The Acoustic Dissection of Cough: Diving Into Machine Listening-based COVID-19 Analysis and Detection

*,†Zhao Ren, ‡Yi Chang, *,§,||Katrin D. Bartl-Pokorny, *,§,||Florian B. Pokorny, and *,‡Björn W. Schuller, *Augsburg, and †Hannover, Germany, ‡London, United Kingdom, and §||Graz, Austria

SUMMARY: Objectives. The coronavirus disease 2019 (COVID-19) has caused a crisis worldwide. Amounts of efforts have been made to prevent and control COVID-19′s transmission, from early screenings to vaccinations and treatments. Recently, due to the spring up of many automatic disease recognition applications based on machine listening techniques, it would be fast and cheap to detect COVID-19 from recordings of cough, a key symptom of COVID-19. To date, knowledge of the acoustic characteristics of COVID-19 cough sounds is limited but would be essential for structuring effective and robust machine learning models. The present study aims to explore acoustic features for distinguishing COVID-19 positive individuals from COVID-19 negative ones based on their cough sounds.

Methods. By applying conventional inferential statistics, we analyze the acoustic correlates of COVID-19 cough sounds based on the COMPARE feature set, i.e., a standardized set of 6,373 acoustic higher-level features. Furthermore, we train automatic COVID-19 detection models with machine learning methods and explore the latent features by evaluating the contribution of all features to the COVID-19 status predictions.

Results. The experimental results demonstrate that a set of acoustic parameters of cough sounds, e.g., statistical functionals of the root mean square energy and Mel-frequency cepstral coefficients, bear essential acoustic information in terms of effect sizes for the differentiation between COVID-19 positive and COVID-19 negative cough samples. Our general automatic COVID-19 detection model performs significantly above chance level, i.e., at an unweighted average recall (UAR) of 0.632, on a data set consisting of 1,411 cough samples (COVID-19 positive/negative: 210/1,201).

Conclusions. Based on the acoustic correlates analysis on the COMPARE feature set and the feature analysis in the effective COVID-19 detection approach, we find that several acoustic features that show higher effects in conventional group difference testing are also higher weighted in the machine learning models.

Key words: Acoustics−Automatic disease detection−Computational paralinguistics−Cough−COVID-19.

INTRODUCTION

A novel coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused a disease that quickly spread worldwide at the end of 2019 and the beginning of 2020. In February 2020, the World Health Organization (WHO) named the disease COVID-19 and shortly after that declared the COVID-19 outbreak a global pandemic. Globally, as of February 2022, more than 434,150,000 confirmed cases of COVID-19, including more than 5,940,000 deaths were reported to the WHO.[1]

Both the presenting symptoms and the symptom severity vary considerably from patient to patient, ranging from asymptomatic infections or a mild flu-like clinical picture to severe illness or even death. Commonly reported symptoms of COVID-19 include (1) respiratory and ear-nose-throat symptoms such as cough, shortness of breath, sore throat and headache, (2) systemic symptoms such as fever, muscle pain, and weakness, as well as (3) loss of smell and/or taste.[2] Less common ear-nose-throat symptoms associated with COVID-19 are pharyngeal erythema, nasal congestion, tonsil enlargement, rhinorrhea, and upper respiratory tract infection.[3]

Diagnostic approaches

The early detection of a COVID-19 infection in a patient is essential to prevent the transmission of the virus to other hosts and provide the patient with appropriate and early treatment. A series of laboratory diagnosis instruments have been proposed to test for COVID-19, e.g., computed tomography (CT), real-time reverse transcription polymerase chain reaction (rRT-PCR) tests, and serological methods.[4,5] CT and X-ray detect COVID-19 based on chest images.[6-9] An rRT-PCR test focuses on analyzing the virus' ribonucleic acid (RNA) and synthesized complementary deoxyribonucleic acid (cDNA) from a nasopharyngeal swab and/or an oropharyngeal swab.[10] Serological instruments measure antibody responses to the corresponding infection and confirm the COVID-19 status.[5] However, the instruments mentioned above are costly and/or not always

available, since they can only be conducted by professionals and require special equipment and a certain time of analysis. Even though rapid antigen and molecular tests are more and more used by non-professionals/the test person him- or herself to quickly detect COVID-19, e.g. in everyday life settings, they result in a huge amount of waste due to the testing kits as well as their packing. Thus, it is essential to develop low-cost, real-time, easy-to-apply, and eco-friendly screening instruments that are ready-to-use every day and basically everywhere.

Disease detection based on bioacoustic signals

A promising approach for a screening tool fulfilling these requirements could be based on bioacoustic signals such as speech sounds or cough sounds.[11-14] Several studies have reported acoustic peculiarities in the speech of patients who have diseases associated with symptoms affecting anatomical correlates of speech production, such as bronchial asthma[15,16] or vocal cord disorders.[17-19] Differences in various acoustic parameters were also found in recent studies comparing speech samples of COVID-19 positive and COVID-19 negative individuals.[20,21] Motivated by acoustic voice peculiarities found for various diseases, machine learning has been increasingly applied to automatically detect medical conditions from voice, such as upper respiratory tract infection,[22] Parkinson's disease,[23] and depression.[24] Recent studies on the automatic detection of COVID-19 from speech signals achieved promising results through both traditional machine learning[13,25-27] and deep learning techniques.[13,28,29] Although research on the automatic detection of diseases based on speech is rapidly expanding, it faces a number of challenges in terms of algorithm generalizability and potential application in real-world scenarios. These challenges include gender and age distribution, the presence of different mother tongues, dialects, sociolects, or cognitive aspects such as individual speech-language and reading competence that may affect various acoustic parameters.[30-36] Studies on COVID-19 face additional challenges related to the fact that COVID-19 is a relatively new and not yet well understood disease with a wide range of symptoms and divergent symptom severity.[37,38] Studies need to consider the symptom heterogeneity of COVID-19 positive patients and the fact that many symptoms may also occur in other diseases such as bronchial asthma or flu. Therefore, it is essential to consider the inclusion of patients with COVID-19-like symptoms but other diagnoses into COVID-19 negative study groups.

In contrast to speech, the acoustic parameters of cough sounds are less dependent on language-related aspects. Therefore, systems based on voluntarily produced cough sounds may be more easily applicable to a broader target group than speech-based systems. Cough is not only a promising bioacoustic signal since it reflects a body function performed by all people regardless of their culture or language competence, but is also one of the most prominent symptoms of COVID-19 and is closely related to the lung primarily affected by COVID-19.

Physiology of cough

Cough is an important defense mechanism of the respiratory system as it cleans the airways through high-velocity airflow from accidentally inhaled foreign materials or materials produced internally in the course of infections. A cough is composed of an inspiratory, a compressive, and an expiratory phase.[39] It is initiated with the inspiration of air (about 50% of vital capacity), followed by a prompt closure of the glottis and the contraction of abdominal muscles and other expiratory muscles. This process allows the compression of the thorax and the increase of subglottic pressure. The next phase of a cough constitutes the rapid opening of the glottis resulting in a high-velocity airflow (peak expiratory airflow phase), followed by a steady-state airflow (plateau phase) for a variable − voluntarily controllable − duration.[40,41] The optional final phase is the interruption of the airflow due to the closure of the glottis.[42] Cough can be classified into two broad categories: wet/productive cough with sputum excreted and dry/non-productive cough without sputum.[43] Cough sounds were found to vary significantly due to a person's body structure, sex, and the kind of sputum.[43] For example, the sound spectrograms of wet coughs contain clear vertical lines that appear once continuous sounds break off. This manifests in audible interruptions. Moreover, the duration of the second cough phase was longer for wet coughs compared to dry coughs, whereas the durations of the first and third cough phase did not differ significantly. Also Hashimoto and colleagues[44] revealed that the ratio of the duration of the second phase to the total cough duration was significantly higher for wet coughs than for dry coughs. Chatrzarrin and colleagues[45] compared acoustic characteristics of wet and dry coughs and found that the number of peaks of the energy envelope of the cough signal and the power ratio of two frequency bands of the second expiratory phase of the cough signal significantly differentiated between the two cough types. Wet cough sounds presented with more peaks and a reduced frequency band power ratio, indicating more spectral variation as compared to dry cough sounds.

Disease detection from cough sounds

A number of researchers have been interested in potential acoustic differences between voluntarily produced cough sounds of patients with pulmonary diseases and healthy individuals. Knocikova and colleagues[46] compared the cough sounds of patients with chronic obstructive pulmonary disease (COPD), patients with bronchial asthma, and healthy controls. They found that patients with COPD had the longest cough duration and the highest power among the three groups. Higher frequencies were detected in the cough sounds of the bronchial asthma group compared with the COPD group. Furthermore, in the bronchial asthma group, the power of cough sound was shifted

to a higher frequency range compared with the control group.[46] Another study[47] found that cough duration, MFCC1 (Mel-frequency cepstral coefficient), and MFCC9 features were the most important acoustic features for classification of pulmonary disease state (i.e., bronchial asthma, COPD, chronic cough, healthy) and disease severity, defined based on a patient's forced expiratory volume in the first second (FEV1) divided through the forced vital capacity (FVC). Similar to the speech/voice domain, various automatic approaches have proved to be effective at detecting pulmonary diseases from cough sounds[47,48]; good performance was even achieved when differentiating between two obstructive pulmonary diseases, namely bronchial asthma and COPD.[49] Furthermore, using acoustic features extracted from cough sounds, Nemati and colleagues[47] automatically classified the symptom severity of patients with pulmonary diseases. In another study,[50] cough sound analysis was used to predict spirometry results, i.e., FVC, FEV1, and FEV1/FVC, for patients with obstructive, restrictive, and combined obstructive-restrictive pulmonary diseases as well as healthy controls. Machine learning algorithms were also applied to distinguish pertussis coughs from croup and other coughs in children.[51] Nemati and colleagues[52] used a random forest algorithm to classify wet and dry coughs based on a comprehensive set of acoustic features and achieved an accuracy of 87%. Notably, the accuracy is calculated as the average of the sensitivity (88%) and specificity (86%) for classification of wet and dry cough sounds. Based on improved reverse MFCCs, Zhu and colleagues[53] achieved an accuracy in the classification of wet and dry coughs of 93.66% using hidden Markov models.

COVID-19 detection based on cough sounds

A set of studies have investigated detecting COVID-19 from cough sounds. Alsabek and colleagues[54] compared MFCC acoustic features in cough, breathing, and voice samples of COVID-19 positive and COVID-19 negative individuals. They found a higher correlation between the COVID-19 positive group and the COVID-19 negative group for the voice samples than for the cough and breathing samples. Therefore, they concluded that the cough and breathing of a patient may be more suitable for detecting a COVID-19 infection than his or her voice. Another study[55] collected cough sounds from public media interviews with COVID-19 positive patients and analyzed them for the number of peaks present in the energy spectrum and power ratio between the first two phases of each cough event. They found the majority of cough sounds to have a low power ratio and a high number of peaks, a characteristic pattern previously reported for wet coughs.[45] Brown and colleagues[11] compared several hand-crafted features extracted from their collected crowd-sourced cough sounds of COVID-19 positive and COVID-19 negative individuals. They found that coughs from COVID-19 positive individuals are longer in total duration, and have more pitch onsets,

higher periods, and lower root mean square (RMS) energy. In contrast, their MFCC features have fewer outliers compared to those of COVID-19 negative individuals.

The reported differences in acoustic features extracted from cough sounds of COVID-19 positive and COVID-19 negative individuals are promising for the automatic detection of COVID-19. To process hand-crafted features, traditional machine learning methods such as support vector machines (SVMs) and extreme gradient boosting were utilised.[11,13,56] End-to-end deep learning models were developed to detect COVID-19 from the log spectrograms of cough sounds, and performed better than the linear SVM baseline.[57] Similarly, deep learning was also successfully used to process MFCCs[28,58,59] or Mel spectrograms[60] of cough sounds. The studies above have raised the potential and shown the effectiveness of machine learning for a cough sound-based detection of COVID-19.

Contributions of this work

Due to advancements in signal processing and machine learning technology, today's computers are able to "listen to" sounds and identify acoustic patterns which often remain hidden for human listeners. The rapidly growing field of machine listening aims to teach computers to automatically process and evaluate audio content for the purpose of a wide range of acoustic detection/classification tasks. In the present study, we analyze acoustic differences in cough sounds produced by COVID-19 positive and COVID-19 negative individuals and further explore the feasibility of machine listening techniques to automatically detect COVID-19. On the one hand, we include COVID-19 positive and COVID-19 negative individuals irrespective of the presence or absence of any symptoms associated with COVID-19. On the other hand, this study aims to address the above-mentioned challenges of symptom heterogeneity of COVID-19 positive patients including asymptomatic COVID-19 infections as well as similarities of symptoms to symptom characteristics of other diseases. Thus, we also investigate the isolated scenarios of COVID-19 positive and COVID-19 negative individuals all of which showing COVID-19-associated symptoms, and of COVID-19 positive and COVID-19 negative individuals all of which not showing any COVID-19-associated symptoms. Data for our experiments is taken from the open COUGHVID database[61] that provides 27,500 cough recordings in conjunction with information about present symptoms. We analyse the acoustic features of the COMPUTATIONAL PARALINGUISTICS CHALLENGE (COMPARE) feature set that recently achieved good performance for COVID-19 detection from cough sounds.[13] Furthermore, we train an effective (i.e., significantly better than the chance level) machine learning classifier based on the extracted COMPARE features. Finally, we investigate the contribution of the acoustic features extracted from the cough sounds to the COVID-19 status predictions of the machine learning classifier.

MATERIALS AND METHODS

In this study, we apply standard audio processing, statistics, and machine learning methodology to recorded cough sounds in order to address different tasks of interest. We set a special focus on explainability. Therefore, our analyses are based on established acoustic features initially extracted from the audio files.

Databases and tasks

Data preprocessing

The dataset in our study is selected from the ongoing crowdsourcing COUGHVID data collection.[61,62] The COUGHVID database is collected via a web interface.[63] Thus, the recordings can be collected with personal computers, laptops, or smartphones. All cough recordings are voluntarily produced by the participants. The participants receive safe coughing instructions on the web page, e.g., holding the smartphone at arm's length and coughing into the crook of the elbow, putting the phone into a plastic zip bag. Each audio recording lasts up to 10 seconds. At the time of analysis for this study, the latest released COUGHVID database consists of 27,550 cough sound files. There are three statuses of a cough sample for each participant to self-report: healthy, symptomatic without COVID-19 diagnosis, and COVID-19. There are 2,800 audio recordings annotated by experts, including a diagnosis, severity level, and whether or not audible health anomalies are present, such as dyspnea, and wheezing. As only a small proportion of the data have expert labels, the expert labels are not used in our work. Additionally, it is optional for each participant to report the geographic location (latitude, longitude), age, gender, and whether she/he has other preexisting respiratory conditions and muscle pain/fever symptoms. Apart from the self-reported information, the data collectors trained an extreme gradient boosting (XGB) classifier on 121 cough sounds and 94 non-cough sounds to predict the probability of a recording containing cough sounds to exclude non-cough recordings.[61]

Since the participants with the symptomatic status did not explicitly report whether they were diagnosed with COVID-19, we only include the samples labelled as healthy (i.e., negative) and COVID-19 (i.e., positive) in the present study. Furthermore, we exclude samples with cough sound probabilities below or equal to 0.99 trying to ensure that each recording contains useful cough sounds. We note that no participant information was released in Orlandic and colleagues.[61] We assume the audio files with the same location, age, and gender were recorded from the same participant. To better implement subject-independent experiments, we aim to include only one cough sound file per participant. Therefore, the audio files with the same location, age, and gender are reduced into a single one by random selection. Moreover, all audio files containing noise and speech are manually excluded. The final dataset for this study contains 1,411 audio files (3.75 hours) with a mean duration of 9.57 ± 3.30 s standard deviation (SD); the audio files are re-sampled into 16 kHz.

TABLE 1.
Total and Gender-Specific Distribution of Number of Cough Samples Across COVID-19 Status and Symptom Conditions.

Status		Symptoms	Total	Gender (f/m)
neg	neg_	no	996	293/703
	neg_+	respiratory only	124	48/76
		muscle pain/fever only	57	20/37
		both symptoms	24	8/16
Σ			1 201	369/832
pos	pos_	no	111	36/75
	pos_+	respiratory only	40	13/27
		muscle pain/fever only	27	12/15
		both symptoms	32	14/18
Σ			210	75/135

neg, COVID-19 negative; pos, COVID-19 positive; f, female; m, male; +, symptomatic; _, asymptomatic.

For each status (i.e., COVID-19 negative, COVID-19 positive), four classes of clinical conditions are considered: no symptoms, respiratory symptoms only, muscle pain/fever symptoms only, and both aforementioned symptoms. For each symptom class, the total number of samples and their gender distributions are listed in Table 1. Similar to other COVID-19 related acoustic databases,[13,64] the sample numbers at the two statuses are imbalanced. Furthermore, the age distribution is examined across all symptom conditions (see Figure 1).

Task definition

This study has three major aims. Firstly, we aim to identify useful acoustic features for COVID-19 detection. To this end, we compare a set of acoustic features extracted from cough sounds of COVID-19 positive and COVID-19 negative individuals by means of conventional inferential statistics (Section: Acoustic feature extraction and analysis). Secondly, we aim to demonstrate basic feasibility of automated COVID-19 detection based on cough sounds by applying machine listening methodology. Thirdly, we aim to investigate explainability of automatic COVID-19 detection by comparing feature relevance in the machine listening approach to feature relevance according to conventional group difference testing. To achieve these aims, we group the dataset based on the related COVID-19 and symptom status into:

- samples of COVID-19 positive participants with respiratory and/or muscle pain/fever symptoms (pos_+)
- samples of COVID-19 positive participants without respiratory and/or muscle pain/fever symptoms (pos_)
- samples of COVID-19 negative participants with respiratory and/or muscle pain/fever symptoms (neg_+)
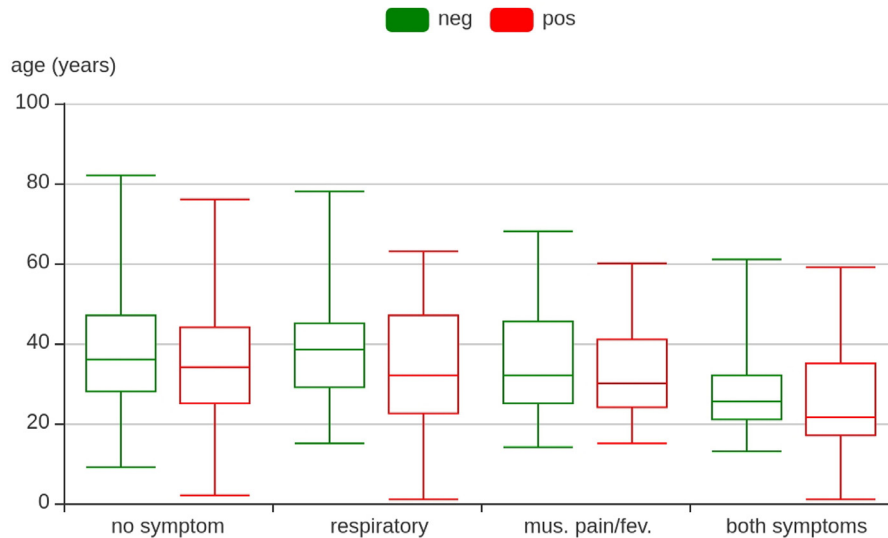
FIGURE 1. The age distribution of the 1,411 cough samples of the dataset for COVID-19 neg(ative) or (pos)itive. Mus., muscle; fev., fever.

- samples of COVID-19 negative participants without respiratory and/or muscle pain/fever symptoms (neg_).

Based on these subgroups, we implement three clinically meaningful tasks:

- *Task 1: COVID-19 positive (pos) vs COVID-19 negative (neg)*. This task addresses the three aims of this study by including all samples of COVID-19 positive participants (210) and all samples of COVID-19 negative participants (1,201) irrespective of the presence or absence of symptoms.
- *Task 2: COVID-19 positive with symptoms (pos_+) vs COVID-19 negative with symptoms (neg_+)*. This task addresses the three aims of this study by including only samples of COVID-19 positive participants with respiratory and/or muscle pain/fever symptoms (99) and samples of COVID-19 negative participants with respiratory and/or muscle pain/fever symptoms (205).
- *Task 3: COVID-19 positive without symptoms (pos_) vs COVID-19 negative without symptoms (neg_)*. This task addresses the three aims of this study by including only samples of asymptomatic COVID-19 positive participants (111) and samples of asymptomatic COVID-19 negative participants (996).

Apart from all three tasks, it is interesting to investigate the performance in these tasks under different genders or age ranges, as previous studies have shown gender and age differences in cough behaviour.[65,66] As shown in Figure 1, the median of the age ranges is around 30, which is used to split the data into two age groups for each task.

Acoustic feature extraction and analysis
Both instrumental phonetic analysis and traditional machine learning build upon acoustic features in the cough signals. We extract features from every audio recording by the open-source toolkit openSMILE[67] according to the ComParE feature set, which is the standard baseline feature set in the INTERSPEECH ComParE series[68] and has proven effective in COVID-19 detection from cough sounds.[13] The ComParE feature set consists of 6,373 features calculated by several supra-segmental functionals, e.g., mean value, over segmental low-level descriptors (LLDs), e.g., loudness ($\rightarrow$ mean [loudness] = mean of loudness). The LLDs, in the form of sequential features, are generated by analysing short-time segments, while the functionals focus on mapping the LLDs into a feature vector through computing statistical features inside each LLD and over multiple LLDs. The details of the ComParE feature set can be found in Schuller and colleagues.[68]

Using the Kolmogorov-Smirnov test, we determine at a 5% significance level that most class-specific feature distributions are unlikely to come from standard normal distributions. Thus, we apply the non-parametric two-sided Mann-Whitney U test to analyze the extracted features for distribution differences between the COVID-19 positive and COVID-19 negative samples in each task. We further compute the effect size r as the correlation coefficient calculated as the z-value divided by the square root of the number of samples. Finally, we rank the features according to the effect size's absolute value. Features showing at least a weak correlation effect ($|r| \geq 0.1$) are considered relevant for the respective task. These features are referred to as top features.

Automatic COVID-19 detection
Classifiers. Based on all 6,373 extracted ComParE features, we apply machine learning methodology to study automatic COVID-19 detection in (three) binary classification tasks. Several classification approaches come into consideration, including linear models and non-linear models.

A linear model learns a linear mapping between the inputs, i.e., the features, and the labels, i.e., the COVID-19 status; a non-linear model learns a non-linear mapping. In our work, a set of models are applied to detect COVID-19 from cough sounds. The employed linear models consist of linear regression models, i.e., least absolute shrinkage and selection operator (LASSO), Ridge, and ElasticNet, and a linear SVM model. Linear regression models, i.e., logistic regression in classification tasks, construct a linear model with different penalties, i.e., L1, L2, and a combination of L1 and L2, leading to the three models: LASSO, Ridge, and ElasticNet, respectively. The coefficients of the features in a linear regression model can be considered as the feature importance. Additionally, an SVM model is trained to find a hyperplane to maximize the margin between two classes. In linear SVM models, the coefficients of this hyperplane can be regarded as weights, whose absolute values indicate the relevance of each feature in the decision function − the larger the absolute value, the more important the respective feature.[69,70] Apart from the linear models, the utilized non-linear models contain decision tree, random forest, and multilayer perceptron (MLP). A decision tree constructs a tree-like model by learning simple decision rules from the features, and a random forest is composed of a number of decision trees for performance improvement. Both decision tree and random forest are able to calculate the feature importance, which is computed as the total reduction of the criterion brought by each feature. A set of hidden layers in MLP leads to a highly non-linear function between the inputs and the labels, which makes it challenging to interpret each feature's role in the model. To learn the feature importance of neural networks (e.g., MLP), a set of methods have been proposed, such as deep learning important features[71] and causal explanation (CXPlain).[72]

The COMPARE features have shown effectiveness in various audio classification tasks, including pathological-speech-related disease detection,[73] on small to medium-sized datasets and represent one of the official machine learning pipelines of the INTERSPEECH COMPARE series.[13,68] In this study, we reapply this well-proven feature set in combination with the aforementioned models to investigate the basic feasibility of detecting COVID-19 from cough sounds. The whole COMPARE features extracted from the audio samples are used as the input of the machine learning models.

Due to the limited size of the data, splitting the dataset into training, development, and test may lead to unreliable results. Hence, a five-fold cross validation strategy is used to generate the predictions over all audio samples. The whole dataset is equally split into five folds, each of which is used as the test set while the other four are employed as the training set. We combine all test sets in the 5-fold cross validation for performance evaluation; the final results are obtained on the combined dataset. The best hyperparameter is then selected corresponding to the best performance on the combined dataset. The linear regression models and the SVM model are optimized from multiple inverse

values of regularization strength and complexity parameters $C \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ respectively. Random forest is optimized from tree numbers in $\{50, 100, 150, 200, 250, 300\}$. The decision tree is learnt with the default parameters in scikit-learn.[74] In the MLP model, three linear layers with the numbers of output neurons 1,024, 256, and 2. To avoid overfitting problems, the first two layers are followed by dropout operations with the probabilities of setting a neuron as 0.2 and 0.3, respectively. Notably, balanced class weights are applied to each SVM model in order to mitigate the data imbalance problem.

Evaluation metrics. We use the unweighted average recall (UAR) to evaluate the classification performance purposefully without considering the data imbalance characteristics. UAR is the average of recalls on all classes. Additionally, we report the area under the receiver operating characteristic curve (AUC-ROC) calculated based on the probability estimates of each audio sample being predicted as the COVID-19 positive class. The AUC score may be inconsistent with the UAR, since the probability of a prediction is calibrated by Platt scaling and fit by an additional cross-validation procedure on the training data. Finally, the confusion matrices for all three classification tasks are depicted to show the detailed performance.

Explainability of automatic COVID-19 detection. To provide an insight into the best performing linear classifier of each task, we export the respective model's feature coefficients and calculate mean weights across all cross-validation folds. We then rank the features for relevance, i.e., according to the absolute value of the mean feature weights. The explainability of automatic COVID-19 detection via linear classification models is quantified by comparing feature relevance in the linear classification model to feature relevance according to the effect size in the non-parametric group difference test.

RESULTS

Feature analysis

The analysis of the extracted 6,373 COMPARE features yields a number of features relevant for the investigated tasks. For Task 1 (pos vs neg), we identify 220 top features, i.e., features with an absolute value of the effect size r greater than or equal to .1 in the non-parametric group difference test. For Task 2 (pos_+ vs neg_+) and Task 3 (pos_- vs neg_-), 1,567 and 46 features are found relevant, respectively. Table 2 reveals the LLDs underlying the respective top features of each task. All LLD categories of the COMPARE set turn out to be relevant for the differentiation of symptomatic participants with and without COVID-19 (Task 2). However, when including asymptomatic participants (Task 1) or exclusively focusing on asymptomatic participants (Task 3), only selected energy-related, spectral, and voicing-related LLDs are found to differ at an absolute value of the effect size $|r| \geq 0.1$. Figure 2 depicts the group-wise probability

TABLE 2.
Categorization of the 65 Low-Level Descriptors (LLDs) of the CₒₘₚₐᵣE Feature Set and Specification of Involvement ($\checkmark$) or Non-Involvement ($\times$) in a Top Feature of the Respective Differentiation Task (Task 1: pos vs neg, Task 2: pos_+ vs neg_+, and Task 3: pos_- vs neg_-).

Group	Energy-related LLDs (4)	Task 1: pos vs neg	Task 2: pos_+ vs neg_+	Task 3: pos_- vs neg_-
Prosodic	Auditory spectrum sum (loudness)	$\checkmark$	$\checkmark$	$\checkmark$
Prosodic	RASTA-filtered auditory spectrum sum	$\checkmark$	$\checkmark$	$\times$
Prosodic	RMS energy	$\checkmark$	$\checkmark$	$\checkmark$
Prosodic	zero-crossing rate	$\times$	$\checkmark$	$\times$
Group	**Spectral LLDs (55)**	**pos vs neg**	**pos_+ vs neg_+**	**pos_- vs neg_-**
Cepstral	MFCC 1−14	$\checkmark$	$\checkmark$	$\checkmark$
Spectral	Psychoacoustic harmonicity	$\checkmark$	$\checkmark$	$\checkmark$
Spectral	Psychoacoustic sharpness	$\times$	$\checkmark$	$\times$
Spectral	Spectral centroid	$\times$	$\checkmark$	$\times$
Spectral	Spectral energy 250−650 Hz, 1−4 kHz	$\checkmark$	$\checkmark$	$\checkmark$
Spectral	Spectral entropy	$\times$	$\checkmark$	$\times$
Spectral	Spectral flux	$\checkmark$	$\checkmark$	$\checkmark$
Spectral	Spectral kurtosis	$\checkmark$	$\checkmark$	$\times$
Spectral	Spectral roll-off point 0.25, 0.50, 0.75, 0.90	$\checkmark$	$\checkmark$	$\times$
Spectral	Spectral skewness	$\times$	$\checkmark$	$\times$
Spectral	Spectral slope	$\checkmark$	$\checkmark$	$\checkmark$
Spectral	Spectral variance	$\times$	$\checkmark$	$\times$
Spectral	RASTA-filtered auditory spectral band 1−26	$\checkmark$	$\checkmark$	$\checkmark$
Group	**Voicing-related LLDs (6)**	**pos vs neg**	**pos_+ vs neg_+**	**pos_- vs neg_-**
Prosodic	Fundamental frequency	$\times$	$\checkmark$	$\times$
Quality	HNR	$\checkmark$	$\checkmark$	$\checkmark$
Quality	Jitter (local and DDP)	$\times$	$\checkmark$	$\times$
Quality	Shimmer	$\times$	$\checkmark$	$\times$
Quality	Voicing probability	$\times$	$\checkmark$	$\times$

DDP, difference of differences of periods; HNR, harmonics-to-noise ratio; MFCC, Mel-frequency cepstral coefficient; neg, COVID-19 negative; pos, COVID-19 positive; RASTA, relative spectral transform; RMS, root mean square; +, symptomatic; −, asymptomatic.

density estimates of the respective top one feature of each task, i.e., the mean inter-peak distance of the RMS energy with $|r| \geq 0.15$ for Task 1, the mean inter-peak distance of the fourth MFCC with $|r| \geq 0.26$ for Task 2, and again the mean inter-peak distance of the RMS energy with $|r| \geq 0.15$ for Task 3. Fourteen out of 6,373 features are found to be jointly relevant for the differentiation between COVID-19 positive and COVID-19 negative in both symptomatic and asymptomatic participants (Table 3).

Automatic COVID-19 detection
The performance of the machine learning models for our three tasks is shown in Table 4. All best UARs for the three tasks significantly exceed chance level (UAR: 0.5) in a one-tailed z-test (pos vs neg: $p < 0.001$; pos_+ vs neg_+: $p < 0.001$; and pos_- vs neg_-: $p < 0.001$). Correspondingly, the confusion matrices of all three results are depicted in Figure 3. All negative classes in the three tasks are classified with high true negative rates, while the true positive rates are around 0.5. Among the three tasks, for the task of pos_- vs neg_- we

achieve the lowest UAR at the highest ratio of COVID-19 positive samples being incorrectly assigned to the negative class (see Figure 3). In Table 5, the performance on the data from the male participants is marginally better than that on data from the female participants. Furthermore, the performance for participants under or equal to 30 years is better for two of the three tasks than the performance for participants over 30 years.

Explainability of the COVID-19 detection model
Our closer look at the weighting of acoustic features in the trained linear models reveals that identified top features according to the effect size in the non-parametric group difference test are also higher weighted in the respective best performing linear classification model for each task, i.e., Ridge for Tasks 1 and 2, and ElasticNet for Task 3 (see Figure 4). The 220 top features of Task 1 have a mean rank of $1,496 \pm 1,640$ SD amongst the features ranked according to the absolute value of the Ridge weights. Forty-nine out of the 220 top features are also among the first 220 Ridge
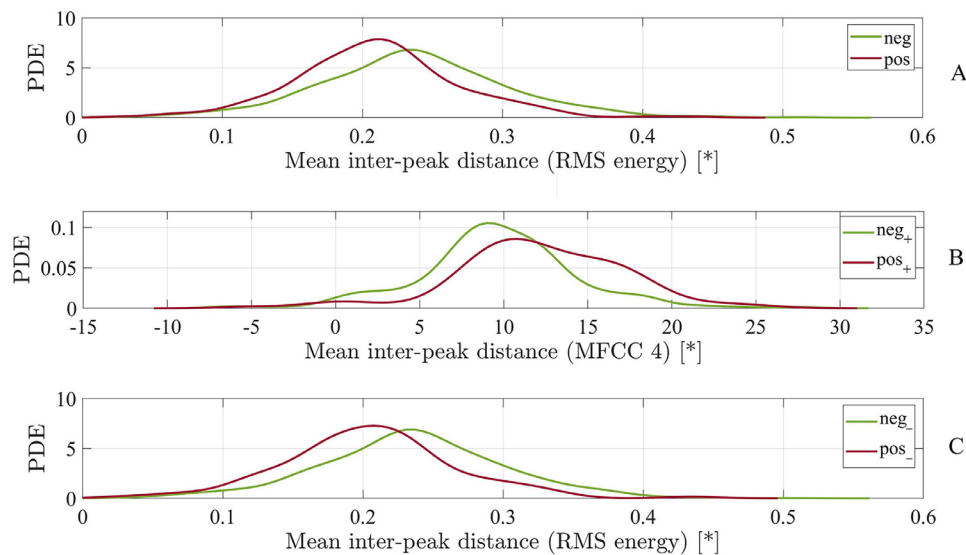
FIGURE 2. Comparison between part A: COVID-19 positive (pos) and COVID-19 negative (neg) participants (Task 1), part B: symptomatic COVID-19 positive (pos$_+$) and symptomatic COVID-19 negative (neg$_+$) participants (Task 2), and part C: asymptomatic COVID-19 positive (pos$_-$) and asymptomatic COVID-19 negative (neg$_-$) participants (Task 3) by means of the probability density estimate (PDE) of the top one feature of the respective differentiation task. MFCC, Mel-frequency cepstral coefficient; RMS, root mean square; *, real measurement unit does not exist as feature values refer to the amplitude of the digital audio signal.

weight ranked features. The 1,567 top features of Task 2 have a mean rank of $1,166 \pm 1,084$ SD amongst the Ridge weight ranked features. Herein, 1,173 out of the 1,567 top features are among the first 1,567 Ridge weight ranked

features. The best performing model for Task 3, i.e., Elastic-Net, only builds upon 250 non-zero feature coefficients. The 250 features with non-zero coefficients include 42 out of the 46 top features of Task 3. Eighteen out of the 46 top features are also among the first 46 ElasticNet weight ranked features.

TABLE 3.
Joint Top Features Between the pos$_+$ vs neg$_+$ (Symptomatic COVID-19 Positive vs Symptomatic COVID-19 Negative) and the pos$_-$ vs neg$_-$ (Asymptomatic COVID-19 Positive vs Asymptomatic COVID-19 Negative) Differentiation Tasks Listed According to Their Mean Ranks Rounded to Integers.

Mean rank	Feature
17	Flatness (Δ spectral energy 250−650 Hz)
19	Flatness (spectral energy 250−650 Hz)
27	Flatness (RMS energy)
33	Flatness (spectral flux)
149	Mean inter-peak distance (RMS energy)
194	Quartile 3 (HNR)
195	IQR 1−3 (HNR)
196	IQR 2−3 (HNR)
202	Mean inter-peak distance (loudness)
291	Mean inter-peak distance (spectral flux)
410	Skewness (Δ RASTA-filtered auditory spectral band 12)
635	Mean value of peaks (RMS energy)
725	Mean inter-peak distance (spectral harmonicity)
786	Mean value of peaks (loudness)

HNR, harmonics-to-noise ratio; IQR, interquartile range; RASTA, relative spectral transform; RMS, root mean square; Δ= first-order derivative.

DISCUSSION

This study considers the presence/absence of COVID-19-associated symptoms when comparing acoustic features extracted from cough sounds produced by COVID-19 positive and COVID-19 negative individuals and when applying machine listening technology to detect COVID-19 automatically. Although the classification performance of the SVM used in our study is significantly better than chance level, there are studies reporting better performances for COVID-19 detection based on cough sounds.[11,57] Brown and colleagues[11] studied three classification tasks: COVID-19/non-COVID, COVID-19 with cough/non-COVID with cough, and COVID-19 with cough/non-COVID asthma cough. The second and the third tasks were based on cough sounds and breath sounds, respectively, whereas the first task was based on both cough and breath sounds. All three tasks achieved over 80%, higher than those in our three tasks, which is possibly caused by two reasons. Firstly, the number of users in Brown and colleagues[11] is quite small and potentially less representative[75] as compared to our study: The number of users in Brown and colleagues[11] for each task is 62/220, 23/29, and 23/18, respectively, whereas the task-wise numbers of samples in our study are 210/1,201, 99/205, and 111/996. Secondly, the first and third tasks utilized breath

TABLE 4.
Classification Performance in Terms of Unweighted Average Recall (UAR) and Area Under the Receiver Operating Characteristic Curve (AUC) for the Three Tasks.

	Task	(1) pos vs neg		(2) pos+ vs neg+		(3) pos− vs neg−	
	Samples (#)	210/1,201		99/205		111/996	
	Models	UAR	AUC	UAR	AUC	UAR	AUC
Linear	LASSO	0.586	0.625	0.573	0.547	0.536	0.549
	Ridge	*0.632*	*0.671*	*0.653*	*0.641*	0.558	0.594
	ElasticNet	0.615	0.650	0.598	0.596	*0.565*	*0.591*
	SVM	0.610	0.642	0.601	0.609	0.563	0.617
Non-linear	Decision Tree	0.521	0.521	0.538	0.538	0.501	0.501
	Random Forest	0.500	0.651	0.544	0.606	0.500	0.600
	MLP	0.558	0.600	0.593	0.606	0.505	0.570

neg, COVID-19 negative; pos, COVID-19 positive; +, symptomatic; −, asymptomatic.

sounds, which perhaps provide some discriminative features. Coppock and colleagues[57] trained deep neural networks on the log spectrograms of both cough and breath sounds from the same crowd-sourced dataset as in Brown and colleagues[11] and achieved better results on the three tasks compared with the baseline, where SVMs processed the COMPARE features. The AUCs in two of the three tasks are above 82% and the UARs are above 76%. Similarly, the better performance of this work could be caused by the limited number of participants (26/245, 23/19, and 62/293) and features from breathing sounds. In addition, an extra task of distinguishing COVID-19 and healthy participants without symptoms was set in Coppock and colleagues.[57] Nevertheless, the performance on COVID-19 positive samples without symptoms was not reported independently. In contrast, such performance is evaluated in our Task 3, which is crucial for preventing COVID-19 transmission. The found minor gender and age differences in detection performance are most probably related to imbalances in the dataset. There are more than twice as many data samples from male than from female participants. With regard to age, it has to be considered that with increasing age the likelihood of

chronic lung and voice diseases/problems also increases which might to some extent mask symptoms caused by an acute respiratory disease.

Our study reveals several acoustic peculiarities in COVID-19. As shown in Table 2, a set of LLDs could be helpful for differentiating COVID-19 positive individuals from COVID-19 negative ones. Across the three tasks, there are common LLDs of high relevance according to the effect size in the non-parametric group difference test, namely loudness, RMS energy, MFCCs, psychoacoustic harmonicity, spectral energy, spectral flux, spectral slope, RASTA-filtered auditory spectral bands, and HNR. Differences in RMS energy and MFCC-related features between the coughs of COVID-19 positive and COVID-19 negative individuals are also reported in Brown and colleagues.[11] In contrast to Tasks 1 and 3, i.e., tasks including asymptomatic participants, in our Task 2 all LLD categories of the COMPARE set are found to bear relevant acoustic information to distinguish between the two groups. This might be due to an increased acoustic variability of symptomatic coughs (as compared to asymptomatic coughs) being reflected in a wider range of acoustic parameters. In other words, the
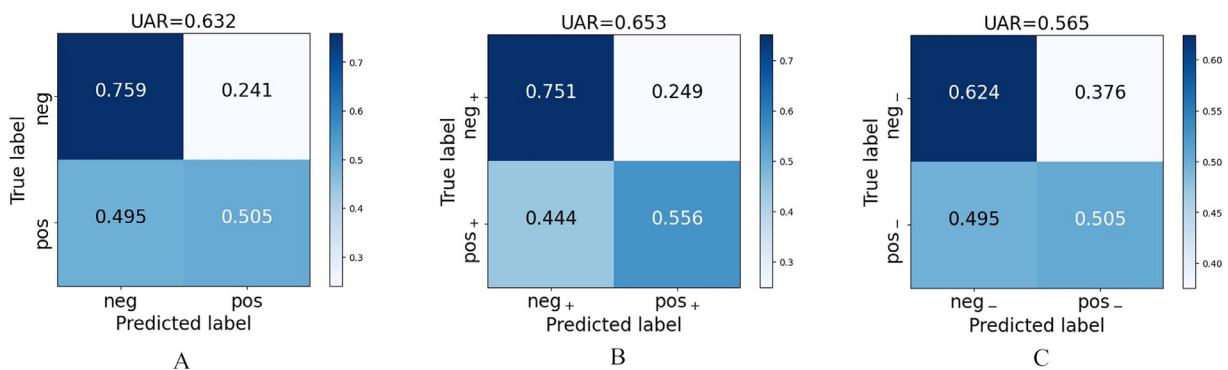

FIGURE 3. Confusion matrices for the three classification tasks complementary to the best results given in Table 4. Part A: (Task 1) pos vs neg, part B: (Task 2) pos+ vs neg+, part C: (Task 3) pos− vs neg−. neg, COVID-19 negative; pos, COVID-19 positive; UAR, unweighted average recall; +, symptomatic; −, asymptomatic.

TABLE 5.
Gender-Wise and Age Group-Wise Classification Performance in Terms of Unweighted Average Recall (UAR) and Area Under the Receiver Operating Characteristic Curve (AUC) for the Three Tasks. The first two tasks (pos vs neg and pos₊ vs neg₊) are achieved by the Ridge models, and the third one (pos₋ vs neg₋) is achieved by the ElasticNet models. neg, COVID-19 negative; pos, COVID-19 positive; y, years; ₊, symptomatic; ₋, asymptomatic.

Task	Female		Male	
	UAR	AUC	UAR	AUC
(1) pos vs neg	0.602	0.620	0.601	0.636
(2) pos₊ vs neg₊	0.573	0.563	0.679	0.676
(3) pos₋ vs neg₋	0.517	0.508	0.604	0.584

Task	Age ≤ 30y		Age > 30y	
	UAR	AUC	UAR	AUC
(1) pos vs neg	0.604	0.604	0.590	0.609
(2) pos₊ vs neg₊	0.699	0.667	0.549	0.502
(3) pos₋ vs neg₋	0.500	0.496	0.551	0.562

difference between coughs of symptomatic individuals with COVID-19 and individuals with symptoms caused by any other disease acoustically manifests more manifold than the difference between COVID-19-related and non-COVID-related coughs in a sample that also or exclusively contains asymptomatic individuals. However, the lower number of available cough samples for Task 2 as compared to the other tasks might also cause a biased distribution of feature values. The analyzed feature weights within the linear

classification models show consistency with the features' effect sizes, i.e., most top features according to the effect size also have higher weights in the linear classification models. This is a relevant finding towards the explainability of the applied machine learning approach. As indicated in Table 2, there are less top LLDs for Task 3. That might be because LASSO tends to use less features due to the nature of L1 regularization. As a combination of LASSO and Ridge, ElasticNet is based on less features compared with Ridge.

As both speech and cough sounds are produced by the respiratory system, we herein compare and analyze peculiar acoustic parameters of patient's speech and cough sounds. When analyzing the acoustic peculiarities of patients with diseases that affect the anatomical correlates of speech production, the related studies reported that the peculiar acoustic parameters of the patients' speech include fundamental frequency (f_o), vowel formants, jitter, shimmer, HNR, and maximum phonation time (MPT).[15,16,18,19] Additionally, the peculiar acoustic parameters of voice samples of COVID-19 positive and COVID-19 negative participants were reported to include f_o standard deviation, jitter, shimmer, HNR, the difference between the first two harmonic amplitudes (H1−H2), MPT, cepstral peak prominence,[20] mean voiced segment length, and the number of voiced segments per second.[21] We can find that there are common acoustic peculiarities between the voice of COVID-19 patients and patients with some other diseases: f_o-related features, jitter, shimmer, HNR, and MPT. In Table 2, several acoustic LLDs of cough sounds have shown potential for distinguishing COVID-19 positive and COVID-19 negative individuals. Particularly, f_o, jitter, shimmer, and HNR have high effective sizes in Task 2, i.e., pos₊ vs neg₊. The above findings indicate that there are similarities in acoustic
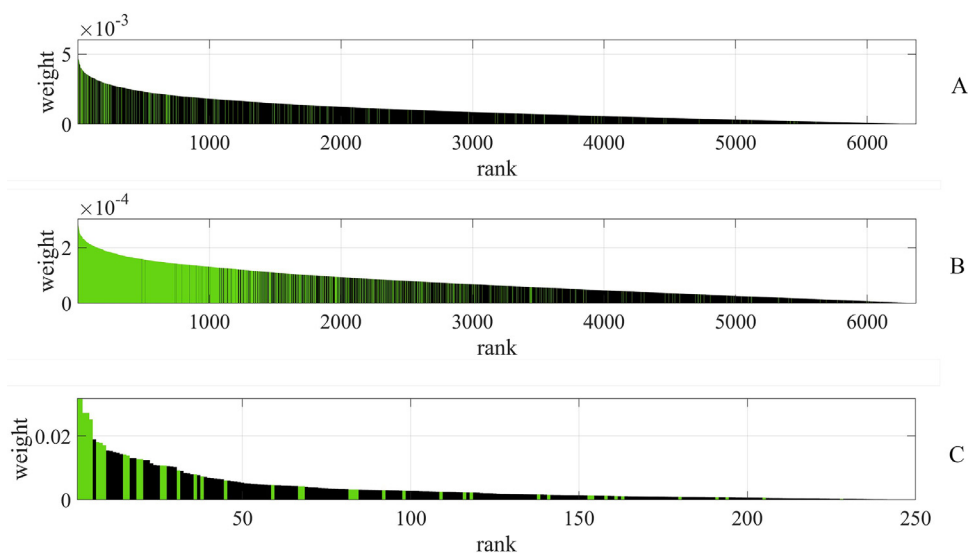

FIGURE 4. Feature ranking according to the absolute value of Ridge feature weights for part A: (Task 1) pos vs neg and part B: (Task 2) pos₊ vs neg₊, as well as of ElasticNet feature weights for part C: (Task 3) pos₋ vs neg₋. Green bars indicate top features according to the effect size in the non-parametric group difference test. A different x-axis scaling is used for (Task 3) as the ElasticNet model only builds upon 250 non-zero feature coefficients.

peculiarities of speech and cough sounds of COVID-19 patients.

Limitations

The classification performance reported in our study needs to be interpreted in the light of the well-known challenges of data collection via crowdsourcing, including data validity, data quality, and participant selection bias.[76-78] The COUGHVID database does not allow to verify the COVID-19 status of the participants, as the participants were not asked to provide a copy or confirmation of their positive or negative COVID-19 test. Another limitation is that the participants have not been instructed to record the data during a defined time window after the positive or negative COVID-19 test. Therefore, it is possible that some participants recorded their cough at the beginning of their infection, whereas others did the recording towards the end of their infection. Interestingly, the disease stage of COVID-19 was found to influence the nature of the cough (shifting from dry at an early disease stage to more wet at a later disease stage), concomitantly affecting acoustic parameters of the cough.[55] Moreover, the participants were asked to answer whether they had respiratory and/or muscle/pain symptoms, but no information on the severity of their symptoms is available. Although the safe recording instructions provided on the web page are reasonable with regard to the transmission of the virus, the suggestion to put the smartphone into a plastic zip bag while recording is suboptimal from an acoustic perspective. Another limitation of our study is that the participants did not receive clear instructions on how to cough, e.g., how often, or whether to take a breath between two coughs. Various audio recording devices and settings are inherent for crowdsourcing; we expect no bias towards one of the participant groups concerning the use of recording devices. We reduced files with the same location, age, and gender into a single one to promote that only one cough sound file per participant is included, but we cannot guarantee that our dataset has only one sample per participant or that we have not mistakenly merged recordings from various individuals living in the same household.

Our target in this work was to explore the hand-crafted features' importance for automatic COVID-19 detection. Some classifiers like k-nearest neighbours were not utilized as it might be difficult for them to output the feature coefficients/importance. Other approaches, such as transfer learning and end-to-end deep learning, were not used, as their inputs are either the original audio waves or simple time-frequency representations. Therefore, it is challenging to explain the features' contribution with these methods. Additionally, we decided to apply a cross-validation schema due to the small dataset size, thus, testing was not entirely independent from the training as hyper-parameters were optimized on the test partitions.

We decided to employ just a single dataset rather than multiple datasets in this study. We selected the COUGH-VID dataset because it contains enough data and sufficient meta information to analyze the effects of COVID-19-related symptoms on acoustic parameters and automatic COVID-19 detection. Coswara[79] was also considered at the start of the experiments. However, the symptom information is not complete and well-organized for our study to analyze the effect of symptoms for detecting COVID-19. We also considered well-structured data, including University of Cambridge dataset collected by the COVID-19 Sounds app,[11] diagnosis of COVID-19 using acoustics (DiCOVA) 2021 challenge data,[64] and INTERSPEECH COMPARE 2021 challenge data.[13] However, these databases did not provide (sufficient) symptom information.

CONCLUSIONS

In this study, we acoustically analyzed cough sounds and applied machine listening methodology to automatically detect COVID-19 on a subset of the COUGHVID database (1,411 cough samples; COVID-19 positive/negative: 210/1,201). Firstly, the acoustic correlates of COVID-19 cough sounds were analyzed by means of conventional statistical tools based on the COMPARE set containing 6,373 acoustic higher-level features. Secondly, machine learning models were trained to automatically detect COVID-19 and evaluate the features' contribution to the COVID-19 status predictions. A number of acoustic parameters of cough sounds, e.g., statistical functionals of the root mean square energy and Mel-frequency cepstral coefficients, were found to be relevant for distinguishing between COVID-19 positive and COVID-19 negative cough samples. Among several linear and non-linear automatic COVID-19 detection models investigated in this work, Ridge linear regression achieved a UAR of 0.632 for distinguishing between COVID-19 positive and COVID-19 negative individuals irrespective of the presence or absence of any symptoms and, thus, performed significantly better than chance level. With regard to explainability, the best performing machine learning models were found to have put higher weight on acoustic features that yielded higher effects in conventional group difference testing.

OUTLOOK

Automatic COVID-19 detection from cough sounds can be helpful for the early screening of COVID-19 infections, saving time and resources for clinics and test centers. Specifically, machine listening applications distinguishing between cough samples of symptomatic COVID-19 positive individuals and those of individuals with other diseases could advise the patient to stay at home and contact her/his doctor by phone before entering clinics/hospitals to meet medical professionals. This would help to prevent the spread of the virus in an especially vulnerable population. By distinguishing between cough samples produced by asymptomatic COVID-19 positive and COVID-19 negative individuals, an easy-to-apply instrument, such as a mobile application and a hand-held testing device, could help to prevent the

unconscious transmission of the virus from asymptomatic COVID-19 positive individuals.

From our point of view, it is highly important for future studies to specify the symptoms more clearly (e.g., severity estimates, onset time of symptoms), to include additional aspects potentially affecting the cough sound, such as smoking and vocal cord dysfunctions, and to differentiate in the COVID-19 negative group between participants with chronic respiratory diseases such as asthma or COPD and patients with a temporary infection such as the flu. Furthermore, it would be interesting for future studies to acoustically analyze the cough phases separately, as previous studies reported certain phase-specific acoustic peculiarities for wet and dry coughs.[44,45] Moreover, it will be encouraging to consider more sound types (e.g., breathing and speech) and evaluate the physical and/or mental status of COVID-19 positive patients (e.g., anxiety) from speech for comprehensive COVID-19 detection and status monitoring applications.[80,81]

From the perspective of machine learning, feature selection methods will be investigated to extract useful features only. Deep learning models shall be explored for better performance due to their strong capability of extracting highly abstract representations. Particularly, when developing real-life applications for COVID-19 detection, it will be more efficient to skip the feature extraction procedure through training an end-to-end deep neural network with the input of audio signals or time-frequency representations. In addition to explaining linear classification models by analyzing the weights of the acoustic features in this study, explaining deep neural networks along the dimension of time frame or frequency will need to be investigated to provide a detailed interpretation for each specific cough sound, i.e., when and at which frequency band a cough sound shows COVID-19-specific acoustic peculiarities. For this purpose, a set of approaches could be employed, e.g., local interpretable model-agnostic explanations,[82] shapley additive explanations (SHAP),[83] and attention mechanisms.[84,85]

DECLARATIONS OF INTEREST

None.

REFERENCES

1. World Health Organization. *WHO Coronavirus (COVID-19) Dashboard*. 2022. Accessed at: February 28, 2022. Accessed from: https://covid19.who.int/.
2. Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, et al. A comprehensive review of COVID-19 characteristics. *Biol Proced Online*. 2020;22:19. https://doi.org/10.1186/s12575-020-00128-2.
3. El-Anwar MW, Elzayat S, Fouad YA. ENT manifestation in COVID-19 patients. *Auris Nasus Larynx*. 2020;47:559–564. https://doi.org/10.1016/j.anl.2020.06.003.
4. Long C, Xu H, Shen Q, et al. Diagnosis of the coronavirus disease (COVID-19): rRT-PCR or CT? *Eur J Radiol*. 2020;126: 108961. https://doi.org/10.1016/j.ejrad.2020.108961.
5. Tang Y-W, Schmitz JE, Persing DH, et al. Laboratory diagnosis of COVID-19: Current issues and challenges. *J Clin Microbiol*. 2020;58: e00512–e00520. https://doi.org/10.1128/JCM.00512-20.
6. Jiang Y, Chen H, Loew M, et al. COVID-19 CT image synthesis with a conditional generative adversarial network. *IEEE J Biomed Health Inform*. 2021;25:441–452. https://doi.org/10.1109/JBHI.2020.3042523.
7. Raptis CA, Hammer MM, Short RG, et al. Chest CT and coronavirus disease (COVID-19): a critical review of the literature to date. *AJR Am J Roentgenol*. 2020;215:839–842. https://doi.org/10.2214/AJR.20.23202.
8. Tabik S, Gómez-Ríos A, Martín-Rodríguez JL, et al. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE J Biomed Health Inform*. 2020;24:3595–3605. https://doi.org/10.1109/JBHI.2020.3037127.
9. Zhang Y, Liao Q, Yuan L, et al. Exploiting shared knowledge from non-COVID lesions for annotation-efficient COVID-19 CT lung infection segmentation. *IEEE J Biomed Health Inform*. 2021;25:4152–4162. https://doi.org/10.1109/JBHI.2021.3106341.
10. Rahbari R, Moradi N, Abdi M. rRT-PCR for SARS-CoV-2: analytical considerations. *Clin Chim Acta*. 2021;516:1–7. https://doi.org/10.1016/j.cca.2021.01.011.
11. Brown C, Chauhan J, Grammenos A, et al. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. *Proc ACM SIGKDD*. 2020:3474–3484. https://doi.org/10.1145/3394486.3412865.
12. Hecker P, Pokorny F, Bartl-Pokorny K, et al. Speaking corona? Human and machine recognition of COVID-19 from voice. *Proc Interspeech*. 2021:1029–1033. https://doi.org/10.21437/Interspeech.2021-1771.
13. Schuller BW, Batliner A, Bergler C, et al. The INTERSPEECH 2021 computational paralinguistics challenge: COVID-19 cough, COVID-19 speech, escalation & primates. *Proc Interspeech*. 2021:431–435. https://doi.org/10.21437/Interspeech.2021-19.
14. Schuller BW, Schuller DM, Qian K, et al. COVID-19 and computer audition: an overview on what speech & sound analysis could contribute in the SARS-CoV-2 Corona crisis. *Front Digit Health*. 2021;3: 564906. https://doi.org/10.3389/fdgth.2021.564906.
15. Balamurali BT, Hee HI, Teoh OH, et al. Asthmatic versus healthy child classification based on cough and vocalised /a:/sounds. *J Acoust Soc Am*. 2020;148:EL253. https://doi.org/10.1121/10.0001933.
16. Dogan M, Eryuksel E, Kocak I, et al. Subjective and objective evaluation of voice quality in patients with asthma. *J Voice*. 2007;21:224–230. https://doi.org/10.1016/j.jvoice.2005.11.003.
17. Falk S, Kniesburges S, Schoder S, et al. 3D-FV-FE aeroacoustic larynx model for investigation of functional based voice disorders. *Front Physiol*. 2021;12: 616985. https://doi.org/10.3389/fphys.2021.616985.
18. Jesus LMT, Martinez J, Hall A, et al. Acoustic correlates of compensatory adjustments to the glottic and supraglottic structures in patients with unilateral vocal fold paralysis. *Biomed Res Int*. 2015;704121. https://doi.org/10.1155/2015/704121.
19. Petrović-Lazić M, Babac S, Vuković M, et al. Acoustic voice analysis of patients with vocal fold polyp. *J Voice*. 2011;25:94–97. https://doi.org/10.1016/j.jvoice.2009.04.002.
20. Asiaee M, Vahedian-Azimi A, Atashi SS, et al. Voice quality evaluation in patients with COVID-19: an acoustic analysis. *J Voice*. 2020; S0892-1997:30368-4. https://doi.org/10.1016/j.jvoice.2020.09.024.

21. Bartl-Pokorny KD, Pokorny FB, Batliner A, et al. The voice of COVID-19: acoustic correlates of infection in sustained vowels. *J Acoust Soc Am.* 2021;149:4377–4383. https://doi.org/10.1121/10.0005194.

22. Albes M, Ren Z, Schuller B, et al. Squeeze for sneeze: compact neural networks for cold and flu recognition. *Proc Interspeech.* 2020:4546–4550. https://doi.org/10.21437/Interspeech.2020-2531.

23. Yaman O, Ertam F, Tuncer T. Automated Parkinson's disease recognition based on statistical pooling method using acoustic features. *Med Hypotheses.* 2020;135: 109483. https://doi.org/10.1016/j.mehy.2019.109483.

24. Ringeval F, Schuller B, Valstar M, et al. AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI, and cross-cultural affect recognition. *Proc AVEC.* 2019:3–12. https://doi.org/10.1145/3347320.3357688.

25. Han J, Brown C, Chauhan J, et al. Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data. *Proc ICASSP.* 2021:8328–8332. https://doi.org/10.1109/ICASSP39728.2021.9414576.

26. Shimon C, Shafat G, Dangoor I, et al. Artificial intelligence enabled preliminary diagnosis for COVID-19 from voice cues and questionnaires. *J Acoust Soc Am.* 2021;149:1120–1124. https://doi.org/10.1121/10.0003434.

27. Stasak B, Huang Z, Razavi S, et al. Automatic detection of COVID-19 based on short-duration acoustic smartphone speech analysis. *J Healthc Inform Res.* 2021;5:201–217. https://doi.org/10.1007/s41666-020-00090-4.

28. Hassan A, Shahin I, Alsabek MB. COVID-19 detection system using recurrent neural networks. *Proc CCCI.* 2020:1–5. https://doi.org/10.1109/CCCI49893.2020.9256562.

29. Pinkas G, Karny Y, Malachi A, et al. SARS-CoV-2 detection from voice. *IEEE Open J Eng Med Biol.* 2020;1:268–274. https://doi.org/10.1109/OJEMB.2020.3026468.

30. Mendonça Alves L, Reis C, Pinheiro Â. Prosody and reading in dyslexic children. *Dyslexia.* 2015;21:35–49. https://doi.org/10.1002/dys.1485.

31. Goyal K, Singh A, Kadyan V. A comparison of Laryngeal effect in the dialects of Punjabi language. *JAIHC.* 2021. https://doi.org/10.1007/s12652-021-03235-4.

32. Nagumo R, Zhang Y, Ogawa Y, et al. Automatic detection of cognitive impairments through acoustic analysis of speech. *Curr Alzheimer Res.* 2020;17:60–68. https://doi.org/10.2174/1567205017666200213094513.

33. Procter T, Joshi A. Cultural competency in voice evaluation: considerations of normative standards for sociolinguistically diverse voices. *J Voice.* 2020;S0892-1997. https://doi.org/10.1016/j.jvoice.2020.09.025. 30369-6.

34. Rojas S, Kefalianos E, Vogel A. How does our voice change as we age? A systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age. *J Speech Lang Hear Res.* 2020;63:533–551. https://doi.org/10.1044/2019_JSLHR-19-00099.

35. Sun L. Using prosodic and acoustic features for Chinese dialects identification. *Proc IPMV.* 2020:118–123. https://doi.org/10.1145/3421558.3421577.

36. Taylor S, Dromey C, Nissen SL, et al. Age-related changes in speech and voice: spectral and cepstral measures. *J Speech Lang Hear Res.* 2020;63:647–660. https://doi.org/10.1044/2019_JSLHR-19-00028.

37. Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol.* 2021;19:141–154. https://doi.org/10.1038/s41579-020-00459-7.

38. Tu H, Tu S, Gao S, et al. Current epidemiological and clinical features of COVID-19; a global perspective from China. *J Infect.* 2020;81:1–9. https://doi.org/10.1016/j.jinf.2020.04.011.

39. McCool FD. Global physiology and pathophysiology of cough: ACCP evidence-based clinical practice guidelines. *Chest.* 2006;129(1 suppl):48S–53S. https://doi.org/10.1378/chest.129.1_suppl.48S.

40. Kelemen SA, Cseri T, Marozsan I. Information obtained from tussigrams and the possibilities of their application in medical practice. *Bull Eur Physiopathol Respir.* 1987;23(Suppl 10):51s–56s.

41. Lee KK, Davenport PW, Smith JA, et al. CHEST Expert Cough Panel. Global physiology and pathophysiology of cough: part 1: cough phenomenology - CHEST guideline and expert panel report. *Chest.* 2021;159:282–293. https://doi.org/10.1016/j.chest.2020.08.2086.

42. Lee KK, Matos S, Ward K, et al. Sound: a non-invasive measure of cough intensity. *BMJ Open Respir Res.* 2017;4: e000178. https://doi.org/10.1136/bmjresp-2017-000178.

43. Murata A, Taniguchi Y, Hashimoto Y, et al. Discrimination of productive and non-productive cough by sound analysis. *Intern Med.* 1998;37:732–735. https://doi.org/10.2169/internalmedicine.37.732.

44. Hashimoto Y, Murata A, Mikami M, et al. Influence of the rheological properties of airway mucus on cough sound generation. *Respirology.* 2003;8:45–51. https://doi.org/10.1046/j.1440-1843.2003.00432.x.

45. Chatrzarrin H, Arcelus A, Goubran R, et al. Feature extraction for the differentiation of dry and wet cough sounds. *Proc MEMEA.* 2011:162–166. https://doi.org/10.1109/MeMeA.2011.5966670.

46. Knocikova J, Korpas J, Vrabec M, et al. Wavelet analysis of voluntary cough sound in patients with respiratory diseases. *J Physiol Pharmacol.* 2008;59(Suppl 6):331–340.

47. Nemati E, Rahman MJ, Blackstock E, et al. Estimation of the lung function using acoustic features of the voluntary cough. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020:4491–4497. https://doi.org/10.1109/EMBC44109.2020.9175986.

48. Infante C, Chamberlain DB, Kodgule R, et al. Classification of voluntary coughs applied to the screening of respiratory disease. *Annu Int Conf IEEE Eng Med Biol Soc.* 2017:1413–1416. https://doi.org/10.1109/EMBC.2017.8037098.

49. Infante C, Chamberlain D, Fletcher R, et al. Use of cough sounds for diagnosis and screening of pulmonary disease. *Proc IEEE GHTC.* 2017:1–10. https://doi.org/10.1109/GHTC.2017.8239338.

50. Sharan RV, Abeyratne UR, Swarnkar VR, et al. Predicting spirometry readings using cough sound features and regression. *Physiol Meas.* 2018;39: 095001. https://doi.org/10.1088/1361-6579/aad948.

51. Parker D, Picone J, Harati A, et al. Detecting paroxysmal coughing from pertussis cases using voice recognition technology. *PLoS One.* 2013;8:e82971. https://doi.org/10.1371/journal.pone.0082971.

52. Nemati E, Rahman MM, Nathan V, et al. A comprehensive approach for classification of the cough type. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020:208–212. https://doi.org/10.1109/EMBC44109.2020.9175345.

53. Zhu C, Liu B, Li P, et al. Automatic classification of dry cough and wet cough based on improved reverse Mel frequency cepstrum coefficients. *J Biomed Eng.* 2016;33:239–243. https://doi.org/10.7507/1001-5515.20160042.

54. Alsabek MB, Shahin I, Hassan A. Studying the similarity of COVID-19 sounds based on correlation analysis of MFCC. *Proc CCCI.* 2020:1–5.

55. Cohen-McFarlane M, Goubran R, Knoefel F. Novel coronavirus cough database: NoCoCoDa. *IEEE Access.* 2020;8:154087–154094.

56. Mouawad P, Dubnov T, Dubnov S. Robust detection of COVID-19 in cough sounds: using recurrence dynamics and variable Markov model. *SN Comput Sci.* 2021;2:34. https://doi.org/10.1007/s42979-020-00422-6.

57. Coppock H, Gaskell A, Tzirakis P, et al. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. *BMJ Innov.* 2021;7:356–362. https://doi.org/10.1136/bmjinnov-2021-000668.

58. Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings in sustained vowels. *IEEE Open J Eng Med Biol.* 2020;1:275–281. https://doi.org/10.1109/OJEMB.2020.3026928.

59. Pahar M, Klopper M, Warren R, et al. COVID-19 cough classification using machine learning and global smartphone recordings. *Comput Biol Med.* 2021;135: 104572. https://doi.org/10.1016/j.compbiomed.2021.104572.

60. Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked.* 2020;20: 100378. https://doi.org/10.1016/j.imu.2020.100378.

61. Orlandic L, Teijeiro T, Atienza D. The COUGHVID crowdsourcing dataset: a corpus for the study of large-scale cough analysis algorithms. *Sci Data*. 2021;8:156. https://doi.org/10.1038/s41597-021-00937-4.

62. Orlandic L, Teijeiro T, Atienza D. The COUGHVID crowdsourcing dataset: a corpus for the study of large-scale cough analysis algorithms, v2.0. *Zenodo*. 2021. https://doi.org/10.5281/zenodo.4498364.

63. COUGHVID Web Interface. 2021. Accessed at: October 8, 2021. Accessed from: https://coughvid.epfl.ch/.

64. Muguli A, Pinto L, R N, et al. DiCOVA challenge: dataset, task, and baseline system for COVID-19 diagnosis using acoustics. *Proc Interspeech*. 2021:901–905. https://doi.org/10.21437/Interspeech.2021-74.

65. Bai H, Sha B, Xu X, et al. Gender difference in chronic cough: are women more likely to cough? *Front Physiol*. 2021;12: 654797. https://doi.org/10.3389/fphys.2021.654797.

66. Ioan I, Poussel M, Coutier L, et al. What is chronic cough in children? *Front Physiol*. 2014;5:322. https://doi.org/10.3389/fphys.2014.00322.

67. Eyben F, Wöllmer M, Schuller B. OpenSMILE: the Munich versatile and fast open-source audio feature extractor. *Proc ACM Multimedia*. 2010:1459–1462. https://doi.org/10.1145/1873951.1874246.

68. Schuller B, Steidl S, Batliner A, et al. The INTERSPEECH 2014 computational paralinguistics challenge: cognitive & physical load. *Proc Interspeech*. 2014:427–431.

69. Chang Y-W, Lin C-J. Feature ranking using linear SVM. *Proc WCCI Workshop on the Causation and Prediction Challenge*. 2008;3:53–64.

70. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46:389–422. https://doi.org/10.1023/A:1012487302797.

71. Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. *Proc ICML*. 2017:3145–3153.

72. Schwab P, Karlen W. CXPlain: causal explanations for model interpretation under uncertainty. *Proc NeurIPS*. 2019:10220–10230.

73. Cummins N, Pan Y, Ren Z, et al. A comparison of acoustic and linguistics methodologies for Alzheimer's dementia recognition. *Proc INTERSPEECH*. 2020:2182–2186. https://doi.org/10.21437/Interspeech.2020-2635.

74. Scikit-learn. 2022. Accessed at: March 3, 2022. Accessed from: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html/.

75. Brain D, Webb G. On the effect of data set size on bias and variance in classification learning. *Proc AKAW*. 2000:117–128 Accessed from: http://i.giwebb.com/wp-content/papercite-data/pdf/brainwebb99.pdf. Accessed 24 May 2022.

76. Afshinnekoo E, Ahsanuddin S, Mason CE. Globalizing and crowdsourcing biomedical research. *Br Med Bull*. 2016;120:27–33. https://doi.org/10.1093/bmb/ldw044.

77. Khare R, Good BM, Leaman R, et al. Crowdsourcing in biomedicine: challenges and opportunities. *Brief Bioinform*. 2016;17:23–32. https://doi.org/10.1093/bib/bbv021.

78. Porter ND, Verdery AM, Gaddis SM. Enhancing big data in the social sciences with crowdsourcing: data augmentation practices, techniques, and opportunities. *PLoS One*. 2020;15: e0233154. https://doi.org/10.1371/journal.pone.0233154.

79. Sharma N, Krishnan P, Kumar R, et al. Coswara − A database of breathing, cough, and voice sounds for COVID-19 diagnosis. *Proc INTERSPEECH*. 2020:4811–4815. https://doi.org/10.21437/Interspeech.2020-2768.

80. Han J, Qian K, Song M, et al. An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety. *ArXiv Preprint*. 2020. https://doi.org/10.48550/ARXIV.2005.00096.

81. Qian K, Schmitt M, Zheng HY, et al. Computer audition for fighting the SARS-CoV-2 corona crisis − Introducing the Multitask Speech Corpus for COVID-19. *IEEE Internet Things J*. 2021;8:16035–16046. https://doi.org/10.1109/JIOT.2021.3067605.

82. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?": explaining the predictions of any classifier. *Proc SIGKDD*. 2016:1135–1144. https://doi.org/10.1145/2939672.2939778.

83. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *Proc NIPS*. 2017. Accessed from: https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf. Accessed 24 May 2022.

84. Ren Z, Kong Q, Han J, et al. CAA-Net: conditional Atrous CNNs with attention for explainable device-robust acoustic scene classification. *IEEE Trans Multimedia*. 2020;23:4131–4142. https://doi.org/10.1109/TMM.2020.3037534.

85. Zhao Z, Bao Z, Zhao Y, et al. Exploring deep spectrum representations via attention-based recurrent and convolutional neural networks for speech emotion recognition. *IEEE Access*. 2019;7:97515–97525. https://doi.org/10.1109/ACCESS.2019.292862.