



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results

Vladimir Despotovic^{a,*}, Muhannad Ismael^b, Maël Cornil^b, Roderick Mc Call^b, Guy Fagherazzi^c

^a University of Luxembourg, Department of Computer Science, Esch-sur-Alzette, Luxembourg

^b Luxembourg Institute of Science and Technology, IT for Innovation in Services Department, Esch-sur-Alzette, Luxembourg

^c Luxembourg Institute of Health, Department of Population Health, Deep Digital Phenotyping Research Unit, Strassen, Luxembourg

ARTICLE INFO

Keywords:

Voice
Cough
COVID-19
Digital biomarker
Artificial intelligence

ABSTRACT

COVID-19 heavily affects breathing and voice and causes symptoms that make patients' voices distinctive, creating recognizable audio signatures. Initial studies have already suggested the potential of using voice as a screening solution. In this article we present a dataset of voice, cough and breathing audio recordings collected from individuals infected by SARS-CoV-2 virus, as well as non-infected subjects via large scale crowdsourced campaign. We describe preliminary results for detection of COVID-19 from cough patterns using standard acoustic features sets, wavelet scattering features and deep audio embeddings extracted from low-level feature representations (VGGish and OpenL3). Our models achieve accuracy of 88.52%, sensitivity of 88.75% and specificity of 90.87%, confirming the applicability of audio signatures to identify COVID-19 symptoms. We furthermore provide an in-depth analysis of the most informative acoustic features and try to elucidate the mechanisms that alter the acoustic characteristics of coughs of people with COVID-19.

1. Introduction

COVID-19 caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has been declared as a global pandemic by the World Health Organization (WHO), and has rapidly spread over more than 200 countries worldwide [1]. The main symptoms include fever, dry cough, sore throat, dyspnea, fatigue, headache and in severe cases multiple organ failure [2–4]. Implicitly, voice is also affected resulting in lack of energy to produce sound and loss of voice caused by shortness of breath and upper airway congestion. Recurrent dry coughs can further influence changes in vocal cords affecting voice quality. Recent study has reported changes in the acoustic parameters of voice caused by the insufficient airflow through the vocal tract as a consequence of pulmonary and laryngological involvements in people with COVID-19 [5]. Therefore, all these respiratory conditions caused by COVID-19 can make patients' voice distinctive, creating identifiable voice signatures.

Voice has already proved to be a potent digital biomarker for early detection and monitoring the disease progress of various medical conditions, with the most prominent examples being the neurological disorders such as Parkinson's disease [6,7], Mild Cognitive Impairment and Alzheimer's disease [8,9], Multiple Sclerosis [10] and Amyotrophic Lateral Sclerosis [11,12]. Other conditions that affect voice include

Rheumatoid Arthritis [13,14], which may lead to voice hoarseness due to cricoarytenoid joint involvement, or Diabetes Mellitus that provokes vocal fatigue caused by decreased laryngeal muscle strength in the presence of neuropathy [15]. For an exhaustive list of medical conditions that may invoke voice disorders the reader is referred to Ref. [16].

With the onset of COVID-19, various efforts were made to develop efficient solutions for automatic diagnosis that would supplement the standard testing methods. While the Real-time Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR) test serves as the gold standard for COVID-19 detection, it requires an in-person visit to hospital or laboratory for taking an upper respiratory specimen (nasopharyngeal and oropharyngeal swab). Local clinics may also lack RT-qPCR facilities and skilled staff, requiring transport of specimens and further delaying test results. The Rapid Antigen Test (RAT) is an alternative that does not require laboratory processing and alleviates the time constraint of RT-qPCR, but unfortunately its sensitivity decreases with lower viral loads, thus providing false negative results in people with lower levels of the SARS-CoV-2 virus [17]. Chest Computed Tomography (CT) may complement RT-qPCR to compensate false-negative tests for COVID-19 suspected patients with pneumonia [18]. An automatic system for discrimination between COVID-19 pneumonias, non-COVID-19 pneumonias and controls based on deep convolutional

* Corresponding author.

E-mail address: vladimir.despotovic@uni.lu (V. Despotovic).

neural networks (Deep COVID DeteCT) that uses 3D model of entire CT chest is proposed in Ref. [19].

Development of remote diagnosis solutions for COVID-19 detection and disease progress monitoring would enable minimizing physical contact between medical staff and patients and avoiding their exposure to SARS-CoV-2; however, this research has still not reached full maturity. Attempts are made to create low-cost wearable sensing technologies, that would allow remote monitoring of physiological signals and biochemical markers, for either early detection of COVID-19 cases or tracking the recovery process during self-isolation at home [20]. Another promising technology relies on connection between the sounds produced by patients (voice, cough or breathing) and respiratory system disorders caused by COVID-19. Analyzing coughs has been successfully used for detection of pneumonia, asthma, bronchiolitis and croup [21], pertussis [22], chronic obstructive pulmonary disease [23] and tuberculosis [24].

Several studies have already explored the usability of voice, cough and breathing for detection and screening of COVID-19 [25–31]. Crowdsourced dataset of cough and breathing samples is collected and used for distinguishing between individuals tested positive and negative to COVID-19, as well as participants diagnosed with asthma [25]. Handcrafted acoustic features and features automatically extracted using the VGGish model were used in combination with a simple logistic regression classifier. Similar feature extraction approach was followed in Ref. [28] in combination with Recurrent Neural Networks for the classification task. Five-level empirical mode decomposition and 5-level discrete wavelet transform, as well as deep features extracted from scalogram images using ResNet50 and MobileNet networks, were extracted from user coughs and classified into positive and negative to COVID-19 using the Support Vector Machines classifier in Ref. [29]. Large-scale crowdsourced cough dataset externally validated and labeled by expert physicians was collected within the COUGHVID study. The labels include a diagnosis, severity level, and existence of audible anomalies in cough sounds, such as dyspnea, wheezing or nasal congestion [30]. Vocal biomarkers initially designed for detection of Alzheimer's disease were successfully used for identification of COVID-19 from forced cough recordings, showing the ability to almost perfectly detect even the asymptomatic cases [26]. This is in contrast to results given in Ref. [31], where the substantial performance decrease is observed for distinguishing asymptomatic positive cases from healthy participants. Although the preliminary results for identification of COVID-19 from voice and respiratory sounds are promising, pointing out to the relevance of using audio signatures to detect COVID-19 symptoms, the results are still inconclusive and further efforts are required to reach the maturity and confirm the effectiveness of the proposed models over different datasets and a variety of voice sounds.

We add to existing efforts of the community by collecting a dataset of speech, cough and breathing samples of both people diagnosed positive to COVID-19 and non-infected individuals via large scale public involvement. While cough and breathing are language independent, the collected speech data is multilingual in 8 different languages (English, German, French, Spanish, Portuguese, Arabic, Luxembourgish and Serbian) and represents, to the best of our knowledge, the first dataset that takes into account cross-linguistic variations. It can be used to assess to what extent the language a person speaks might affect the performance of a speech-based COVID-19 predictive model. Recent study has shown that despite the universality of speech motor system involved in speech production there are language-specific differences in aspects such as phonation or prosody that influence the perception of speech impairment in Parkinson's disease [32], that are also reported in some cases of people with COVID-19 [33,34].

While cross-linguistic analysis of speech stands as a mid-to long-term objective, we report in this paper preliminary results on detection of COVID-19 from cough patterns, since dry cough was identified as one of the most frequent symptoms in our dataset. In contrast to other related research which mostly tries to show that a particular machine learning

model is appropriate for a COVID-19 detection task, we go a step further and provide an in-depth analysis of the most informative acoustic features, trying to elucidate the exact mechanisms that alter the acoustic characteristics of coughs of people with COVID-19. Furthermore, we show that wavelet scattering transform is a very promising feature extraction method that is robust to noise present in the data, but also able to learn features from the limited data resources. This work comprises a step toward the development of low-cost and easy-to-use computer aided tools for the automatic assessment of respiratory symptoms related to COVID-19 and remote patient monitoring during the recovery process.

Section 2 gives a detailed description of the collected dataset, data preprocessing and feature extraction techniques, as well as classification models for COVID-19 detection used in this paper. Section 3 provides experimental results and discusses the efficiency of various audio feature extraction and modelling approaches, while section 4 gives concluding remarks and directions for further research.

2. Material and methods

2.1. Dataset collection via CDCVA study

The ability to successfully identify COVID-19 patients from their voices heavily depends on collection of a large dataset that contains speech, coughs and breathing of people diagnosed positive to COVID-19, non-infected individuals, but also people who might suffer from other respiratory conditions. Although there is no standard protocol for data collection, we propose within CDCVA¹ (COVID-19 Detection by Cough and Voice Analysis) study to collect the sounds emitted from human's mouth using five vocal tasks: sustained phonation of a vowel/aaaa/, coughing (3 times), breathing deeply in and out through mouth (3 times), number counting from 1 to 20, and reading a specified text. Counting and reading task are provided in a language of the participant's choice. The data is collected using a multilingual web-based platform which is designed in 8 languages (English, German, French, Spanish, Portuguese, Arabic, Luxembourgish and Serbian) to cover a large range of potential participants, but also the most used languages in the greater region of Luxembourg. Besides vocal data, information about the participant's age, gender, country of residence, native language, weight, height, smoking and drinking habits is also collected, as well as COVID-19 related symptoms and comorbidities. Health status of the participant (positive or negative to COVID-19) is determined based on the self-declaration confirmed by the standard RT-qPCR or RAT test, with the date of testing. The participants can also provide the information whether they are currently in home isolation or hospitalized.

As the data collected is highly personal, extensive discussions were undertaken on the ethical approach of the project. The participants involved are asked to give their informed consent. Information relating to data protection issues is also provided. From a design point of view, the user interface was developed to be as simple to use as possible and to work across a range of browsers and operating systems.

The dataset is entirely crowdsourced; therefore, it is uncontrolled and we have to rely on the participants in our study to infer ground truth labels. Another challenge is related to the use of different devices, microphones, web browsers and recording conditions for data collection, resulting in data stored in different formats and with different quality of recordings. Hence, preprocessing techniques have to be applied to unify the recording samples prior to further use.

After removing items with either missing audio files or demographic data (therefore labels could not be retrieved), a total of 1103 participants were identified in the study, out of which 92.38% declared as negative to COVID-19 (1019) and 7.62% as positive (84), as shown in Table 1. Approximately half of participants are in the age range between

¹ <https://cdcva.list.lu/>.

Table 1
Statistics of the CDCVA dataset.

	Positive	Negative	Total
Participants	84 (7.62%)	1019 (92.38%)	1103 (100%)
Gender			
Male	38 (45.24%)	580 (56.92%)	618 (56.03%)
Female	46 (54.76%)	437 (42.88%)	483 (43.79%)
Other	0 (0%)	2 (0.20%)	2 (0.18%)
Age			
≤20	3 (3.57%)	62 (6.08%)	65 (5.89%)
21–30	16 (19.05%)	180 (17.66%)	196 (17.77%)
31–40	28 (33.33%)	260 (25.51%)	288 (26.11%)
41–50	20 (23.81%)	243 (23.85%)	263 (23.85%)
51–60	12 (14.29%)	174 (17.08%)	186 (16.86%)
61–65	3 (3.57%)	51 (5.01%)	54 (4.90%)
>65	2 (2.38%)	49 (4.81%)	51 (4.62%)
Country			
Luxembourg	50 (59.52%)	415 (40.73%)	465 (42.16%)
Serbia	20 (23.81%)	371 (36.41%)	391 (35.45%)
France	3 (3.57%)	42 (4.12%)	45 (4.08%)
Germany	1 (1.19%)	32 (3.14%)	33 (2.99%)
Other	10 (11.91%)	159 (15.60%)	169 (15.32%)
Language			
English	25 (29.76%)	400 (39.49%)	425 (38.74%)
French	24 (28.57%)	206 (20.34%)	230 (20.97%)
Serbian	8 (9.53%)	208 (20.53%)	216 (19.69%)
German	10 (11.90%)	97 (9.57%)	107 (9.75%)
Luxembourgish	11 (13.10%)	96 (9.48%)	107 (9.75%)
Other	6 (7.14%)	6 (0.59%)	12 (1.10%)

31 and 50 years, with more male participants (56.03%). A large majority of participants originate from Luxembourg and Serbia (77.61%), as a consequence of the media campaign that was done in these two countries. The choice of the survey language reflects the international and multilingual structure of population in Luxembourg, where most of the participants in the study come from, but is partly influenced by the fact that the default language of the web platform is English. Hence, English was a dominant choice with 38.74% of participants, followed by French (20.97%) and Serbian (19.69%). Note that for 6 participants in the study it was not possible to determine the language, since they provided only voice samples for language independent tasks (coughing, breathing and/or sustained vowel phonation).

Distribution of symptoms across participants declared as positive to COVID-19 in the study reveals that loss of smell was the most prevalent

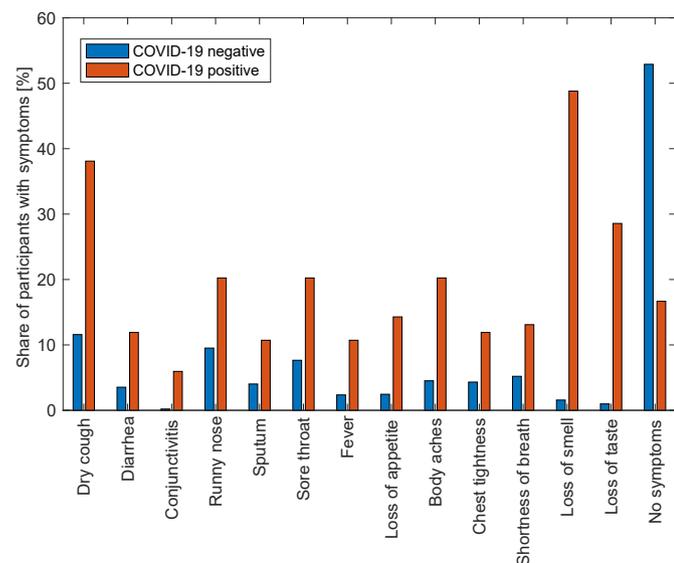


Fig. 1. Distribution of symptoms across participants in CDCVA dataset.

symptom (48.81%), followed by dry cough (38.1%) and loss of taste (28.57%), as shown in Fig. 1. Approximately 17% of COVID-19 positive participants were asymptomatic and 5% were hospitalized. Majority of COVID-19 negative participants did not report any symptoms (58%), as expected. Interestingly, 16 participants negative to COVID-19 reported loss of smell and 10 reported loss of taste, which are typical COVID-19 symptoms rarely observed in general population. Taking into account that 32 non-COVID-19 participants also reported being in home isolation, suggests that some of the participants were very likely to be positive to COVID-19 according to symptoms, but were probably not tested or were falsely negative, leading to incorrect labeling in our dataset. These participants are excluded from further experiments. Distribution of comorbidities that might also affect voice is similar in both populations and not significant to affect overall results (see Fig. 2).

2.2. Data preprocessing

Preprocessing is performed to account for the potential missing, incomplete or noisy data in the dataset. Various problems were observed in the dataset, such as missing data instances for particular vocal tasks, substitution of vocal tasks (e.g. coughing recorded instead of breathing), or incomplete and missing demographic data.

All audio data instances were converted to WAV audio format with 44.1 kHz sampling rate and 32-bit floating point bit-depth. Stereo recordings were converted to mono before further processing. Leading and trailing silences were removed from all audio recordings. Since cough audio recordings consisted of 2–5 single coughs (the participants were instructed to cough 3 times), all cough recordings were segmented to individual cough signals, keeping the track about the participant that produced each of the individual coughs.

Some recordings contained an ambient noise (e.g. chatting, laughing, TV, traffic noise in the background etc.); hence, noise reduction was applied prior to further signal processing. Audacity v2.4.2 was used to remove the background noise using the spectral noise gating. Quiet sound segment with background ambient noise was preselected to estimate the spectrum of pure tones that make up the background noise, to create a fingerprint of the background noise in the audio file. When applied to the entire audio signal, the noise reduction algorithm reduces all pure tones that are not sufficiently louder than their average levels in

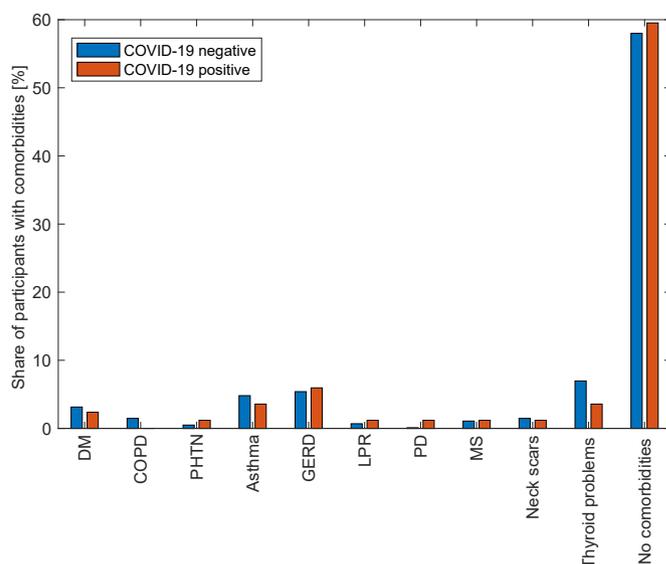


Fig. 2. Distribution of comorbidities across participants in CDCVA dataset. DM: Diabetes Mellitus; COPD: Chronic obstructive pulmonary disease; PHTN: Pulmonary hypertension; GERD: Gastroesophageal reflux disease; LPR: Laryngopharyngeal reflux; PD: Parkinson’s disease; MS: Multiple sclerosis; Neck scars: Scars from neck surgery or from trauma to the front of the neck.

the fingerprint; thus, preserving the useful signal (e.g. voice or cough) and minimizing noise. An example of the waveform and the spectrogram of the cough signal before and after the noise reduction is provided in Fig. 3. The clicking noise present in Fig. 3a is mostly suppressed in both time and frequency directions, as shown in Fig. 3b. Note that noise reduction was applied before segmentation and silence removal.

2.3. Feature extraction and feature selection

Instead of using handcrafted acoustic features as in Ref. [25] or Mel frequency Cepstral Coefficients (MFCCs) as in Ref. [27], we opt to experiment with standard acoustic feature sets, such as the Geneva Minimalistic Acoustic Parameter Set (GeMaps), extended Geneva Minimalistic Acoustic Parameter Set (eGeMaps) and ComParE feature set, which are used as baseline feature sets for various acoustic tasks. They extract a large number of potentially useful acoustic features, and we further apply a data driven approach to automatically derive the relevant features and their correlations to patient's COVID-19 status. All used feature sets comprise of supra-segmental features, meaning that the acoustic low-level descriptors (LLDs) are summarized over variable length audio segments by applying statistical functionals (e.g. mean, standard deviation, skewness, kurtosis, quartiles, etc.) to obtain a feature vector of constant length [35].

GeMaps feature set is a minimalistic set of voice parameters selected according to their potential to address physiological changes during voice production. It contains 62 acoustic features including 18 energy/amplitude, frequency and spectral parameters and 6 temporal features, as well as statistical functionals applied to these parameters (arithmetic mean, standard deviation, percentile etc.) [36]. eGeMaps is an extended GeMaps feature set which contains 26 additional acoustic parameters, leading to a set of 88 acoustic features [36]. Finally, ComParE is a brute force acoustic feature set composed of 65 LLDs and various statistical functionals applied to these LLDs, leading to a total of 6737 features extracted for each of the variable length data instances. This feature set is used as a part of Computational Paralinguistics Challenge that takes place each year at the INTERSPEECH conference [37]. All features are extracted using openSMILE v3.0 software [38].

We furthermore experiment with the wavelet scattering features which are used to extract low-variance representations from audio signal by applying a wavelet scattering transform, i.e. a series of wavelet decompositions and modulus operators. The resulting structure is similar to a convolutional neural network, but does not require training. The obtained time-averaged signal representations are invariant over large time scales, much longer than in MFCCs. Wavelet scattering features provide state-of-the-art results in musical genre classification and phone segment classification tasks [39]. We define the wavelet

scattering network with two wavelet filter banks: the first one having 8 wavelets per octave and the second one having 1 wavelet per octave, whereas the invariant time scale is 0.5 s. Each audio sample is divided into a number of scattering time windows across 411 paths in the scattering transform, leading to a total 411 features extracted per each time window. After the acoustic or wavelet scattering features are extracted, only a subset of most relevant features are used for the classification task based on the mutual information criterion, which measures the dependency between two random variables. It is equal to zero if the variables are independent, whereas higher values denote the higher dependency [40]. We sort the features in descending order according to the mutual information and use only the subset of features that have the largest influence on classification task. All features were standardized before further use, i.e. the mean was removed and they were scaled to unit variance.

2.4. Models for detection of COVID-19 from coughs

Features extracted in Section 2.3 are further fed into three ensemble models: random forests, boosted and bagged decision trees. Random forests classifier with 250 trees was used to perform classification into participants positive and negative to COVID-19. Fully grown and unpruned trees are used with Gini index as the criterion to split the node. The features are randomly permuted at each split. Bootstrap aggregation or bagging model is configured with 250 decision trees with the maximum depth equal to 4. Finally, AdaBoost.M1 algorithm was used for adaptive boosting with 250 decision stumps, i.e. one-level trees with two leaf nodes.

Beside the ensemble models, Multi-Layer Perceptron (MLP) is applied for classification using the same standard acoustic feature sets and wavelet scattering features. MLP architecture consists of an input layer with the number of nodes that corresponds to the number of selected features, three hidden layers with 128, 64 and 16 neurons respectively, and an output layer with two neurons that correspond to the number of classes to perform the classification task. Softmax activation function is used in the output layer, while the Rectified Linear Unit (ReLU) activation function, which allows to reduce the likelihood of vanishing gradient, is used in the hidden network layers. The Adam optimization algorithm is applied to update the network weights iteratively during the training, with cross-entropy as an objective function.

Beside features presented in Section 2.3, we also extract deep audio embeddings using VGGish and OpenL3 models. VGGish feature extractor extracts 128-dimensional audio embeddings directly from non-overlapping 960 ms frames of audio signal. Each frame is further divided into 25 ms windows with 10 ms overlap and short-time Fourier transform is computed to obtain the spectrograms integrated in 64 mel-

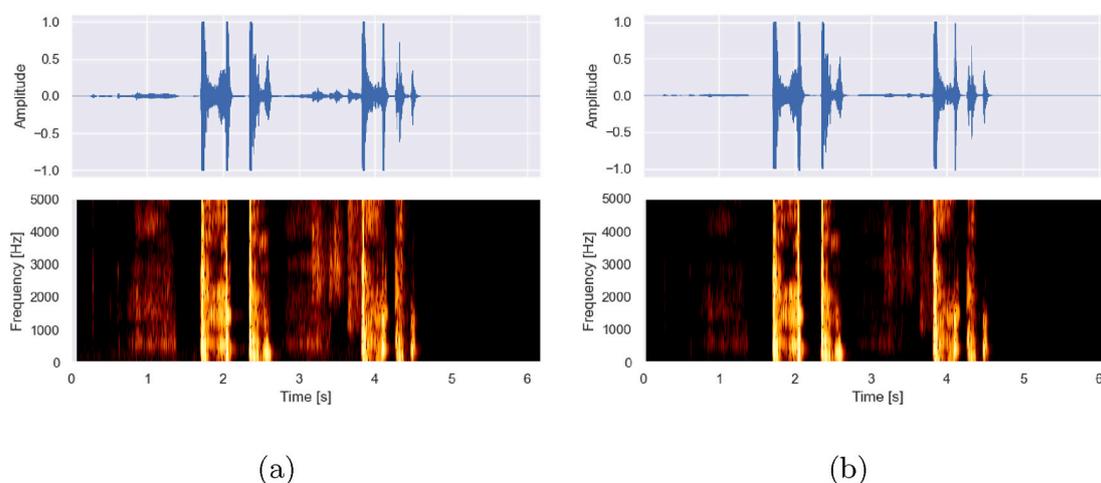


Fig. 3. The waveforms and the spectrograms of the cough signal a) before and b) after noise reduction.

spaced frequency bins, resulting in 96×64 bins log-mel spectrograms [41]. The input audio signals are resampled to 16 kHz to match the requirements of the VGGish network. VGGish network is composed of a series of convolutional, ReLU activation layers and max pooling layers, followed by three fully connected layers. The network is pretrained on the Audio Set large-scale dataset for audio event classification which is composed of over 1.7 millions of 10 s excerpts from YouTube videos (4971 h of audio in total) classified into 632 audio events [42]. We remove the output layer from pretrained VGGish network and add instead one fully connected layer with 128 nodes, followed by a ReLU activation and a dropout layer with dropout rate equal to 0.5 to prevent the overfitting. Finally, we add an output layer with 2 nodes to perform the classification.

For extracting OpenL3 embeddings we use model pretrained on the environmental subset of AudioSet, which includes human and animal sounds, as well as other sounds in natural acoustic environments. The model uses a Mel-spectrogram time-frequency representation with 256 Mel bands and extracts a 6144-dimensional audio embedding per each signal frame [43]. Each cough file is segmented into 1 s overlapping frames with a hop size of 0.5 s, leading to multiple embeddings for each cough file. Majority voting is used to determine the label of a single cough file, based on the labels of its corresponding frames. Classification is performed using MLP with two hidden layers with sizes 512 and 128 nodes respectively, each followed by a dropout layer with dropout rate equal to 0.5, and finally an output layer with 2 nodes.

2.5. Hyperparameter optimization

Grid search was used for tuning the hyperparameters in the proposed classification models, as shown in Table 2. For tuning the bagging classifier the number of base estimators (individual trees) and the maximal number of samples to train each base estimator, given as the share of total number of samples, were used as the hyperparameters. In addition to this two hyperparameters, the maximal number of features to consider for each split, given as the share of total number of features, is used for random forests. The number of base estimators was the only hyperparameter optimized in case of the boosting classifier.

For training OpenL3 and VGGish networks, the number of fully connected hidden layers on top feature extraction layers, the number of nodes in the fully connected layers, the mini-batch size, as well as the dropout rate were tuned using the grid search.

Concerning MLP, the optimal number of hidden layers, as well as the optimal number of nodes in the hidden layers are evaluated using the grid search.

3. Results and discussion

3.1. Experimental setup

We perform preliminary experiments for detection of COVID-19 using cough signals only. Given the highly imbalanced dataset that we collected, we create a balanced subset that includes coughs of all participants positive to COVID-19 and the same number of non-infected participants matched by age and gender, without the medical history of conditions that affect voice, and without symptoms to COVID-19.

Table 2
Hyperparameter optimization for the classifiers.

Hyperparameter	Classifier	Range
# base estimators	Bagging, Boosting, RF	[100–500] in steps of 50
Max # samples	Bagging, RF	[0.6 – 1.0] in steps of 0.1
Max # features	RF	[0.6 – 1.0] in steps of 0.1
# hidden layers	OpenL3, VGGish, MLP	1–3
# nodes per layer	OpenL3, VGGish, MLP	16, 64, 128, 256, 512, 1024
Dropout rate	OpenL3, VGGish	[0 – 0.5] in steps of 0.1
Mini-batch size	OpenL3, VGGish	16, 32, 64, 128

Note that the balanced subset contains 82 participants positive to COVID-19, out of 84 in the initial dataset, since cough signals were missing for two users. Participants positive to COVID-19 were between 14 and 69 years (mean 40.11, standard deviation 12.21), whereas non-infected participants were between 15 and 68 years (mean 40.26, standard deviation 12.1). To make the experiment more realistic, smoking was not taken as an exclusion criterion from the balanced subset. Prevalence of smoking was higher in control group (28.05%) than in the COVID-19 positive group (17.07%), as shown in Table 3.

Although we instructed the users to cough three times, the number of single coughs per audio sample varied from 2 to 5. Therefore, we decided to segment the balanced subset of coughs, and analyse instead the individual cough signals, leading to a dataset that was composed of 496 cough signals in total, 249 of which were produced by COVID-19 positive participants and 247 by non-infected subjects.

We provide data visualization for the segmented subset of coughs using t-distributed Stochastic Neighbor Embedding (t-SNE) [44]. t-SNE generates a reduced feature space where the similar features are represented by points close in space and vice versa. Fig. 4 describes an example of two-dimensional graph for selected features extracted using the ComParE features set, as the most comprehensive set of acoustic features. The graph could be split into two regions: 1) Top-left highlighted with a red rectangle, where the subjects diagnosed positive to COVID-19 and described by red points are mostly present; 2) Non-right highlighted with blue rectangle for non-infected individuals. Although both regions contain overlapping data points, data distribution suggests that it is possible to discriminate between two groups of participants in our study.

Given a limited dataset size, 5-fold stratified cross-validation is used for splitting the data into five subsets (folds), four being used for training, and the remaining one for testing. Each class is approximately equally distributed across folds, preserving the class distribution of the original dataset. The procedure is repeated 5 times to obtain performance estimates for each of the test folds, and then averaged over all folds. Since the dataset contains multiple cough signals from each of the participants, splitting into folds was done with respect to the participants, meaning that the coughs of the same speaker cannot appear in both training and test subset.

The model performance was assessed using accuracy, sensitivity and specificity as performance measures. Accuracy is the ratio of the number of correctly classified data instances and the total number of data instances. Sensitivity represents the ability of the model to correctly detect COVID-19 positive subjects, whereas the specificity is the model ability to correctly identify non-infected participants.

Table 3

Statistics of the balanced CDCVA subset used for detection of COVID-19 from coughs.

	Positive			Negative		
	#	Age Mean (SD)	Min- Max	#	Age Mean (SD)	Min- Max
Male	36 (43.90%)	41.19 (12.62)	14–67	36 (43.90%)	40.41 (12.26)	15–68
Female	46 (56.10%)	39.26 (11.95)	18–69	46 (56.10%)	39.43 (11.77)	18–68
Total	82 (100%)	40.11 (12.21)	14–69	82 (100%)	40.26 (12.10)	15–68
	Smoking habits			Smoking habits		
Gender	Smoking	Non-smoking		Smoking	Non-smoking	
Male	5 (13.89%)	31 (86.11%)		10 (27.78%)	26 (72.22%)	
Female	9 (19.57%)	37 (80.43%)		13 (28.26%)	33 (71.74%)	
Total	14 (17.07%)	68 (82.93%)		23 (28.05%)	59 (71.95%)	

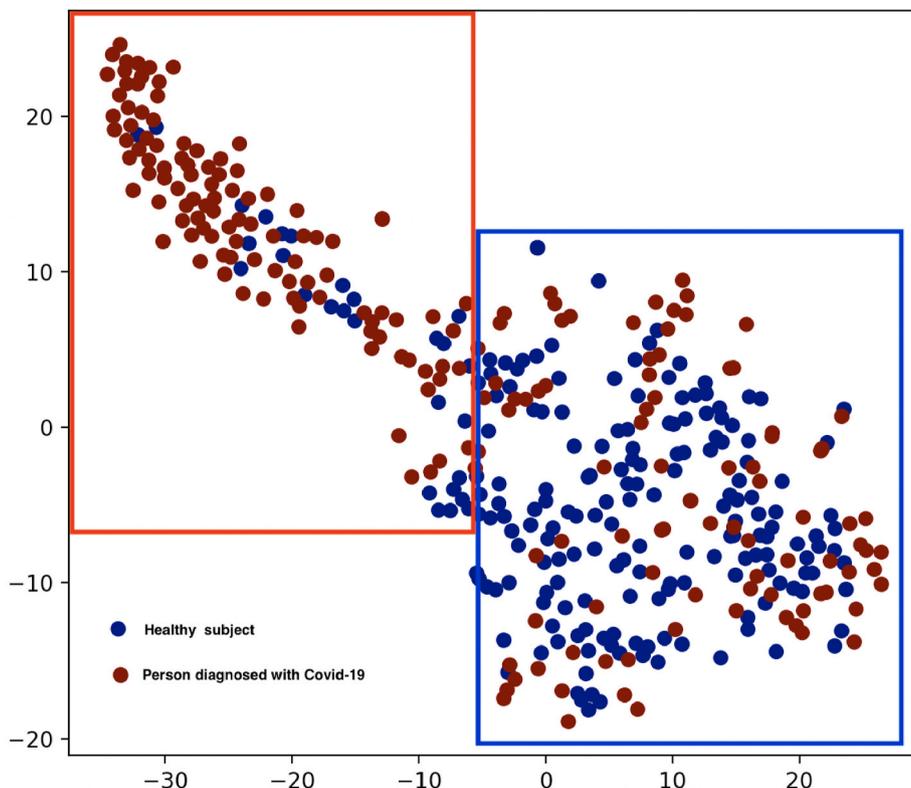


Fig. 4. Two-dimensional graph for selected features generated by t-distributed Stochastic Neighbor Embedding approach. Features are extracted using ComParE feature set and selected using mutual information criterion.

3.2. Numerical results

Performance is evaluated using three ensemble models (random forests, bagging and boosting), MLP, VGGish and OpenL3 networks. Three standard feature sets are used for feature extraction with ensemble models, i.e. GeMaps, eGeMaps and ComParE, as explained in Section 2.3. Since GeMaps and eGeMaps feature sets are relatively small no feature selection was applied. Considering the size of the ComParE feature set only 1% of the most informative features are kept based on the mutual information criterion. Obtained results provided in Table 4 reveal that although minimalistic sets of acoustic features, such as GeMaps and eGeMaps, are capable of learning intrinsic features from

Table 4 Performance evaluation using standard acoustic features sets, wavelet scattering features and audio embeddings.

		Accuracy	Sensitivity	Specificity
Random Forests	GeMaps	72.35%	70.21%	74.06%
	eGeMaps	73.32%	71.25%	74.52%
	ComParE	87.31%	87.37%	87.72%
	Wavelet	87.50%	86.94%	87.74%
Bagging	GeMaps	73.96%	65.33%	82.20%
	eGeMaps	74.16%	66.56%	81.35%
	ComParE	87.12%	83.71%	90.87%
	Wavelet	84.92%	88.75%	81.06%
Boosting	GeMaps	69.51%	68.72%	70.64%
	eGeMaps	68.65%	65.44%	71.86%
	ComParE	82.70%	78.79%	87.33%
	Wavelet	88.52%	87.19%	89.82%
MLP	GeMaps	64.79%	61.6%	67.1%
	eGeMaps	62.2%	56.8%	67.6%
	ComParE	70.0%	66.2%	74.19%
	Wavelet	80.0%	72.2%	87.6%
VGGish		76.73%	80.84%	72.86%
OpenL3		76.65%	81.89%	71.12%

coughs, substantially better results are obtained using a brute force audio feature extraction approach with ComParE features, leading to accuracy and sensitivity approximately equal to 87% for random forests, whereas specificity goes up to 90.87% in case of bagging. Random forests and bagging outperform boosting for all standard acoustic feature sets, which may be explained by the principle differences in design of these ensemble models. Unlike random forests and bagging where trees are grown independently, in boosting each subsequent model improves the performance of the previous one by learning from its misclassifications. Therefore, boosting is by nature more sensitive to outliers that might be present in our dataset due to crowdsourced data collection. MLP provides in general lower performance results than the ensemble models.

We want to furthermore contribute to the explainability of the ensemble models by trying to discover the exact mechanisms that alter the acoustic parameters of coughs in people with COVID-19. In order to do that we analyse in Fig. 5 the ten most informative features in ComParE acoustic feature set. The best indicator of COVID-19 coughs according to the mutual information criterion is the root mean square signal frame energy smoothed by a moving average filter with window length equal to 3 (pcm_RMSenergy_sma_perc1.0). Being an explosive sound, cough has bursts of energy increase in a short interval of time which are more evident in signals produced by people with COVID-19. Since dry cough is the second most prevalent symptom observed in almost 40% of participants positive to COVID-19 (see Fig. 1), higher energy levels in comparison to forced coughs of non-infected individuals are expected.

Spectral harmonicity, the second most informative feature, describes the harmonic structure of an audio signal in which the sound frequencies are integer multiples of the fundamental frequency. It is typically observed in voiced speech, but recent study has shown that 2 to 3 harmonics can be observed in forced coughs of healthy individuals, whereas no clear pattern in the harmonics' structure is visible in wet coughs [45].

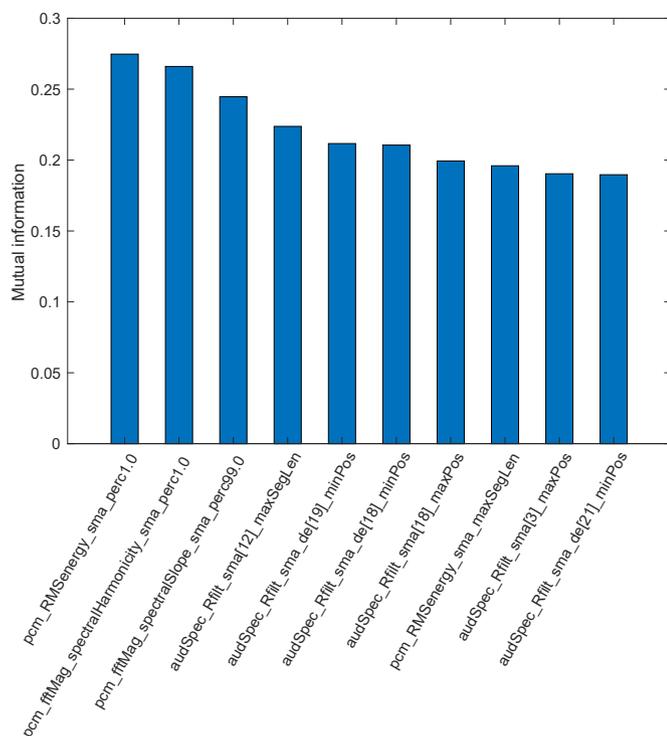


Fig. 5. The list of 10 most informative features in ComParE acoustic feature set based on the mutual information criterion.

Our study confirms that spectral harmonicity may be also used to distinguish between coughs of people positive to COVID-19 and non-infected subjects.

The cough signal has a typical physiological mechanism which starts with inspiratory phase, followed by closure of glottis and rise in intrathoracic pressure in the compression phase, and sudden glottis opening resulting in fast expiratory airflow and cough sound in the expiratory phase. Several subsequent partial glottis closures may lead to additional voiced sounds [46]. Spectral slope as the third most informative feature is closely related to the speed of glottis closure during the cough production, resulting in a steeper spectral slope for slower closure of glottis [47].

RASTA processing uses band-pass filters on time trajectories of speech feature vectors to suppress the spectral components that change more quickly or slowly than in a typical speech signal; thus, it can be used to suppress the additive noise in speech recognition or speech enhancement techniques [48]. Although RASTA features were not previously studied in cough signal processing, 6 out of 10 most informative features in our study are related to either the absolute position of the maximum or minimum value (in frames), or the maximum segment length of one of 26 RASTA filtered auditory bands, suggesting that it is a very potent indicator of coughs produced by people with COVID-19.

Results obtained with wavelet scattering features and ensemble models (accuracy 88.52%, sensitivity 88.75%, specificity 89.82%) suggest that they are very powerful data representations which are able to minimize within-class differences, emphasizing at the same time dissimilarity between coughs of people positive to COVID-19 and non-infected subjects. Although similar in structure to convolutional neural networks, wavelet scattering transform does not learn the filters during optimization, but applies them a priori, which may be beneficial when training data is limited, as in our case. Considering the nonstationary and stochastic nature of cough, wavelet transform enables decomposing the cough signal in time and frequency domain, while focusing on the local signal structures to capture efficiently frequency changes at any time moment [49]. In contrast to standard acoustic features, boosting provides slightly improved performance for wavelet scattering features

in comparison to other ensemble models, indicating that wavelet filtering is robust to noise in the data [50]. Note that the results provided in Table 4 for wavelet scattering features are obtained after applying feature selection with 10% of most informative features based on mutual information criterion.

In order to explore the influence of feature selection, the experiments are repeated with wavelet scattering features and boosting as the best performing individual model, but with different sizes of feature vectors. The results given in Table 5 show that using a large number of irrelevant features adds uncertainty to the prediction and reduces the overall model performance. Best individual results for accuracy, specificity and sensitivity are obtained using only 10% most relevant features (88.52%, 87.19% and 89.82%, respectively). Further decrease of feature size degrades the performance. Beside improving the performance, feature selection reduces also the computational costs of the model.

We furthermore experiment with high-level audio feature embeddings extracted from cough recordings using an existing pretrained VGGish model. The idea is to transfer the knowledge from the domain where data is easier to collect by pretraining the model on a large-scale audio event classification dataset and benefit from learning the generic audio features that can be shared across domains. The model is trained using Adam optimizer with mini-batch size of 64, L2 regularisation parameter equal to 0.005 and adaptive learning rate which drops by a factor of 10 if no improvement is observed after 3 iterations (initial learning rate was equal to 0.001). The model converges after 6 epochs and achieves accuracy equal to 76.73%, sensitivity 80.84% and specificity 72.86%, showing the ability to learn reasonable feature representations directly from audio signals and avoiding the need for hand-crafted feature extraction. However, the model is still not able to reach state-of-the-art performance achieved with wavelet scattering or ComParE features. A reason for lower performance might be the fact that due to the requirements of the VGGish model, the signal was down-sampled to 16 kHz, whereas the sampling frequency for the wavelet scattering features and the standard acoustic features sets was 44.1 kHz.

Similar performance is obtained with OpenL3 cough embeddings (accuracy 76.65%, sensitivity 81.89%, specificity 71.12%). The model is trained for 50 epochs using Adam optimizer with mini-batch size of 64. We expect that by collecting more data and combining different voice modalities beside the cough, better audio embeddings can be learned, the impact of transferred learning will be larger, and the deep learning models will generalise better to a target dataset.

4. Conclusions

A new crowdsourced multilingual dataset of speech, cough and breathing samples of people diagnosed positive to COVID-19 and non-infected individuals is presented in this study, as well as preliminary results for detection of COVID-19 from cough signal, which point out to the relevance of using audio signatures to detect COVID-19 symptoms. In contrast to other related research which mostly tries to prove that a particular machine learning model is appropriate for a COVID-19 detection task, we go a step further and provide an in-depth analysis of the most informative acoustic features, trying to elucidate the mechanisms that alter the acoustic characteristics of coughs of people with COVID-19.

Although not intended to be a reliable diagnostic system for COVID-19 that can replace RT-qPCR or RAT tests, the proposed solution can be

Table 5

Performance evaluation using wavelet scattering features and boosting with different dimensions of feature vectors.

Feature size	1%	5%	10%	30%	50%	100%
Accuracy	84.53%	87.31%	88.52%	85.71%	86.29%	82.47%
Sensitivity	82.27%	85.08%	87.19%	81.61%	83.52%	79.29%
Specificity	86.50%	89.46%	89.82%	89.32%	88.63%	85.36%

used for monitoring of people on a very large scale, in a short period of time (e.g. at airports or at border controls). Challenges related to disambiguation with other respiratory pathologies with similar symptoms remain to be addressed. The system is still highly experimental; however, with enough training examples it can be tuned to a reasonable performance, and potentially become a valuable tool for disease screening.

Future work will include other vocal modalities collected within this study, such as breathing, speech and sustained vowel phonations, in addition to cough analysed in this paper. We expect that by combining various signals affected by COVID-19 the model performance may be further improved.

Role of the funding source

The study funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Author contributions

Vladimir Despotovic has contributed to the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, and final approval of the version to be submitted.

Muhannad Ismael has contributed to the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, and final approval of the version to be submitted.

Mael Cornil has contributed to the acquisition of data, analysis and interpretation of data, critical revision of the article for important intellectual content, and final approval of the version to be submitted.

Roderick Mc Call has contributed to the conception and design of the study, the acquisition of data, critical revision of the article for important intellectual content, and final approval of the version to be submitted.

Guy Fagherazzi has contributed to the analysis and interpretation of data, critical revision of the article for important intellectual content, and final approval of the version to be submitted.

Acknowledgements

(Muhannad Ismael, Mael Cornil, Roderick Mc Call) and Vladimir Despotovic are supported by Luxembourg Institute of Science and Technology (LIST) and University of Luxembourg (UL) respectively, as well as by Luxembourg National Research Fund (FNR) (CDCVA, Grant Number 14856868). Guy Fagherazzi is supported by the Luxembourg Institute of Health, FNR (Predi-COVID, Grant Number 14716273) and the Andre Losch Fondation. The authors declare no competing interests.

References

- [1] J. Sanders, M. Monogue, T. Jodlowski, J. Cutrell, Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review, *J. Am. Med. Assoc.* 323 (18) (2020) 1824–1836.
- [2] S. Kujawski, K. Wong, J. Collins, et al., Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States, *Nat. Med.* 26 (2020) 861–868.
- [3] D. Chang, M. Lin, L. Wei, L. Xie, G. Zhu, C.S.D. Cruz, L. Sharma, Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside Wuhan, China, *J. Am. Med. Assoc.* 323 (11) (2020) 1092–1093.
- [4] P. Sun, X. Lu, C. Xu, W. Sun, B. Pan, Understanding of COVID-19 based on current evidence, *J. Med. Virol.* 92 (6) (2020) 548–551.
- [5] M. Asiaee, A. Vahedian-azimi, S.S. Atashi, A. Keramatfar, M. Nourbakhsh, Voice quality evaluation in patients with COVID-19: an acoustic analysis, *J. Voice* (2021). In press.
- [6] D. Hemmerling, M. Wojcik-Pedziwiatr, Prediction and estimation of Parkinson's disease severity based on voice signal, *J. Voice* (2021). In press.
- [7] V. Despotovic, T. Skovranek, C. Schommer, Speech based estimation of Parkinson's disease using Gaussian processes and automatic relevance determination, *Neurocomputing* 401 (2020) 173–181.
- [8] L. Toth, I. Hoffmann, G. Gosztolya, V. Vincze, G. Szatloczki, Z. Banreti, M. Pakaski, J. Kalman, A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech, *Curr. Alzheimer Res.* 15 (2) (2018) 130–138.
- [9] A. Koenig, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. Robert, R. David, Automatic speech analysis for the assessment of patients with pre-dementia and Alzheimer's disease, *Alzheimers Dement (Amst)* 1 (1) (2015) 112–124.
- [10] J. Ruzs, B. Benova, H. Ruzickova, M. Novotny, T. Tykalova, J. Hlavnicka, T. Uher, M. Vaneckova, M. Anellova, K. Novotna, L. Kadrnokova, D. Horakova, Characteristics of motor speech phenotypes in multiple sclerosis, *Multiple Sclerosis Related Disord.* 19 (2018) 62–69.
- [11] H. Vieira, N. Costa, T. Sousa, S. Reis, L. Coelho, Voice-based classification of amyotrophic lateral sclerosis: where are we and where are we going? a systematic review, *Neurodegener. Dis.* 19 (2019) 163–170.
- [12] A. Silbergleit, A. Johnson, B. Jacobson, Acoustic analysis of voice in individuals with amyotrophic lateral sclerosis and perceptually normal vocal quality, *J. Voice* 11 (2) (1997) 222–231.
- [13] R. Speyer, I. Speyer, M. Heijnen, Prevalence and relative risk of dysphonia in rheumatoid arthritis, *J. Voice* 22 (2) (2008) 232–237.
- [14] A. Hamdan, D. Saredidine, Laryngeal manifestations of rheumatoid arthritis, *Autoimmune Dis.* 2013 (2013). Article ID 103081.
- [15] A. Hamdan, Z. Kurban, S. Azar, Prevalence of phonatory symptoms in patients with type 2 diabetes mellitus, *Acta Diabetol.* 50 (5) (2013) 731–736.
- [16] G. Fagherazzi, A. Fischer, M. Ismael, V. Despotovic, Voice for health: the use of vocal biomarkers from research to clinical practice, *Digit. Biomarkers* 5 (1) (2021) 78–88.
- [17] A. Scohy, A. Anantharajah, M. Bodeus, B. Kabamba-Mukadi, A. Verroken, H. Rodriguez-Villalobos, Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis, *J. Clin. Virol.* 129 (2020). Article ID 104455.
- [18] M. Kassem, M. Masallat, Clinical application of chest computed tomography (CT) in detection and characterization of coronavirus (Covid-19) pneumonia in adults, *J. Digit. Imag.* 34 (2) (2021) 273–283.
- [19] E. Lee, J. Zheng, E. Colak, et al., Deep COVID DeteCT: an international experience on COVID-19 lung detection and prognosis using chest CT, *npj Digit. Med.* 4 (11). (2021).
- [20] H.C. Ates, F.G. Ali K. Yetisen, C. Dincer, Wearable devices for the detection of COVID-19, *Nat. Electron.* 4 (2021) 13–14.
- [21] P. Porter, U. Abeyratne, V. Swarnkar, J. Tan, T. wan Ng, J.M. Brisbane, D. Speldewinde, J. Choveaux, R. Sharan, K. Kosasih, P. Della, 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.* 20 (2019). Article ID 81.
- [22] R.X.A. Pramono, S.A. Intiaz, E. Rodriguez-Villegas, A cough-based algorithm for automatic diagnosis of pertussis, *PLoS One* 9 (2016) 1–20.
- [23] A. Windmon, M. Minakshi, P. Bharti, S. Chellappan, M. Johansson, B.A. Jenkins, P. R. Athilingam, 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 Inf.* 23 (4) (2018) 1566–1573.
- [24] G. Botha, G. Theron, R. Warren, M. Klopfer, K. Dheda, P.V. Holden, T. Niesler, Detection of tuberculosis by automatic cough sound analysis, *Physiol. Meas.* 39 (4) (2018). Article ID 045005.
- [25] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. X. profile imageTong Xia, P. Cicuta, C. Mascolo, Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data, in: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, CA, USA, 2020*, pp. 3474–3484.
- [26] J. Laguarda, F. Hueto, B. Subirana, COVID-19 artificial intelligence diagnosis using only cough recordings, *IEEE Open J. Eng. Med. Biol.* 1 (2020) 275–281.
- [27] P. Mouawad, T. Dubnov, S. Dubnov, Robust detection of COVID-19 in cough sounds, *SN Comput. Sci.* 2 (2021). Article ID 34.
- [28] A. Hassan, I. Shahin, M.B. Alsabek, Covid-19 detection system using recurrent neural networks, in: *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 2020, pp. 1–5.
- [29] Y.E. Erdogan, A. Narin, Covid-19 detection with traditional and deep features on cough acoustic signals, *Comput. Biol. Med.* 136 (2021) 104765.
- [30] L. Orlandic, T. Teijeiro, D. Atienza, The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms, *Sci Data* 8 (2021) 156.
- [31] J. Han, C. Brown, J. Chauhan, A. Grammenos, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, C. Mascolo, Exploring automatic covid-19 diagnosis via voice and symptoms from crowdsourced data, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021)*, 2021, pp. 8328–8332.
- [32] S. Pinto, A. Chan, I. Guimaraes, R. Rothe-Neves, J. Sadat, A cross-linguistic perspective to the study of dysarthria in Parkinson's disease, *J. Phonetics* 64 (2017) 156–167.
- [33] M.S. Cotelli, M. Cotelli, F. Manelli, G. Bonetti, R. Rao, A. Padovani, B. Borroni, Effortful speech with distortion of prosody following sars-cov-2 infection, *Neurol. Sci.* 41 (12) (2020) 3767–3768.
- [34] R. Bhagat, S. Narayanan, B. Karki, W. Liu, K. Rimmel, A case of isolated dysarthria in a covid-19 infected stroke patient: a nondisabling neurological symptom with grave prognosis, *Cureus* 12 (8) (2020). Article ID e9921.
- [35] F. Eyben, A. Batliner, B. Schuller, Towards a standard set of acoustic features for the processing of emotion in speech, in: *Proceedings of the 159th Meeting Acoustical Society of America/NOISE-CON 2010, Baltimore, Maryland, USA, 2010*, pp. 1–12.
- [36] F. Eyben, K. Scherer, B. Schuller, J. Sundberg, E. Andre, C. Busso, L. Devillers, J. Epps, P. Laukka, S. Narayanan, K. Truong, The Geneva minimalistic acoustic

- parameter set (GeMaps) for voice research and affective computing, *IEEE Trans. Affective Comput.* 7 (2) (2016) 190–202.
- [37] B.W. Schuller, A. Batliner, C. Bergler, F.B. Pokorny, J. Krajewski, M. Cychosz, R. Vollmann, S.-D. Roelen, S. Schnieder, E. Bergelson, A. Cristia11, A. Seidl1, A. S. Warlaumont, L. Yankowitz, E. Noeth, S. Amiriparian, S. Hantke, M. Schmitt, The interspeech 2019 computational paralinguistics challenge: styrian dialects, continuous sleepiness, baby sounds & orca activity, in: *INTERSPEECH 2019*, Graz, Austria, 2019, pp. 2378–2382.
- [38] F. Eyben, M. Woellmer, B. Schuller, Opensmile - the munich versatile and fast open-source audio feature extractor, in: *ACM Multimedia Conference*, Florence, Italy, 2010, pp. 1459–1462.
- [39] J. Anden, S. Mallat, Deep scattering spectrum, *IEEE Trans. Signal Process.* 62 (16) (2014) 4114–4128.
- [40] A. Kraskov, H. Stoegebauer, P. Grassberger, Estimating mutual information, *Phys. Rev. E, Statist. Nonlinear Soft Matter Phys.* 69 (2004). Article ID 066138.
- [41] S. Hershey, S. Chaudhuri, D.P. Ellis, J.F. Gemmeke, A. Jansen, R.C. Moore, M. Plakal, D. Platt, R.A. Saurous, B. Seybold, M. Slaney, R.J. Weiss, K. Wilson, CNN architectures for large-scale audio classification, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017, pp. 776–780.
- [42] J.F. Gemmeke, D.P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R.C. Moore, M. Plakal, M. Ritter, Audio Set: an ontology and human-labeled dataset for audio events, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017, pp. 776–780.
- [43] J. Cramer, H. Wu, J. Salamon, J. Bello, Look, listen, and learn more: design choices for deep audio embeddings, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 3852–3856.
- [44] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (86) (2008) 2579–2605.
- [45] V. Singh, J. Rohith, V. Mittal, Preliminary analysis of cough sounds, in: *2015 Annual IEEE India Conference (INDICON)*, New Delhi, India, 2015, pp. 1–6.
- [46] M. Polverino, F. Polverino, M. Fasolino, F. Ando, A. Alfieri, F. D. Blasio, Anatomy and neuro-pathophysiology of the cough reflex arc, *Multidiscip. Respir. Med.* 7 (5).
- [47] S. A. Memon, Acoustic Correlates of the Voice Qualifiers: A Survey, *ArXiv abs/2010.15869*.
- [48] H. Hermansky, N. Morgan, RASTA processing of speech, *IEEE Trans. Speech Audio Process.* 2 (4) (1994) 578–589.
- [49] K. Kosasih, U. Abeyratne, V.S. V, R. Triasih, Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 62 (4) (2015) 1185–1194.
- [50] B. Zhou, A. J. Ruggles, E. Huang, J. H. Frank, Wavelet-based algorithm for correction of beam-steering artefacts in turbulent flow imaging at elevated pressures, *Exp. Fluid* 60 (136).