model performance and confirm which classes the model predicts better or worse [25]. Multiple metric scores were then estimated from this matrix, particularly Sensitivity, Specificity, Positive and Negative Predictive Values (PPV, NPV), Accuracy, and F1-Score. The area under the receiver operating characteristic curve (AUROC) was also computed to establish a broader evaluation of Sensitivity and Specificity.

### 3. Results

A total of 322 people agreed to participate in Phase I. After quality-wise filtering, the training set was reduced by 51.6%, leaving 166 subjects effectively selected for training the classification model (34.3% SARS-CoV-2 positive). In Phase II, the test group was reduced by 52.8%, leaving 58 subjects available to evaluate the algorithm's performance (69.0% SARS-CoV-2 positive). In Phase I, a 5-fold cross-validation strategy was employed, with the dataset being split into 80% for training and 20% for testing at each iteration. For Phase II, the pipeline selected in the previous stage was evaluated on the entire dataset (test set).

Observing the outcomes of Table 2, approximately 60.0% of the participants in both phases were female ( $p = 0.948$ ), and the majority were under 50 years old ( $p = 0.470$ ). The prevalent clinical symptoms reported in Phase I were flu-like, cough, fever, and sore throat. In Phase II, flu-like symptoms, cough, muscle/body aches, and headaches were the most frequent ones. A distinct symptomatic profile seemed to show up, being confirmed by the significant group differences identified in the Symptoms domain ( $p$ -value: 0.003). Infection rates also revealed distinct profiles, with Phase I subjects displaying a more significant proportion of non-infected subjects than Phase II (65% vs. 27%,  $p$ -value < 0.001).

**Table 2.** Population characterisation by gender, age, and symptoms, regarding both Phase I and Phase II groups.

Demographic and Clinical Data		Distribution by Groups		$p$ -Value <sup>a</sup>
		Phase I (N = 166)	Phase II (N = 58)	
Gender	Female	99	35	0.948
	Male	66	23	
	N.A.	1	0	
Age (years)	18–29	70	17	0.470
	30–39	30	20	
	40–49	35	8	
	50–59	21	9	
	60–69	8	3	
	70–79	2	1	
Symptoms <sup>b</sup>	Flu-like	49	35	0.003 *
	Cough	45	33	
	Fever	33	12	
	Sore throat	33	19	
	Body aches/myalgias	30	32	
	Headaches	27	23	
	Loss of smell/taste	18	1	
	Shortness of breath	8	5	
SARS-CoV-2	Diarrhoea	8	1	<0.001 *
	Infected	57	40	
	Not Infected	109	18	

N.A.: Not answered; \*—significance value at 5%; <sup>a</sup>— $p$ -values were obtained using a non-parametric one-way ANOVA/Chi-squared test; <sup>b</sup>—multiple-choice items (does not total 100%).

During this first quality-wise filtering step, recordings were evaluated by the YAMNet model, and eligible subjects were moved into the next stage. Then, breathing and cough recordings were split into segments, where each was evaluated and accepted/rejected by the event detector model. After this step, the feature extraction phase was carried out, followed by model training and evaluation. During the training stage, five distinct combinations of Phase I participants were defined to iteratively fine-tune each model (subject-wise 5-fold cross-validation) [22], thus mitigating the influence of random splits in the allocation of subjects and model selection. The pipeline that achieved the highest F1-score (macro) was selected, as this score is a suitable alternative to accuracy for datasets with unbalanced classes. After this process, a Gaussian Naïve Bayes classification model, combined with a One Hot Encoder scaler and a Family-wise Error rate feature selector, provided the best results, achieving an averaged 88.2% macro F1-score (validation score). Results from the training and validation stages are provided in detail in Table 3, where ten distinct obtained pipelines are displayed in descending order of performance.

**Table 3.** Ranking of top 10 pipelines returned by TPOT using data from Phase I participants. F1-score (macro average) was defined to establish the pipeline selection criteria score. Scores are displayed in descending order and results from an average over five distinct subject train/test folds. The best-performing and selected pipeline is bolded and underscored.

Dataset	Rank	TPOT Pipeline			F1-Score (%) *
		Scaler	Selector	Classifier	
OSCAR	<u>1</u>	<u>OneHot</u>	<u>FwE</u>	<u>Gauss. NB</u>	<u>88.2</u>
	2	OneHot	FwE	Multi. NB	86.3
	3	Binarizer	Var. Thresh	SVM Linear	84.5
	4	Polynomial	Percentile	Extra Trees	83.7
	5	Robust	Var. Thresh	Random Forest	80.1
	6	RBF Sampler	FwE	SVM Linear	74.3
	7	RBF Sampler	FwE	MLP	70.5
	8	Max. Abs.	FwE	Extra Trees	66.4
	9	Robust	FwE	KNN	60.5
	10	Robust	RFE	MLP	52.9

\*—F1-score (macro) value from a 5-fold average of distinct participant splits; Max. Abs.—Maximum Absolute scaler; One Hot—One Hot Encoder; Polynomial—Polynomial feature generation; Var. Thresh—Variance Threshold; Percentile—Percentile selector; FwE—Family-wise Error rate selector; RFE—Recursive Feature Elimination; MLP—Multi-Layer Perceptron; Gauss. NB—Gaussian Naïve Bayes; Multi. NB—Multinomial Naïve Bayes; KNN—K Nearest Neighbors.

The real performance of the selected pipeline was finally assessed using a hold-out testing set (Phase II subjects). The confusion matrix depicting the results obtained from the evaluation of Phase II is shown in Table 4. Regarding performance metrics (Table 5), a sensitivity of 85.0%, specificity of 88.9%, PPV of 94.4%, NPV of 72.7%, and F1-score of 84.7% were obtained. The overall AUROC score reached 94.0%. When evaluating the same score by gender, the model achieved 87.0% for males and 97.0% for females. Considering the age factor, the model obtained 95.0%, 91.0%, 100%, 89.0%, and 100% for the age ranges of 18–29 (N = 17), 30–39 (N = 20), 40–49 (N = 8), 50–59 (N = 9), and 60–69 (N = 3), respectively. When considering the number of days between taking the SARS-CoV-2 diagnostic test and recording audio samples, the model performed less accurately when the test and recording were done on the same day (78.0%). This performance improved when audio samples were recorded one day (93%), three days (100%), and five days (89%) later.

**Table 4.** Classification model raw performance on Phase II testing set. The confusion matrix reports the raw behaviour of the Naïve Bayes algorithm trained on Phase I subjects and tested on Phase II participants.

		Expected	
		COVID-19	Without COVID-19
Predicted	COVID-19	34	2
	Without COVID-19	6	16

**Table 5.** Classification model performance scores on Phase II testing set. In this refined version, seven evaluation metrics are computed and reported (%) to capture a global perspective of the Naïve Bayes model performance.

Dataset	Sensitivity	Specificity	Performance Metrics (%)				AUROC	Accuracy
			PPV	NPV	F1-Score *			
OSCAR Dataset	85.0	88.9	94.4	72.7	84.7	94.0	86.2	

\*—macro score; PPV—positive predictive value; NPV—negative predictive value; AUROC—area under the ROC curve.

To check the validity of the extracted audio features, we conducted a preliminary experiment using only the features extracted from the audios and discarding the self-reported data to obtain a final model. In general, the results decreased slightly, with the model misclassifying 11 out of 58 subjects when tested with Phase II subjects. Sensitivity, specificity, and F1-score reached 80.0%, 83.3%, and 79.3%, respectively.

#### 4. Discussion

The development of digital voice biomarkers holds significant potential in different medical fields [26,27], as they provide non-invasive, cost-effective, and accessible means of monitoring and diagnosing health conditions. This innovative approach uses digital technology and machine learning to improve healthcare outcomes. It has already been shown to be useful in detecting cognitive impairment in preclinical and prodromal Alzheimer’s dementia [28], in Parkinson’s disease [29], as well as in identifying exacerbations of asthma [30–32] and chronic obstructive pulmonary diseases [33].

Different research groups have tried to develop more convenient and accessible methods to diagnose COVID-19 [7]. One of the methodologies used is the analysis of voice recordings using machine learning to design predictive algorithms [12,34,35]. It is anticipated that these algorithms can be used as a screening tool that any person can use before accessing a clinical facility asking for a common antigen or RT-PCR test.

Even though the World Health Organisation no longer considers COVID-19 a global health emergency, it is still crucial for States Parties to consider how they can improve their preparedness for future outbreaks. In this sense, continuous research is needed to find easily implementable detection methods [36].

In Phase I, we developed an algorithm for screening COVID-19-compatible symptoms using a real-world set of community-based participants from the population. Sensitivity and specificity were two clinically relevant metrics to evaluate the model’s performance in the test group (Phase II). Consequently, we expected the model to be sensitive to COVID-19 cases and specific enough not to classify all cases as COVID-19. Specificity should, therefore, ensure that a minimum number of negative cases (without COVID-19) are classified as COVID-19 (false positive).

It is worth noting that the two recruitment phases were conducted at different times, and different SARS-CoV-2 variants prevailed. In Phase I, the Delta variant was most prevalent in Portugal, while in Phase II, a higher proportion of Omicron-linked cases were reported. Table 2 shows that although gender and age appeared to be similarly

distributed in both phases, there were significant differences in symptomatic profiles. Although the model was developed based on Phase I data, it was still able to distinguish well enough between infected and uninfected participants in Phase II. This could give rise to the hypothesis that the vocal characteristics of COVID-19 patients may be stable across different virus variants.

Our results show that the algorithm provides relevant insights by exhibiting appropriate behaviour and misclassifying only 8 out of 58 subjects available for testing. Furthermore, it shows adequate sensitivity in identifying positive and negative cases (85.0% sensitivity, 88.9% specificity) and a high positive predictive value (up to 94.4%). These results suggest that the model predicts positive SARS-CoV-2 cases more reliably than negative cases (considering the lower negative predictive value of 72.7). Gender differences led to a change in the model's AUROC values of about 10%, with the female group performing better than the male group. This fact should be considered in future studies with regard to the need for separate models to account for such differences.

Regarding age differences, the AUROC scores did not show significant variations, except for a slight decrease in performance observed in the 50–60 age range. COVID-19 fatality rates are much higher in older people than in the youth population [37]. If age-related factors influence the model's decision-making process, priority and increased importance should be given to older age groups during algorithm development, considering their lower likelihood of survival.

When looking at the time between performing the test and recording the audio samples, the model performed better the more days passed between the two events. This may indicate that the features capture symptom patterns that do not appear or intensify until several days after diagnosis [38]. This can have practical implications for the timing of the diagnosis, as a substantial part of the SARS-CoV-2 virus transmission occurs during the presymptomatic phase [39]. However, a more comprehensive validation using a larger data set would help to confirm and strengthen all hypotheses and results, and to better understand the biological mechanisms behind these patterns.

The model's performance also decreased when the self-reported data were removed from the feature set (84.7% to 79.3% macro-F1 score), although it still showed appropriate behavior. This suggests that subjects' self-assessment may complement the audio-based patterns, which may not be able to identify some difficult cases (e.g., already with symptoms but without manifestations on the vocal tract).

In terms of performance, asymptomatic participants (with symptoms but not infected) may also explain the model's worse performance, as we found more misclassified COVID-19-infected than non-infected individuals. Even so, the model achieved a macro F1-score of 84.7%, accuracy of 86.2%, and AUROC of 94.2%, which is remarkable. Even though most studies described previously [12–14] achieved notable performance scores (AUROC ranging from 84.6% to 98.6%), those also relied on their training and evaluation strategies in crowdsourced, non-curated audio data obtained from uncontrolled settings and associated with non-validated SARS-CoV-2 labels, which might have masked the validity of potential outcomes and the achievement of identical performance scores. Alternative studies conducted by Robotti et al. and Costantini et al. [17,18] designed and conducted their own data collection protocol, revealing more consistent outcomes with optimal AUROC scores of 94.0% obtained against a hold-out testing set of 30 and 36 subjects, respectively. Both the performance (94.0% vs. 94.2%) and the test set size (30/36 vs. 58) were similar and comparable, suggesting consistency across the three studies, all of which designed a precise and controlled data collection protocol with validated labels as ground truth. Some differentiating factors of our study, such as the meticulous data collection, the data curation procedure (quality analysis, segmentation, and event detection), and the use of validated SARS-CoV-2 tests to target participants as positive or negative, might have contributed to obtaining such results and definitely supported the validity of these findings.

From an ethical perspective, the use of voice data and algorithms in healthcare requires careful consideration of patient autonomy and consent. Voice data is inherently personal

and can reveal sensitive information. For this reason, robust encryption and data protection protocols are needed to prevent unauthorised access or misuse. Moreover, regulations such as the GDPR in Europe impose strict requirements for the collection, storage, and sharing of personal data. The AI Act regulation will also require accountability and transparency on the limitations, misuses, and validation of predictive models used in clinical practice. Such regulations will help ensure the protection of patient data and safety of AI systems upon their usage in real-world settings.

In future work, validation in a larger population group should be performed to prove the efficiency of the approach at a national scale. Our testing set does not fully represent the target population, so performance is expected to change as long as the sample size increases. Additionally, in a follow-up study, our team plans to assess the contribution of each data source to the model's performance using ablation techniques. This will help identify the most important features that could serve as potential digital biomarkers for COVID-19 screening or broader respiratory health assessments. Moreover, such endpoints might also reveal links between underlying biological mechanisms in the vocal tract, symptoms, and characteristics of vocal sounds, being of great interest to inform the medical community, in particular, pulmonology specialists.

## 5. Conclusions

Voice screening algorithms offer an attractive approach to assess respiratory diseases, including COVID-19. This study demonstrates a rapid and convenient method that has the potential to improve the screening strategy to identify positive cases and reduce disease transmission. With sensitivity and specificity rates of 85.0% and 88.9%, respectively, the algorithm correctly identifies most cases, so its performance in a real-world scenario is promising.

One of the key findings of this study is the potential for expansion beyond COVID-19, where identical approaches could be explored to monitor and characterise other respiratory diseases such as asthma or chronic obstructive pulmonary disease (COPD), as well as respiratory infections like seasonal influenza and tuberculosis. The easy accessibility and simplicity of such a data collection method offers significant advantages for extending such approaches to any clinical unit, as audio recordings are non-invasive, inexpensive, and easy to collect.

Our findings highlight the promising potential of voice-based screening strategies, but also leave room for future research in the field of audio signal processing in combination with ML as a complementary method for screening respiratory diseases, especially in controlled clinical settings. Further research should focus on refining such algorithms and processing pipelines, investigating explainability techniques to ensure clinical interpretability of the results, and validating outcomes in large population cohorts to better account for limitations and the impact of specific sources of bias (e.g., gender, age, ethnicity) and maximise clinical utility. Advances in this area will improve diagnostic procedures and better support early intervention in patients.

**Author Contributions:** Conceptualization—P.C.-M., N.N., P.P., H.G. and J.S.; methodology—P.C.-M., N.N., H.G., J.S., P.G. and P.P.; software—H.G., P.M., F.S. and P.G.; validation—H.G., P.M., F.S. and P.G.; formal analysis—H.G., P.M., F.S. and P.G.; investigation—P.C.-M., N.N., P.P., I.C., H.G., P.M., F.S. and J.S.; resources—J.S., N.N., H.G. and F.S.; data curation—N.N., H.G., P.C.-M. and I.C.; writing—P.C.-M., P.M., H.G. and P.P.; writing—review and editing—P.C.-M., P.P., N.N., I.C., H.G., F.S., P.M., P.G. and J.S.; visualization—P.C.-M., P.M., H.G. and N.N.; supervision—N.N. and H.G.; project—N.N. and J.S.; funding acquisition—N.N., J.S. and H.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** The present publication was funded by Fundação Ciência e Tecnologia, IP national support through CHRC (UIDP/04923/2020). The project "OSCAR—vOice Screening of Corona viRus" (POCI-01-02B7-FEDER-051277) leading to this work is co-financed by ERDF-European Regional Fund through the Operational Program for Competitiveness and Internationalization, and by LISBOA 2020—Regional Operational Program for Lisbon and Vale do Tejo.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of NOVA Medical School/Faculdade de Ciências Médicas, Universidade Nova de Lisboa (78/2021/CEFCM, 31 May 2021).

**Informed Consent Statement:** Digitally signed informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data not available due to ethical restrictions.

**Acknowledgments:** The authors thank the participants and the centres' staff responsible for the SARS-CoV-2 tests.

**Conflicts of Interest:** Author Joana Sousa was employed by NOS Inovação, the remaining authors declare that the re-search was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Windmon, A.; Minakshi, M.; Bharti, P.; Chellappan, S.; Johansson, M.; Jenkins, B.A.; Athilingam, P.R. TussisWatch: A Smart-Phone System to Identify Cough Episodes as Early Symptoms of Chronic Obstructive Pulmonary Disease and Congestive Heart Failure. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1566–1573. [CrossRef] [PubMed]
2. Botha, G.H.R.; Theron, G.; Warren, R.M.; Klopper, M.; Dheda, K.; van Helden, P.; Niesler, T.R. Detection of tuberculosis by automatic cough sound analysis. *Physiol. Meas.* **2018**, *39*, 45005. [CrossRef]
3. Porter, P.; Abeyratne, U.; Swarnkar, V.; Tan, J.; Ng, T.-W.; Brisbane, J.M.; Speldewinde, D.; Choveaux, J.; Sharan, R.; Kosasih, K.; et al. A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children. *Respir. Res.* **2019**, *20*, 81. [CrossRef]
4. Matias, P.; Costa, J.; Carreiro, A.V.; Gamboa, H.; Sousa, I.; Gomez, P.; Sousa, J.; Neuparth, N.; Carreiro-Martins, P.; Soares, F. Clinically Relevant Sound-Based Features in COVID-19 Identification: Robustness Assessment With a Data-Centric Machine Learning Pipeline. *IEEE Access* **2022**, *10*, 105149–105168. [CrossRef]
5. Mohammed, E.A.; Keyhani, M.; Sanati-Nezhad, A.; Hejazi, S.H.; Far, B.H. An ensemble learning approach to digital corona virus preliminary screening from cough sounds. *Sci. Rep.* **2021**, *11*, 15404. [CrossRef] [PubMed]
6. Deshpande, G.; Batliner, A.; Schuller, B.W. AI-Based human audio processing for COVID-19: A comprehensive overview. *Pattern Recognit.* **2022**, *122*, 108289. [CrossRef]
7. Carreiro-Martins, P.; Paixão, P.; Caires, I.; Rodrigues, A.; Matias, P.; Gamboa, H.; Carreiro, A.; Soares, F.; Gomez, P.; Sousa, J.; et al. Diagnosis of COVID-19 by sound-based analysis of vocal recordings. *Pulmonology* **2023**, *29*, 455–456. Available online: <https://www.journalpulmonology.org/en-diagnosis-covid-19-by-sound-based-analysis-articulo-S2531043723000570> (accessed on 27 June 2024). [CrossRef]
8. Sharma, N.; Krishnan, P.; Kumar, R.; Ramoji, S.; Chetupalli, S.R.; Ghosh, P.K.; Ganapathy, S. Coswara—A database of breathing, cough, and voice sounds for COVID-19 diagnosis. Proceedings of the Annual Conference of the International Speech Communication Association. *Interspeech* **2020**, *2020*, 4811–4815.
9. Xia, T.; Spathis, D.; Ch, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Bondareva, E.; Dang, T.; Floto, A.; Cicuta, P.; et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. In Proceedings of the Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), Virtual Event, 4–28 August 2021; pp. 1–13.
10. Brown, C.; Chauhan, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Spathis, D.; Xia, T.; Cicuta, P.; Mascolo, C. 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, Virtual Event, 6–10 July 2020; ACM: New York, NY, USA, 2020; pp. 3474–3484. Available online: <https://dl.acm.org/doi/10.1145/3394486.3412865> (accessed on 28 June 2024).
11. Anupam, A.; Mohan, N.J.; Sahoo, S.; Chakraborty, S. Preliminary Diagnosis of COVID-19 Based on Cough Sounds Using Machine Learning Algorithms. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1391–1397.
12. Coppock, H.; Gaskell, A.; Tzirakis, P.; Baird, A.; Jones, L.; Schuller, B. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: A pilot study. *BMJ Innov.* **2021**, *7*, 356–362. Available online: <https://innovations.bmj.com/content/7/2/356> (accessed on 12 July 2024). [CrossRef]
13. Pahar, M.; Klopper, M.; Warren, R.; Niesler, T. COVID-19 cough classification using machine learning and global smartphone recordings. *Comput. Biol. Med.* **2021**, *135*, 104572. Available online: <https://www.sciencedirect.com/science/article/pii/S0010482521003668> (accessed on 30 June 2024). [CrossRef]
14. Melek, M. Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound. *Neural Comput. Appl.* **2021**, *33*, 17621–17632. [CrossRef] [PubMed]
15. Meister, J.A.; Nguyen, K.A.; Luo, Z. Audio Feature Ranking for Sound-Based COVID-19 Patient Detection. In *Progress in Artificial Intelligence*; Marreiros, G., Martins, B., Paiva, A., Ribeiro, B., Sardinha, A., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 146–158.

16. Budd, J.; Baker, K.; Karoune, E.; Coppock, H.; Patel, S.; Payne, R.; Cañadas, A.T.; Titcomb, A.; Hurley, D.; Egglestone, S.; et al. A large-scale and PCR-referenced vocal audio dataset for COVID-19. *Sci. Data* **2024**, *11*, 700. [CrossRef]
17. Robotti, C.; Costantini, G.; Saggio, G.; Cesarini, V.; Calastri, A.; Maiorano, E.; Piloni, D.; Perrone, T.; Sabatini, U.; Ferretti, V.V.; et al. Machine Learning-based Voice Assessment for the Detection of Positive and Recovered COVID-19 Patients. *J. Voice* **2021**, *38*, 796.e1–796.e13. Available online: <https://linkinghub.elsevier.com/retrieve/pii/S089219972100388X> (accessed on 19 July 2024). [CrossRef]
18. Costantini, G.; Cesarini, V.; Robotti, C.; Benazzo, M.; Pietrantonio, F.; Di Girolamo, S.; Pisani, A.; Canzi, P.; Mauramati, S.; Bertino, G.; et al. Deep learning and machine learning-based voice analysis for the detection of COVID-19: A proposal and comparison of architectures. *Knowl. Based Syst.* **2022**, *253*, 109539. Available online: <https://linkinghub.elsevier.com/retrieve/pii/S0950705122007754> (accessed on 4 July 2024). [CrossRef]
19. Hassan, A.; Shahin, I.; Alsabek, M.B. COVID-19 Detection System using Recurrent Neural Networks. In Proceedings of the 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), Sharjah, United Arab Emirates, 3–5 November 2020; pp. 1–5. Available online: <https://ieeexplore.ieee.org/document/9256562/> (accessed on 4 July 2024).
20. Modi, B.; Sharma, M.; Hemani, H.; Joshi, H.; Kumar, P.; Narayanan, S.; Shah, R. Analysis of Vocal Signatures of COVID-19 in Cough Sounds: A Newer Diagnostic Approach Using Artificial Intelligence. *Cureus* **2024**, *16*, e56412. [CrossRef]
21. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN architectures for large-scale audio classification. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; Available online: <https://arxiv.org/pdf/1609.09430> (accessed on 5 July 2024).
22. Le, T.T.; Fu, W.; Moore, J.H. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics* **2020**, *36*, 250–256. [CrossRef] [PubMed]
23. Brian, M.; Raffel, C.; Liang, D.; Ellis, D.P.W.; McVicar, M.; Battenberg, E.; Nieto, O. Librosa: Audio and Music Signal Analysis in Python. In Proceedings of the 14th Python in Science Conference, Austin, TX, USA, 6–12 July 2015; pp. 18–25.
24. Tougui, I.; Jilbab, A.; El Mhamdi, J. Impact of the Choice of Cross-Validation Techniques on the Results of Machine Learning-Based Diagnostic Applications. *Healthc. Inform. Res.* **2021**, *27*, 189–199. Available online: <http://e-hir.org/journal/view.php?doi=10.4258/hir.2021.27.3.189> (accessed on 21 June 2024). [CrossRef] [PubMed]
25. Olivetti, E.; Greiner, S.; Avesani, P. Statistical independence for the evaluation of classifier-based diagnosis. *Brain Inform.* **2015**, *2*, 13–19. [CrossRef]
26. Rogers, H.P.; Hseu, A.; Kim, J.; Silberholz, E.; Jo, S.; Dorste, A.; Jenkins, K. Voice as a Biomarker of Pediatric Health: A Scoping Review. *Children* **2024**, *11*, 684. [CrossRef]
27. Fagherazzi, G.; Fischer, A.; Ismael, M.; Despotovic, V. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. *Digit. Biomark.* **2021**, *5*, 78–88. [CrossRef]
28. Hajjar, I.; Okafor, M.; Choi, J.D.; Moore, E.; Abrol, A.; Calhoun, V.D.; Goldstein, F.C. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer’s disease. *Alzheimer’s Dement. Diagn. Assess. Dis. Monit.* **2023**, *15*, e12393. Available online: <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/dad2.12393> (accessed on 12 July 2024). [CrossRef] [PubMed]
29. Suppa, A.; Costantini, G.; Asci, F.; Di Leo, P.; Al-Wardat, M.S.; Di Lazzaro, G.; Scalise, S.; Pisani, A.; Saggio, G. Voice in Parkinson’s Disease: A Machine Learning Study. *Front. Neurol.* **2022**, *13*, 831428. Available online: <https://www.frontiersin.org/articles/10.3389/fneur.2022.831428/full> (accessed on 12 July 2024). [CrossRef] [PubMed]
30. Iqbal, M.A.; Devarajan, K.; Ahmed, S.M. An optimal asthma disease detection technique for voice signal using hybrid machine learning technique. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6856. [CrossRef]
31. Molfino, N.A.; Turcatel, G.; Riskin, D. Machine Learning Approaches to Predict Asthma Exacerbations: A Narrative Review. *Adv. Ther.* **2024**, *41*, 534–552. [CrossRef]
32. Jayamini, W.K.D.; Mirza, F.; Naeem, M.A.; Chan, A.H.Y. Investigating Machine Learning Techniques for Predicting Risk of Asthma Exacerbations: A Systematic Review. *J. Med. Syst.* **2024**, *48*, 49. [CrossRef]
33. Shen, X.; Liu, H. Using machine learning for early detection of chronic obstructive pulmonary disease: A narrative review. *Respir. Res.* **2024**, *25*, 336. [CrossRef]
34. Benmalek, E.; Elmhamdi, J.; Jilbab, A.; Jbari, A. Automatic COVID-19 detection using machine learning and voice recording. *Res. Biomed. Eng.* **2023**, *39*, 597–612. [CrossRef]
35. Verde, L.; De Pietro, G.; Sannino, G. Artificial Intelligence Techniques for the Non-invasive Detection of COVID-19 Through the Analysis of Voice Signals. *Arab. J. Sci. Eng.* **2023**, *48*, 11143–11153. [CrossRef]
36. WHO. Statement on the Fifteenth Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Coronavirus Disease (COVID-19) Pandemic. 2023. Available online: [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic) (accessed on 15 July 2024).

37. Camacho Moll, M.E.; Mata Tijerina, V.L.; Silva Ramírez, B.; Peñuelas Urquides, K.; González Escalante, L.A.; Escobedo Guajardo, B.L.; Cruz Luna, J.E.; Corrales Pérez, R.; Gómez García, S.; Bermúdez de León, M. Sex, Age, and Comorbidities Are Associated with SARS-CoV-2 Infection, COVID-19 Severity, and Fatal Outcome in a Mexican Population: A Retrospective Multi-Hospital Study. *J. Clin. Med.* **2023**, *12*, 2676. Available online: <https://www.mdpi.com/2077-0383/12/7/2676>. (accessed on 16 July 2024). [[CrossRef](#)]
38. Drain, P.K.; Dalmat, R.R.; Hao, L.; Bemer, M.J.; Budiawan, E.; Morton, J.F.; Ireton, R.C.; Hsiang, T.Y.; Marfatia, Z.; Prabhu, R.; et al. Duration of viral infectiousness and correlation with symptoms and diagnostic testing in non-hospitalized adults during acute SARS-CoV-2 infection: A longitudinal cohort study. *J. Clin. Virol.* **2023**, *161*, 105420. Available online: <https://www.sciencedirect.com/science/article/pii/S1386653223000422> (accessed on 16 July 2024). [[CrossRef](#)]
39. He, X.; Lau, E.H.Y.; Wu, P.; Deng, X.; Wang, J.; Hao, X.; Lau, Y.C.; Wong, J.Y.; Guan, Y.; Tan, X.; et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **2020**, *26*, 672–675. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.