COVID-19 cough classification using machine learning and global smartphone recordings

Madhurananda Pahar [a,*], Marisa Klopper [b], Robin Warren [b], Thomas Niesler [a]

[a] *Department of Electrical and Electronic Engineering, Stellenbosch University, South Africa*
[b] *SAMRC Centre for Tuberculosis Research, DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa*

A R T I C L E   I N F O

A B S T R A C T

We present a machine learning based COVID-19 cough classifier which can discriminate COVID-19 positive coughs from both COVID-19 negative and healthy coughs recorded on a smartphone. This type of screening is non-contact, easy to apply, and can reduce the workload in testing centres as well as limit transmission by recommending early self-isolation to those who have a cough suggestive of COVID-19. The datasets used in this study include subjects from all six continents and contain both forced and natural coughs, indicating that the approach is widely applicable. The publicly available Coswara dataset contains 92 COVID-19 positive and 1079 healthy subjects, while the second smaller dataset was collected mostly in South Africa and contains 18 COVID-19 positive and 26 COVID-19 negative subjects who have undergone a SARS-CoV laboratory test. Both datasets indicate that COVID-19 positive coughs are 15%–20% shorter than non-COVID coughs. Dataset skew was addressed by applying the synthetic minority oversampling technique (SMOTE). A leave-p-out cross-validation scheme was used to train and evaluate seven machine learning classifiers: logistic regression (LR), k-nearest neighbour (KNN), support vector machine (SVM), multilayer perceptron (MLP), convolutional neural network (CNN), long short-term memory (LSTM) and a residual-based neural network architecture (Resnet50). Our results show that although all classifiers were able to identify COVID-19 coughs, the best performance was exhibited by the Resnet50 classifier, which was best able to discriminate between the COVID-19 positive and the healthy coughs with an area under the ROC curve (AUC) of 0.98. An LSTM classifier was best able to discriminate between the COVID-19 positive and COVID-19 negative coughs, with an AUC of 0.94 after selecting the best 13 features from a sequential forward selection (SFS). Since this type of cough audio classification is cost-effective and easy to deploy, it is potentially a useful and viable means of non-contact COVID-19 screening.

1. Introduction

COVID-19 (**CO**rona **VI**rus **D**isease of 20**19**), caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) virus, was declared a global pandemic on February 11, 2020 by the World Health Organisation (WHO). It is a new coronavirus but similar to other coronaviruses, including SARS-CoV (severe acute respiratory syndrome coronavirus) and MERS-CoV (Middle East respiratory syndrome coronavirus) which caused disease outbreaks in 2002 and 2012, respectively [1,2].

The most common symptoms of COVID-19 are fever, fatigue and dry coughs [3]. Other symptoms include shortness of breath, joint pain, muscle pain, gastrointestinal symptoms and loss of smell or taste [4]. At the time of writing, there were 142.1 million active cases of COVID-19 globally, and there had been 3 million deaths, with the USA reporting the highest number of cases (31.7 million) and deaths (567,729) [5]. The scale of the pandemic has caused some health systems to be overrun by the need for testing and the management of cases.

Several attempts have been made to identify early symptoms of COVID-19 through the use of artificial intelligence applied to images. The 50-layer residual neural network (Resnet50) architecture has been shown to perform better than other pre-trained models such as AlexNet, GoogLeNet and VGG16 in these tasks. For example, it has been demonstrated that COVID-19 can be detected from computed tomography (CT) images with an accuracy of 96.23% by using a Resnet50 architecture [6]. The same architecture was shown to detect pneumonia

---

* Corresponding author.
E-mail addresses: mpahar@sun.ac.za (M. Pahar), marisat@sun.ac.za (M. Klopper), rw1@sun.ac.za (R. Warren), trn@sun.ac.za (T. Niesler).
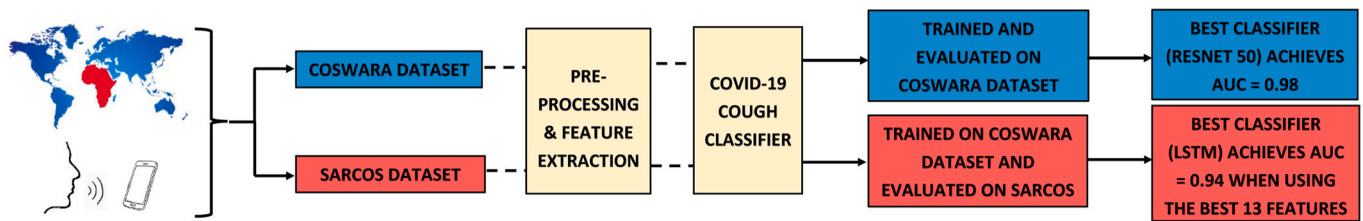
Fig. 1. Location of participants in the Coswara and the Sarcos datasets: Participants in the Coswara dataset were located on five different continents, excluding Africa. The majority (91%) of participants in the Coswara dataset are from Asia, as indicated in Fig. 2. Sarcos participants who supplied geographical information are mostly (75%) from South Africa, as shown in Fig. 3.

due to COVID-19 with an accuracy of 96.7% [7] and to detect COVID-19 from x-ray images with an accuracy of 96.30% [8].

Coughing is one of the predominant symptoms of COVID-19 [9] and also a symptom of more than 100 other diseases, and its effect on the respiratory system is known to vary [10]. For example, lung diseases can cause the airway to be either restricted or obstructed and this can influence the acoustics of the cough [11]. It has also been postulated that the glottis behaves differently under different pathological conditions [12,13] and this makes it possible to distinguish between coughs due to TB [14,15], asthma [16], bronchitis and pertussis (whooping cough) [17–20].

Respiratory data such as breathing, sneezing, speech, eating behaviour and coughing can be processed by machine learning algorithms to diagnose respiratory illness [21–23]. Simple machine learning tools, like binary classifiers, are able to distinguish COVID-19 respiratory sounds from healthy counterparts with an area under the ROC curve (AUC) exceeding 0.80 [24]. Detecting COVID-19 by analysing only the cough sounds is also possible. AI4COVID-19 is a mobile app that records 3 s of cough audio which is analysed automatically to provide an indication of COVID-19 status within 2 min [25]. A deep neural network (DNN) was shown to distinguish between COVID-19 and other coughs with an accuracy of 96.83% on a dataset containing 328 coughs from 150 patients of four different classes: COVID-19, asthma, bronchitis and healthy [26]. There appear to be unique patterns in COVID-19 coughs that allow a pre-trained Resnet18 classifier to identify COVID-19 coughs with an AUC of 0.72. In this case, cough samples were collected over the phone from 3621 individuals with confirmed COVID-19 [27]. COVID-19 coughs were classified with a higher AUC of 0.97 (sensitivity = 98.5% and specificity = 94.2%) by a Resnet50 architecture, trained on coughs from 4256 subjects and evaluated on 1064 subjects that included both COVID-19 positive and COVID-19 negative subjects by implementing four biomarkers [28]. A high AUC exceeding 0.98 was also achieved in Ref. [29] when discriminating COVID-19 positive coughs from COVID-19 negative coughs on a clinically validated dataset consisting of 2339 COVID-19 positive and 6041 COVID-19 negative subjects using DNN based classifiers.

Data collection from COVID-19 patients is challenging and the datasets are often not publicly available. Nevertheless, efforts have been made to compile such datasets. For example, a dataset consisting of coughing sounds recorded during or after the acute phase of COVID-19 from patients via public media interviews has been developed in Ref. [30]. The Coswara dataset is publicly available and collected in a more controlled and targeted manner [31]. At the time of writing, this dataset included useable 'deep cough' i.e. loud coughs recordings from 92 COVID-19 positive and 1079 healthy subjects. We have also begun to compile our own dataset by collecting recordings from subjects who have undergone a SARS-CoV laboratory test in South Africa. This Sarcos (**SAR**S **CO**VID-19 **S**outh Africa) dataset is currently still small and includes only 44 subjects (18 COVID-19 positive and 26 COVID-19 negative).

Both the Coswara and Sarcos datasets are imbalanced since COVID-19 positive subjects are outnumbered by non-COVID-19 subjects. Nevertheless, collectively these two datasets contain recordings from all six continents, as shown in Fig. 1. To improve machine learning

classification performance, we have applied the synthetic minority over-sampling technique (SMOTE) to balance our datasets. Furthermore, we have found that the COVID-19 positive coughs are 15%–20% shorter than non-COVID coughs. Hence, feature extraction is designed to preserve the time-domain patterns over an entire cough. Classifier hyperparameters were optimised by using the leave-*p*-out cross-validation, followed by training and evaluation of machine learning approaches, namely logistic regression (LR), k-nearest neighbour (KNN), support vector machine (SVM), multilayer perceptron (MLP) and deep neural networks (DNN) such as a convolutional neural network (CNN), long short-term memory (LSTM) and Resnet50. The Resnet50 produced the highest AUC of $0.976 \approx 0.98$ when trained and evaluated on the Coswara dataset, outperforming the baseline results presented in Ref. [32]. No classifier has been trained on the Sarcos dataset due to its small size. It can also not be combined with Coswara as it contains slightly different classes. Instead, this dataset has been used for an independent validation of the best-performing DNN classifiers developed on the Coswara dataset. In these validation experiments, it was found that the highest AUC of $0.938 \approx 0.94$ is achieved when using the best 13 features identified using the greedy sequential forward selection (SFS) algorithm and an LSTM classifier. We conclude that it is possible to identify COVID-19 on the basis of cough audio recorded using a smartphone. Furthermore, this discrimination between COVID-19 positive and both COVID-19 negative and healthy coughs is possible for audio samples collected from subjects located all over the world. Additional validation is however still required to obtain approval from regulatory bodies for use as a diagnostic tool.
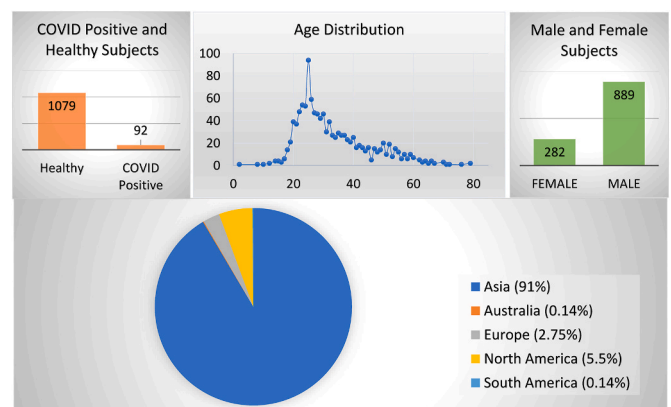
Fig. 2. Coswara dataset at the time of experimentation: There are 1079 healthy and 92 COVID-19 positive subjects in the pre-processed dataset, used for feature extraction and classifier training. Most of the subjects are aged between 20 and 50. There are 282 female and 889 male subjects and most of them are from Asia. Subjects are from five continents: **Asia** (Bahrain, Bangladesh, China, India, Indonesia, Iran, Japan, Malaysia, Oman, Philippines, Qatar, Saudi Arabia, Singapore, Sri Lanka, United Arab Emirates), **Australia**, **Europe** (Belgium, Finland, France, Germany, Ireland, Netherlands, Norway, Romania, Spain, Sweden, Switzerland, Ukraine, United Kingdom), **North America** (Canada, United States), and **South America** (Argentina, Mexico).
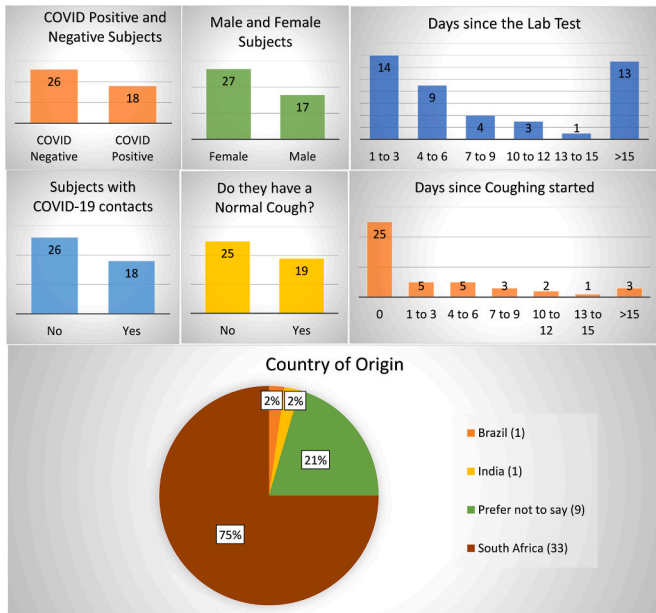
Fig. 3. Sarcos dataset at the time of experimentation: There are 26 COVID-19 negative and 18 COVID-19 positive subjects in the processed dataset. Unlike the Coswara dataset, there are more female than male subjects. Most of the subjects had their lab test performed within two weeks of participation. Only 19 of the subjects reported coughing as a symptom, and for these the reported duration of coughing symptoms was variable. There were 33 subjects from **Africa** (South Africa), 1 from **South America** (Brazil), 1 from **Asia** (India) and the rest declined to specify their geographic location.

2. Data

We have used two datasets in our experimental evaluation: the Coswara dataset and the Sarcos dataset.

2.1. The Coswara dataset

The Coswara project is aimed at developing a diagnostic tool for COVID-19 based on respiratory, cough and speech sounds [31]. Public participants were asked to contribute cough recordings via a web-based data collection platform using their smartphones (https://coswara.iisc.ac.in). The collected audio data includes fast and slow breathing, deep and shallow coughing, phonation of sustained vowels and spoken digits.

Age, gender, geographical location, current health status and pre-existing medical conditions are also recorded. Health status includes 'healthy', 'exposed', 'cured' or 'infected'. Audio recordings were sampled at 44.1 KHz and subjects were from all continents except Africa, as shown in Fig. 2. In this study, we have made use of the raw audio recordings and applied pre-processing as described in Section 2.3.

2.2. The Sarcos dataset

A similar initiative in South Africa encouraged participants to allow the voluntarily recording their coughs using an online platform (https://coughtest.online) under the research project name: 'COVID-19 screening by cough sound analysis'. Only coughs were collected as audio samples, and only subjects who had recently undergone a SARS-CoV laboratory test were asked to participate. The sampling rate for the audio recordings was 44.1 KHz. In addition to the cough audio recordings, subjects were presented with a voluntary and anonymous questionnaire, providing informed consent. The questionnaire prompted for the following information.

- Age and gender.
- Whether tested by an authorised COVID-19 testing centre.
- Days since the test was performed.
- Lab result (COVID-19 positive or negative).
- Country of residence.
- Known contact with COVID-19 positive patient.
- Known lung disease.
- Symptoms and temperature.
- Whether they are a regular smoker.
- Whether they have a current cough and for how many days.

Among the 44 participants, 33 (75%) subjects asserted that they are South African residents and therefore represent the African continent, as shown in Fig. 3. As there were no subjects from Africa in the Coswara dataset, together the Coswara and Sarcos dataset include subjects from all six continents.

2.3. Data pre-processing

The raw cough audio recordings from both datasets have the sampling rate (μ) of 44.1 KHz and is subjected to some simple pre-processing steps, described below. We note, time-window length (λ) as 0.05 s and amplitude threshold value (Φ) as 0.005, where both of these values were determined manually and interactively, as the silence removal was validated by visual inspection in all cases.
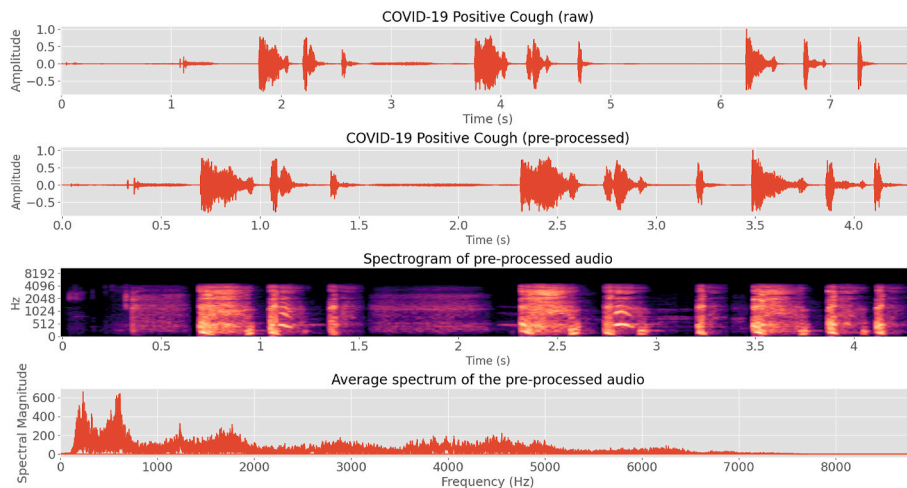

Fig. 4. A processed COVID-19 cough audio which is shorter than the original cough audio but keeps all spectrum resolution. Amplitudes are normalised and extended silences are removed in the pre-processing.

Table 1

Summary of the Coswara and Sarcos Datasets: In the Coswara dataset, there were 1171 subjects with useable 'deep cough' recordings, 92 of whom were COVID-19 positive while 1079 were healthy. This amounts to a total of 1.05 h of cough audio recordings (after pre-processing) that will be used for experimentation. The Sarcos dataset contains data from a total of 44 subjects, 18 of whom are COVID-19 positive and 26 who are not. This amounts to a total of 2.45 min of cough audio recordings (after pre-processing) that has been used for experimentation. COVID-19 positive coughs are 15%–20% shorter than non-COVID coughs.

Dataset	Label	Subjects	Total audio	Average per subject	Standard deviation
Coswara	COVID-19 Positive	92	4.24 min	2.77 s	1.62 s
	Healthy	1079	0.98 h	3.26 s	1.66 s
	Total	1171	1.05 h	3.22 s	1.67 s
Sarcos	COVID-19 Positive	18	0.87 min	2.91 s	2.23 s
	COVID-19 Negative	26	1.57 min	3.63 s	2.75 s
	Total	44	2.45 min	3.34 s	2.53 s

The original cough audio $c_i(t)$ is normalised by following Equation (1).

$$c_i(t) = 0.9 \times \frac{c_i(t)}{|max(c_i(t))|} \tag{1}$$

The processed final cough audio is shown in Fig. 4 and noted as: $C(t)$. Here, I denotes the time-window and we define:

$$C_I(t) = C_{j\mu\lambda}(t)\cdots C_{(j+1)\mu\lambda}(t) \tag{2}$$

For example, when $j = 0$; C_I will be the portion of signal where $C_0\cdots C_{2205}$, as $\mu = 44100$ Hz and $\lambda = 0.05$ s. $0 \leq j \leq \left\lfloor \frac{\Xi}{\mu\lambda} \right\rfloor$, where Ξ is the length of signal $c_i(t)$. $C(t)$ is calculated by following Equation (3).

$$C(t) = C(t) \oplus C_I\{ \text{ if } C_I(t) \geq \Phi \tag{3}$$

where, $\oplus$ means concatenation and, $C_I(t) \geq \Phi$, if $C_{I_i}(t) \geq \Phi$, where $\forall i \in I$.

Thus, the amplitudes of the raw audio data in the Coswara and the Sarcos dataset were normalised, after which periods of silence were removed from the signal to within a 50 ms margin using a simple energy detector. Fig. 4 shows an example of the original raw audio, as well as the pre-processed audio.

After pre-processing, the Coswara dataset contains 92 COVID-19 positive and 1079 healthy subjects and the Sarcos dataset contains 18 COVID-19 positive and 26 COVID-19 negative subjects, as summarised in Table 1. In both datasets, COVID-19 positive coughs are 15%–20% shorter than non-COVID coughs.

2.4. Dataset balancing

Table 1 shows that COVID-19 positive subjects are underrepresented in both datasets. To compensate for this imbalance, which can detrimentally affect machine learning [33,34], we have applied SMOTE data balancing to create equal number of COVID-19 positive coughs during training [35,36]. This technique has previously been successfully applied to cough detection and classification based on audio recordings [15,18,37].

SMOTE oversamples the minor class by generating synthetic examples, instead of for example random oversampling. In our dataset, for each COVID-19 positive cough, 5 other COVID-19 positive coughs were randomly chosen and the closest in terms of the Euclidean distance is identified as $\mathbf{x}^{NN}$. Then the synthetic COVID-19 positive samples are created using Equation (4).

$$\mathbf{x}^{SMOTE} = \mathbf{x} + u\cdot(\mathbf{x}^{NN} - \mathbf{x}) \tag{4}$$
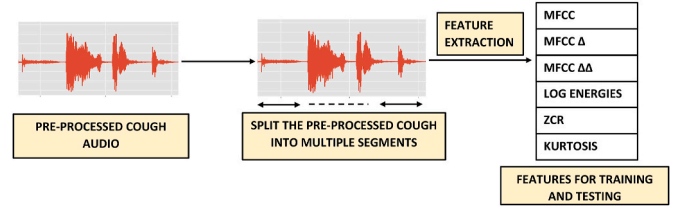

Fig. 5. Feature Extraction: Pre-processed cough audio recordings, shown in Fig. 4, are split into individual segments after which features such as MFCCs, MFCCs velocity (Δ), MFCCs acceleration (ΔΔ), log frame energies, ZCR and kurtosis are extracted. So, for $\mathcal{M}$ number of MFCCs and $\mathcal{S}$ number of segments, the final feature matrix has $(3\mathcal{M} + 3, \mathcal{S})$ dimensions.

Table 2

Feature extraction hyperparameters optimised using the leave-*p*-out cross-validation as described in Section 5.2

Hyperparameter	Description	Range
MFCC ($\mathcal{M}$)	Number of lower-order MFCCs to keep	$13 \times k_1$, where $k_1 = 1, 2, 3, 4, 5$
Frame ($\mathcal{F}$)	Frame-size in which audio is segmented	2^{k_2} where $k_2 = 8, \ldots, 12$
Seg ($\mathcal{S}$)	Number of frames extracted from the audio	$10 \times k_3$, where $k_3 = 5, 7, 10, 12, 15$

The multiplicative factor u is uniformly distributed between 0 and 1 [38].

We have also implemented other extensions of SMOTE such as borderline-SMOTE [39,40] and adaptive synthetic sampling [41]. However, the best results were obtained by using SMOTE without any modification.

3. Feature extraction

The feature extraction process is illustrated in Fig. 5. Features such as mel-frequency cepstral coefficients (MFCCs), log frame energies, zero crossing rate (ZCR) and kurtosis are extracted. MFCCs have been used very successfully as features in audio analysis and especially in automatic speech recognition [42,43]. They have also been found to be useful in differentiating dry coughs from wet coughs [44] and classifying tuberculosis coughs [45]. We have used the traditional MFCC extraction method considering higher resolution MFCCs along with the velocity (first-order difference, Δ) and acceleration (second-order difference, ΔΔ) as adding these has shown classifier improvement in the past [46]. Log frame energies can improve the performance in audio classification tasks [47]. The ZCR [48] is the number of times a signal changes sign within a frame, indicating the variability present in the signal. The kurtosis [49] indicates the tailedness of a probability density. For the samples of an audio signal, it indicates the prevalence of higher amplitudes. These features have been extracted by using the hyperparameters described in Table 2 for all cough recordings.

We have extracted features in a way that preserves the information regarding the beginning and the end of a cough event to allow time-domain patterns in the recordings to be discovered while maintaining the fixed input dimensionality expected by, for example, a CNN. From every recording, we extract a fixed number of features $\mathcal{S}$ by distributing the fixed-length analysis frames uniformly over the time-interval of the cough. The input feature matrix for the classifiers then always has the dimension of $(3\mathcal{M} + 3, \mathcal{S})$ for $\mathcal{M}$ number of MFCCs along with $\mathcal{M}$ number of velocity (Δ) and $\mathcal{M}$ number of acceleration (ΔΔ), as illustrated in Fig. 5. If Λ is the number of samples in the cough audio, we can calculate the number of samples between consecutive frames δ using Equation (5).

Table 3
Classifier hyperparameters, optimised using the leave-*p*-out cross-validation as described in Section 5.2

Hyperparameter	Description	Classifier	Range
ν_1	Regularisation strength	LR	10^{i_1} where $i_1 = -7, -6, \ldots,$ $6, 7$ (10^{-7} to 10^7)
ν_2	*l*1 penalty	LR	0 to 1 in steps of 0.05
ν_3	*l*2 penalty	LR	0 to 1 in steps of 0.05
ξ_1	Number of neighbours	KNN	10 to 100 in steps of 10
ξ_2	Leaf size	KNN	5 to 30 in steps of 5
ζ_1	Regularisation strength	SVM	10^{i_3} where $i_3 = -7, -6, \ldots,$ $6, 7$ (10^{-7} to 10^7)
ζ_2	Kernel Coefficient	SVM	10^{i_4} where $i_4 = -7, -6, \ldots,$ $6, 7$ (10^{-7} to 10^7)
η_1	No. of neurons	MLP	10 to 100 in steps of 10
η_2	*l*2 penalty	MLP	10^{i_2} where $i_2 = -7, -6, \ldots,$ $6, 7$ (10^{-7} to 10^7)
η_3	Stochastic gradient descent	MLP	0 to 1 in steps of 0.05
α_1	No. of Conv filters	CNN	3×2^{k_4} where $k_4 = 3, 4, 5$
α_2	Kernel size	CNN	2 and 3
α_3	Dropout rate	CNN, LSTM	0.1 to 0.5 in steps of 0.2
α_4	Dense layer size	CNN, LSTM	2^{k_5} where $k_5 = 4, 5$
β_1	LSTM units	LSTM	2^{k_6} where $k_6 = 6, 7, 8$
β_2	Learning rate	LSTM	10^{k_7} where $k_7 = -2, -3, -4$
β_3	Batch Size	CNN, LSTM	2^{k_8} where $k_8 = 6, 7, 8$
β_4	No. of epochs	CNN, LSTM	10 to 250 in steps of 20

$$\delta = \left\lceil \frac{\Lambda}{\mathcal{S}} \right\rceil \tag{5}$$

So, for example a 2.2 s long cough audio event contains 97020 samples, as the sampling rate is 44.1 KHz. If the frame length is 1024 samples and number of segments are 100, then the frame skip (δ) is $\left\lceil \frac{97020}{100} \right\rceil = 971$ samples.

In contrast with the more conventionally applied fixed frame rates, this way of extracting features ensures that the entire recording is captured within a fixed number of frames, allowing especially the CNN classifiers to discover more useful temporal patterns and provide better classification performance. This particular method of feature extraction has also shown promising result in classifying COVID-19 breath and speech [37].

4. Classifier architectures

We have trained and evaluated seven machine learning classifiers in total. LR models have been found to outperform other more complex classifiers such as classification trees, random forests, SVM in some clinical prediction tasks [14,50,51]. We have used gradient descent weight regularisation as well as lasso (*l*1 penalty) and ridge (*l*2 penalty) estimators during training [52,53]. This LR classifier has been intended primarily as a baseline against which any improvements offered by the more complex architectures can be measured. A KNN classifier bases its decision on the class labels of the *k* nearest neighbours in the training set and in the past has been able to both detect [54–56] and classify [17,45, 57] sounds such as coughs and snores successfully. SVM classifiers have also performed well in both detecting [58,59] and classifying [60] cough events. The independent term in kernel functions is chosen as a hyperparameter while optimising the SVM classifier. An MLP, a neural network with multiple layers of neurons separating the input and output [61], is capable of learning non-linear relationships and have for
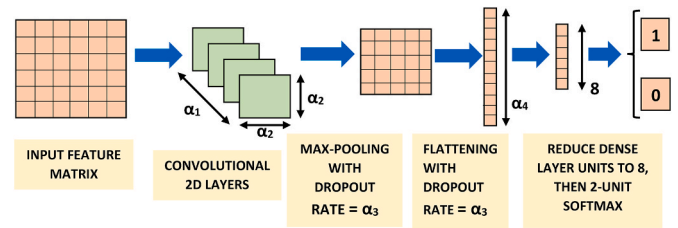

Fig. 6. CNN Classifier: Our CNN classifier uses α_1 two-dimansional convolutional layers with kernel size α_2, rectified linear units as activation functions and a dropout rate of α_3. After max-pooling, two dense layers with α_4 and 8 units respectively and rectified linear activation functions follow. The network is terminated by a two-dimensional softmax where one output (1) represents the COVID-19 positive class and the other (0) healthy or COVID-19 negative class. During training, features are presented to the neural network in batches of size β_3 for β_4 epochs.

Fig. 7. LSTM classifier: Our LSTM classifier has β_1 LSTM units, each with rectified linear activation functions and a dropout rate of α_3. This is followed by two dense layers with α_4 and 8 units respectively and rectified linear activation functions. The network is terminated by a two-dimensional softmax where one output (1) represents the COVID-19 positive class and the other (0) healthy or COVID-19 negative class. During training, features are presented to the neural network in batches of size β_3 for β_4 epochs.
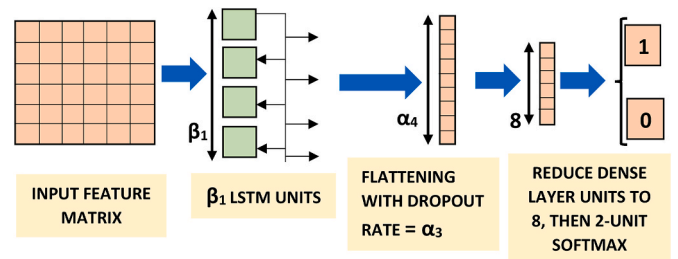
example been shown to be effective when discriminating influenza coughs from other coughs [62]. MLP have also been applied to classify tuberculosis coughs [45,59] and detect coughs in general [63,64]. The penalty ratios, along with the number of neurons are used as the hyperparameters which were optimised using the leave-*p*-out cross-validation process (Fig. 8 and Section 5.2).

A CNN is a popular deep neural network architecture, primarily used in image classification [65]. For example, in the past two decades CNNs were applied successfully to complex tasks such as face recognition [66]. It has also performed well in classifying COVID-19 breath and speech [37]. A CNN architecture [67,68] along with the optimised hyperparameters (Table 3) is shown in Fig. 6. An LSTM model is a type of recurrent neural network whose architecture allows it to remember previously-seen inputs when making its classification decision [69]. It has been successfully used in automatic cough detection [15,70], and also in other types of acoustic event detection [71,72]. The hyperparameters optimised for the LSTM classifier [73] are mentioned in Table 3 and visually explained in Fig. 7. The 50-layer deep residual learning (Resnet50) neural network [74] is a very deep architecture that contains skip layers, and has been found to outperform other very deep architectures such as VGGNet. It performs particularly well on image classification tasks on the dataset such as ILSVRC, the CIFAR10 dataset and the COCO object detection dataset [75]. Resnet50 has already been used in successfully detecting COVID-19 from CT images [6], coughs [28], breath, speech [37] and Alzheimer's [76]. Due to extreme computation load, we have used the default Resnet50 structure mentioned in Table 1 of [74].
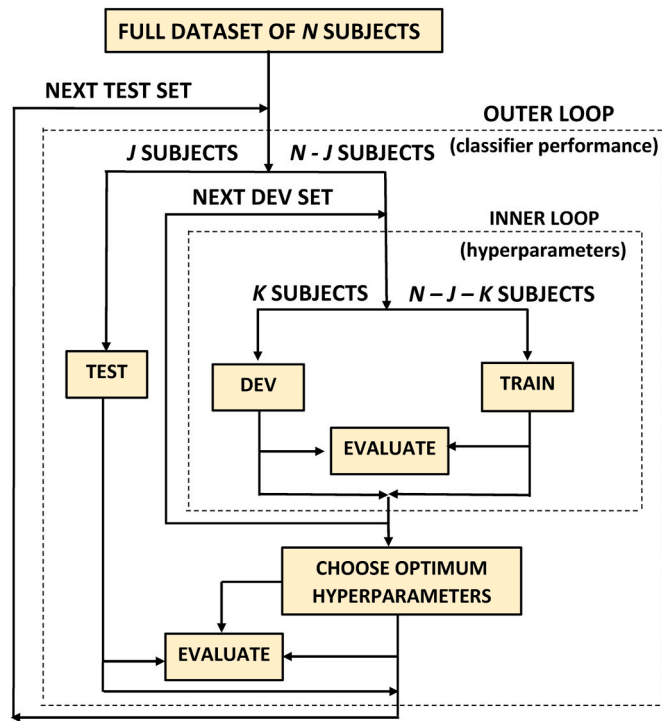
Fig. 8. **Leave p-out cross-validation**, used to train and evaluate the classifiers. The development set (DEV) consisting K subjects has been used to optimise the hyperparameters while training on the TRAIN set, consisted of $N - J - K$ subjects. The final evaluation of the classifiers in terms of the AUC occurs on the TEST set, consisting J subjects.

5. Classification process

5.1. Hyperparameter optimisation

Both feature extraction and classifier architectures have a number of hyperparameters. They are listed in Table 2 and 3 and were optimised by using a leave-p-out cross-validation scheme.

As the sampling rate is 44.1 KHz in both the Coswara and Sarcos dataset, by varying the frame lengths from 2^8 to 2^{12} i.e. 256 to 4096 samples, features are extracted from frames whose duration varies between approximately 5 and 100 ms? Different phases in a cough carry

important features [44] and thus each cough has been divided between 50 and 150 segments with steps of 20–30, as shown in Fig. 5. By varying the number of lower-order MFCCs to keep (from 13 to 65, with steps of 13), the spectral resolution of the features was varied.

5.2. Cross-validation

All our classifiers have been trained and evaluated by using a nested leave-p-out cross-validation scheme, as shown in Fig. 8 [77]. Since only the Coswara dataset was used for training and parameter optimisation, $N = 1171$ in Fig. 8. We have set the train and test split as 4 : 1; as this ratio has been used effectively in medical classification tasks [78]. Thus, $J = 234$ and $K = 187$ in our experiments.

The figure shows that, in an outer loop, J subjects are removed from the complete set of N subjects to be used for later independent testing. Then, a further K subjects are removed from the remaining $N - J$ subjects to serve as a development set to optimise the hyperparameters listed in Table 3. The inner loop considers all such sets of K subjects, and the optimal hyperparameters are chosen on the basis of all these partitions. The resulting optimal hyperparameters are used to train a final system on all $N - J$ subjects which is evaluated on the test set consisting of J subjects. If the $N - J$ subjects in the training portion contain C_1 COVID-19 positive and C_2 COVID-19 negative coughs, then $(C_2 - C_1)$ synthetic COVID-19 positive coughs are created by using SMOTE. AUC has always been the optimisation criterion in this cross-validation. This entire procedure is repeated for all possible non-overlapping test sets in the outer loop. The final performance is evaluated by calculating and averaging AUC over these outer loops.

This cross-validation procedure makes the best use of our small dataset by allowing all subjects to be used for both training and testing purposes while ensuring unbiased hyperparameter optimisation and a strict per-subject separation between cross-validation folds.

5.3. Classifier evaluation

Receiver operating characteristic (ROC) curves were calculated within the inner and outer loops shown in Fig. 8. The area under the ROC curve (AUC) indicates how well the classifier has performed over a range of decision thresholds [79]. From these ROC curves, the decision that achieves an equal error rate (γ_{EE}) was computed. This is the threshold for which the difference between the classifier's true positive rate (TPR) and false positive rate (FPR) is minimised.

We note the mean per-frame probability that a cough is from a

Table 4
Classifier performance when training and evaluating on the Coswara dataset. The best two classifiers along with their feature extraction and optimal classifier hyperparameters are mentioned. The area under the ROC curve (AUC) has been the optimisation criterion during cross-validation. The mean specificity (spec), sensitivity (sens), accuracy (ACC) and standard deviation of AUC (σ_{AUC}) are also shown. The best performance is achieved by the Resnet50.

Classifier	Best Feature Hyperparameters	Optimal Classifier Hyperparameters (Optimised inside nested cross-validation)	Performance				
			Spec	Sens	ACC	AUC	σ_{AUC}
LR	$\mathcal{M} = 13, \mathcal{F} = 1024, \mathcal{S} = 120$	$\nu_1 = 10^{-4}, \nu_2 = 0.25, \nu_3 = 0.75$	57%	94%	75.70%	0.736	0.057
LR	$\mathcal{M} = 26, \mathcal{F} = 1024, \mathcal{S} = 70$	$\nu_1 = 10^{-2}, \nu_2 = 0.45, \nu_3 = 0.55$	59%	74%	66.32%	0.729	0.049
KNN	$\mathcal{M} = 26, \mathcal{F} = 2048, \mathcal{S} = 100$	$\xi_1 = 70, \xi_2 = 20$	65%	83%	74.70%	0.781	0.041
KNN	$\mathcal{M} = 26, \mathcal{F} = 1024, \mathcal{S} = 70$	$\xi_1 = 60, \xi_2 = 25$	64%	81%	73.80%	0.776	0.039
SVM	$\mathcal{M} = 39, \mathcal{F} = 2048, \mathcal{S} = 100$	$\zeta_1 = 10^{-2}, \zeta_2 = 10^{-3}$	74%	71%	72.28%	0.815	0.046
SVM	$\mathcal{M} = 26, \mathcal{F} = 1024, \mathcal{S} = 50$	$\zeta_1 = 10^{-4}, \zeta_2 = 10^2$	74%	74%	73.91%	0.804	0.051
MLP	$\mathcal{M} = 26, \mathcal{F} = 2048, \mathcal{S} = 100$	$\eta_1 = 40, \eta_2 = 10^{-3}, \eta_3 = 0.4$	87%	88%	87.51%	0.897	0.033
MLP	$\mathcal{M} = 13, \mathcal{F} = 1024, \mathcal{S} = 100$	$\eta_1 = 60, \eta_2 = 10^{-1}, \eta_3 = 0.55$	84%	68%	76.02%	0.833	0.041
CNN	$\mathcal{M} = 26, \mathcal{F} = 1024, \mathcal{S} = 70$	$\alpha_1 = 48, \alpha_2 = 2, \alpha_3 = 0.3, \alpha_4 = 16, \beta_3 = 128, \beta_4 = 130$	99%	90%	94.57%	0.953	0.039
CNN	$\mathcal{M} = 39, \mathcal{F} = 1024, \mathcal{S} = 50$	$\alpha_1 = 96, \alpha_2 = 2, \alpha_3 = 0.1, \alpha_4 = 16, \beta_3 = 256, \beta_4 = 170$	98%	90%	94.35%	0.950	0.039
LSTM	$\mathcal{M} = 13, \mathcal{F} = 2048, \mathcal{S} = 70$	$\beta_1 = 128, \beta_2 = 10^{-3}, \alpha_3 = 0.3, \alpha_4 = 32, \beta_3 = 256, \beta_4 = 150$	97%	91%	94.02%	0.942	0.043
LSTM	$\mathcal{M} = 26, \mathcal{F} = 2048, \mathcal{S} = 100$	$\beta_1 = 256, \beta_2 = 10^{-2}, \alpha_3 = 0.3, \alpha_4 = 16, \beta_3 = 256, \beta_4 = 110$	97%	90%	93.65%	0.931	0.041
Resnet50	*$\mathcal{M}=39, \mathcal{F}=1024, \mathcal{S}=50$*	*Default Resnet50 (Table 1 in Ref. [74])*	*98%*	*93%*	*95.33%*	*0.976*	*0.018*
Resnet50	$\mathcal{M} = 26, \mathcal{F} = 1024, \mathcal{S} = 70$	Default Resnet50 (Table 1 in Ref. [74])	98%	93%	95.01%	0.963	0.011

Table 5

Classifier performance when training on the Coswara dataset and evaluating on the Sarcos dataset. The best performance was achieved by the LSTM classifier, and further improvements were achieved by applying SFS.

Classifier	Best Feature Hyperparameters	Optimal Classifier Hyperparameters (trained on Coswara dataset in Table 4)	Performance			
			Spec	Sens	ACC	AUC
CNN	$\mathcal{M} = 26, \mathcal{F} = 1024, \mathcal{S} = 70$	$\alpha_1 = 48, \alpha_2 = 2, \alpha_3 = 0.3, \alpha_4 = 16, \beta_3 = 128, \beta_4 = 130$	61%	85%	73.02%	0.755
LSTM	$\mathcal{M} = 13, \mathcal{F} = 2048, \mathcal{S} = 70$	$\beta_1 = 128, \beta_2 = 10^{-3}, \alpha_3 = 0.3, \alpha_4 = 32, \beta_3 = 256, \beta_4 = 150$	73%	75%	73.78%	0.779
Resnet50	$\mathcal{M} = 39, \mathcal{F} = 1024, \mathcal{S} = 50$	Default Resnet50 (Table 1 in Ref. [74])	57%	93%	74.58%	0.742
LSTM + SFS	*$\mathcal{M}=13, \mathcal{F}=2048, \mathcal{S}=70$*	*$\beta_1=128, \beta_2=10^{-3}, \alpha_3=0.3, \alpha_4=32, \beta_3=256, \beta_4=150$*	*96%*	*91%*	*92.91%*	*0.938*

COVID-19 positive subject by $\widehat{P}$:

$$\widehat{P} = \frac{\sum_{i=1}^{K} P(Y = 1 | X_i, \theta)}{K} \tag{6}$$

where K indicates the number of frames in the cough and $P(Y = 1 | X_i, \theta)$ is the output of the classifier for feature vector X_i and parameters θ for the ith frame. Now we define the indicator variable C as:

$$C = \begin{cases} 1 & \text{if } \widehat{P} \geq \gamma_{EE} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

We then define two COVID-19 index scores ($COVID\_I_1$ and $COVID\_I_2$) in Equations (8) and (9) respectively.

$$COVID\_I_1 = \frac{\sum_{i=1}^{N_1} C}{N_1} \tag{8}$$

$$COVID\_I_2 = \frac{\sum_{i=1}^{N_2} P(Y = 1 | X_i)}{N_2} \tag{9}$$

In Equation (8), N_1 is the number of coughs from the subject in the recording while in Equation (9), N_2 indicates the total number of frames of cough audio gathered from the subject. Hence Equation (6) computes a per-cough average probability while Equation (9) computes a per-frame average probability. For the Coswara dataset, $N_1 = 1$.

The COVID-19 index scores, given by Equation (8) and 9, can both be used to make classification decisions. We have found that for some classifier architectures one will lead to better performance than the other. Therefore, we have made the choice of the scoring function an additional hyperparameter to be optimised during cross-validation.

We have calculated the specificity and sensitivity from these predicted values and then compared them with the actual values and finally calculated the AUC and used it as a method of evaluation. The mean specificity, sensitivity, accuracy and AUC along with the optimal hyperparameters for each classifier are shown in Tables 4 and 5.

6. Results

6.1. Coswara dataset

Classification performance for the Coswara dataset is shown in Table 4. The Coswara results are the average specificity, sensitivity, accuracy and AUC along with its standard deviation calculated over the outer loop test-sets during cross-validation. These tables also show the values of the hyperparameters which produce the highest AUC during cross-validation.

Table 4 shows that all seven classifiers can classify COVID-19 coughs and the Resnet50 classifier exhibits the best performance, with an AUC of 0.976 when using a 120-dimensional feature matrix consisting of 39 MFCCs with appended velocity and acceleration extracted from frames that are 1024 samples long and when grouping the coughs into 50 segments. The corresponding accuracy is 95.3% with sensitivity 93%
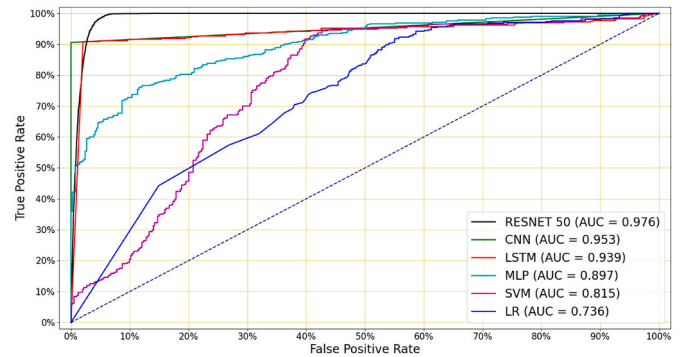

Fig. 9. Mean ROC curves for the classifiers trained and evaluated on the Coswara dataset: The highest AUC of 0.98 was achieved by the Resnet50, while the LR classifier has the lowest AUC of 0.74.

and specificity 98%. The CNN and LSTM classifiers also exhibited good performance, with AUCs of 0.953 and 0.942 respectively, thus comfortably outperforming the MLP, which achieved an AUC of 0.897. The optimised LR and SVM classifiers showed substantially weaker performance, with AUCs of 0.736 and 0.815 respectively. Table 4 also shows that DNN classifiers exhibit lower standard deviation across the folds than other classifiers. This suggests that DNN classifiers are also prone to perform better on new datasets without further hyperparameter optimisation.

The mean ROC curves for the optimised classifier of each architecture are shown in Fig. 9. We see that LSTM, CNN and Resnet50 classifiers achieve better performance than the remaining architectures at most operating points. Furthermore, the figure confirms that the Resnet50 architecture also in most cases achieved better classification performance than the CNN and LSTM. There appears to be a small region of the curve where the CNN outperforms the Resnet50 classifier, but this will need to be verified by future further experimentation with a larger dataset.

We also see from Table 4 that using a larger number of MFCCs consistently leads to improved performance. Since the spectral resolution used to compute the 39-dimensional MFCCs surpasses that of the human auditory system, we conclude that the classifiers are using information not generally perceivable to the human listeners. We have come to similar conclusions in previous work considering the classification of coughing sounds due to tuberculosis [14].

6.2. Sarcos dataset

Classification performance for the Sarcos dataset is shown in Table 5. Here the CNN, LSTM and Resnet50 classifiers trained on the Coswara dataset (as shown in Table 4) were tested on the 44 subjects in Sarcos dataset. No further hyperparameter optimisation was performed and hence Table 5 simply notes the same hyperparameters presented in Table 4. We see that performance has in all cases deteriorated relative to the better-matched Coswara dataset. The best performance was achieved by the LSTM classifier, which achieved an AUC of 0.779. In the
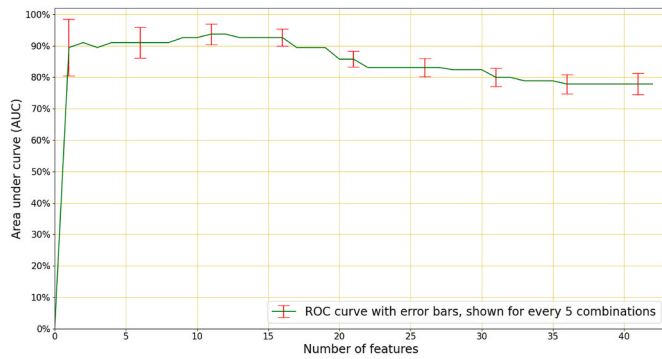
Fig. 10. Sequential Forward Selection, when applied to a feature matrix composed of 13 MFCCs with appended velocity (Δ) and acceleration (ΔΔ), log frame energies, ZCR and kurtosis (Equation (5)). Peak performance is observed after selecting the best 13 features.
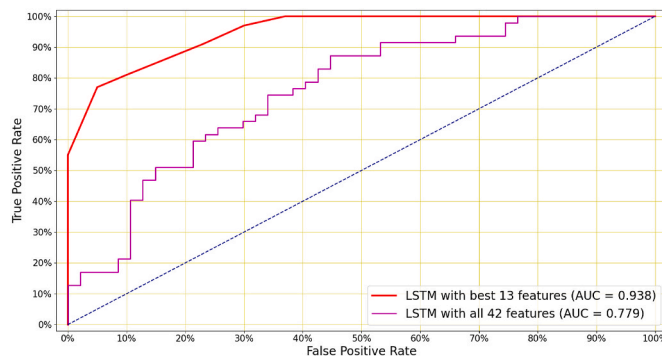

Fig. 11. Mean ROC curve for the best performed LSTM classifier trained on Coswara dataset and evaluated on Sarcos dataset: AUC of 0.78 has been achieved while using all 42 features. After applying SFS and selecting the best 13 features, the AUC has been improved to 0.94.

next section, we improve this classifier by applying feature selection.

6.2.1. Feature selection

As an additional experiment, SFS has been applied to the best-performing system in Table 5, the LSTM. SFS is a greedy selection method for the individual feature dimensions that contribute the most towards the classifier performance [80].

The feature selection hyperparameters in these experiments were 13 MFCCs extracted from 2048 samples (i.e. 0.46 s) long frames while coughs were grouped into 70 segments. Thus, SFS could select from a total of 42 features: MFCCs along with their velocity (Δ) and accelerations (ΔΔ), log frame energy, ZCR and Kurtosis (Equation (5)). After performing SFS to the LSTM classifier, a peak AUC of 0.938 was observed on the Sarcos dataset when using the best 13 features among those 42, as shown in Fig. 10 and Table 5. These 13 selected features led to an improvement of AUC from 0.779 to 0.938 (Fig. 11) and they include MFCCs ranging from 3 to 12 along with their velocity (Δ) and acceleration (ΔΔ), suggesting all dimensions of feature matrix carry equally-important COVID-19 signatures.

7. Conclusion and future work

We have developed COVID-19 cough classifiers using smartphone audio recordings and seven machine learning architectures. To train and evaluate these classifiers, we have used two datasets. The first, larger, dataset is publicly available and contains data from 1171 subjects (92 COVID-19 positive and 1079 healthy) residing on all five continents except Africa. The smaller second dataset contains recordings from 18

COVID-19 positive and 26 COVID-19 negative subjects, 75% of whom reside in South Africa. Thus, together the two datasets include data from subjects residing on all six continents. After pre-processing the cough audio recordings, we have found that the COVID-19 positive coughs are 15%–20% shorter than non-COVID coughs. Then we have extracted MFCCs, log frame energy, ZCR and kurtosis features from the cough audio using a special feature extraction technique which preserves the time-domain patterns and then trained and evaluated those seven classifiers using the nested leave-*p*-out cross-validation. Our best-performing classifier is the Resnet50 architecture and is able to discriminate between COVID-19 coughs and healthy coughs with an AUC of 0.98 on the Coswara dataset. These results outperform the baseline result of the AUC of 0.7 in Ref. [32]. When testing on the Sarcos dataset, the LSTM model trained on the Coswara dataset exhibit the best performance, discriminating COVID-19 positive coughs from COVID-19 negative coughs with an AUC of 0.94 while using the best 13 features determined by sequential forward selection (SFS). Furthermore, since better performance is achieved using a larger number of MFCCs than is required to mimic the human auditory system, we also conclude that at least some of the information used by the classifiers to discriminate the COVID-19 coughs and the non-COVID coughs may not be perceivable to the human ear.

Although the systems we describe require more stringent validation on a larger dataset, the results we have presented are very promising and indicate that COVID-19 screening based on automatic classification of coughing sounds is viable. Since the data has been captured on smartphones, and since the classifier can in principle also be implemented on such device, such cough classification is cost-efficient, easy to apply and deploy. Furthermore, it could be applied remotely, thus avoiding contact with medical personnel.

In ongoing work, we are continuing to enlarge our dataset and to apply transfer learning in order take advantage of the other larger datasets. We are also beginning to consider the best means of implementing the classifier on a readily-available consumer smartphone.

References

[1] WHO, et al., Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. http://www.who.int/csr/sars/country/table2004_04_21/en/index.html, 2003.

[2] R. Miyata, N. Tanuma, M. Hayashi, T. Imamura, J.-i. Takanashi, R. Nagata, A. Okumura, H. Kashii, S. Tomita, S. Kumada, et al., Oxidative stress in patients with clinically mild encephalitis/encephalopathy with a reversible splenial lesion (MERS), Brain Dev. 34 (2) (2012) 124–127.

[3] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China, J. Am. Med. Assoc. 323 (11) (2020) 1061–1069.

[4] A. Carfì, R. Bernabei, F. Landi, et al., Persistent symptoms in patients after acute COVID-19, J. Am. Med. Assoc. 324 (6) (2020) 603–605.

[5] John Hopkins University, COVID-19 dashboard by the center for systems science and engineering (CSSE). https://coronavirus.jhu.edu, Nov. 2020.

[6] S. Walvekar, D. Shinde, et al., Detection of COVID-19 from CT images using Resnet50, in: 2nd International Conference on Communication & Information Processing (ICCIP) 2020, 2020, https://doi.org/10.2139/ssrn.3648863. https://ssrn.com/abstract=3648863.

[7] H. Sotoudeh, M. Tabatabaei, B. Tasorian, K. Tavakol, E. Sotoudeh, A.L. Moini, Artificial intelligence empowers radiologists to differentiate pneumonia induced by COVID-19 versus influenza viruses, Acta Inf. Med. 28 (3) (2020) 190.

[8] M. Yildirim, A. Cinar, A deep learning based hybrid approach for COVID-19 disease detections, Trait. Du. Signal 37 (3) (2020) 461–468.

[9] A. Chang, G. Redding, M. Everard, Chronic wet cough: protracted bronchitis, chronic suppurative lung disease and bronchiectasis, Pediatr. Pulmonol. 43 (6) (2008) 519–531.

[10] T. Higenbottam, Chronic cough and the cough reflex in common lung diseases, Pulm. Pharmacol. Therapeut. 15 (3) (2002) 241–247.

[11] K.F. Chung, I.D. Pavord, Prevalence, pathogenesis, and causes of chronic cough, Lancet 371 (9621) (2008) 1364–1374.

[12] J. Korpáš, J. Sadloňová, M. Vrabec, Analysis of the cough sound: an overview, Pulm. Pharmacol. 9 (5–6) (1996) 261–268.

[13] J. Knocikova, J. Korpas, M. Vrabec, M. Javorka, Wavelet analysis of voluntary cough sound in patients with respiratory diseases, J. Physiol. Pharmacol. 59 (Suppl 6) (2008) 331–340.

[14] G. Botha, G. Theron, R. Warren, M. Klopper, K. Dheda, P. Van Helden, T. Niesler, Detection of tuberculosis by automatic cough sound analysis, Physiol. Meas. 39 (4) (2018) 045005.

[15] M. Pahar, I. Miranda, A. Diacon, T. Niesler, Deep neural network based cough detection using bed-mounted accelerometer measurements, in: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 8002–8006, https://doi.org/10.1109/ICASSP39728.2021.9414744.

[16] M. Al-khassaweneh, R. Bani Abdelrahman, A signal processing approach for the diagnosis of asthma from cough sounds, J. Med. Eng. Technol. 37 (3) (2013) 165–171.

[17] R.X.A. Pramono, S.A. Imtiaz, E. Rodriguez-Villegas, A cough-based algorithm for automatic diagnosis of pertussis, PloS One 11 (9) (2016), e0162128.

[18] 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 Journal of Biomedical and Health Informatics 23 (4) (2018) 1566–1573.

[19] R.V. Sharan, U.R. Abeyratne, V.R. Swarnkar, P. Porter, Automatic croup diagnosis using cough sound recognition, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 66 (2) (2018) 485–495.

[20] G. Rudraraju, S. Palreddy, B. Mamidgi, N.R. Sripada, Y.P. Sai, N.K. Vodnala, S.P. Haranath, Cough sound analysis and objective correlation with spirometry and clinical diagnosis, Informatics in Medicine Unlocked (2020) 100319.

[21] G. Deshpande, B. Schuller, An Overview on Audio, Signal, Speech, & Language Processing for COVID-19, arXiv preprint arXiv:2005.08579, 2020.

[22] A.N. Belkacem, S. Ouhbi, A. Lakas, E. Benkhelifa, C. Chen, End-to-End AI-based point-of-care diagnosis system for classifying respiratory illnesses and early detection of COVID-19: a theoretical framework, Front. Med. 8 (2021) 372.

[23] B.W. Schuller, D.M. Schuller, K. Qian, J. Liu, H. Zheng, X. Li, COVID-19 and Computer Audition: an Overview on what Speech & Sound Analysis Could Contribute in the SARS-CoV-2 Corona Crisis, arXiv preprint arXiv:2003.11117, 2020.

[24] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. 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 & Data Mining, 2020, pp. 3474–3484.

[25] A. Imran, I. Posokhova, H.N. Qureshi, U. Masood, S. Riaz, K. Ali, C.N. John, M. Nabeel, AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app, Informatics in Medicine Unlocked 20 (2020) 100378.

[26] A. Pal, M. Sankarasubbu, Pay attention to the cough: early diagnosis of COVID-19 using interpretable symptoms embeddings with cough sound signal processing. Proceedings of the 36th Annual ACM Symposium on Applied Computing, 2021, pp. 620–628.

[27] P. Bagad, A. Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A. Mahale, S. Rane, N. Agarwal, R. Panicker, Cough against COVID: Evidence of COVID-19 Signature in Cough Sounds, arXiv preprint arXiv:2009.08790, 2020.

[28] J. Laguarta, F. Hueto, B. Subirana, COVID-19 artificial intelligence diagnosis using only cough recordings, IEEE Open Journal of Engineering in Medicine and Biology 1 (2020) 275–281, https://doi.org/10.1109/OJEMB.2020.3026928.

[29] J. Andreu-Perez, H. Pérez-Espinosa, E. Timonet, M. Kiani, M.I. Giron-Perez, A. B. Benitez-Trinidad, D. Jarchi, A. Rosales, N. Gkatzoulis, O.F. Reyes-Galaviz, et al., A generic deep learning based cough analysis system from clinically validated samples for point-of-need covid-19 test and severity levels, IEEE Transactions on Services Computing (1) (2021) 1, https://doi.org/10.1109/TSC.2021.3061402, 1.

[30] M. Cohen-McFarlane, R. Goubran, F. Knoefel, Novel coronavirus cough database: NoCoCoDa, IEEE Access 8 (2020) 154087–154094.

[31] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S.R. Chetupalli, P.K. Ghosh, S. Ganapathy, et al., Coswara–A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis, arXiv preprint arXiv:2005.10548, 2020.

[32] A. Muguli, L. Pinto, N. Sharma, P. Krishnan, P.K. Ghosh, R. Kumar, S. Ramoji, S. Bhat, S.R. Chetupalli, S. Ganapathy, et al., Dicova Challenge: Dataset, Task, and Baseline System for COVID-19 Diagnosis Using Acoustics, arXiv preprint arXiv: 2103.09148, 2021.

[33] J. Van Hulse, T.M. Khoshgoftaar, A. Napolitano, Experimental perspectives on learning from imbalanced data, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 935–942.

[34] B. Krawczyk, Learning from imbalanced data: open challenges and future directions, Progress in Artificial Intelligence 5 (4) (2016) 221–232.

[35] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.

[36] G. Lemaître, F. Nogueira, C.K. Aridas, Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning, J. Mach. Learn. Res. 18 (1) (2017) 559–563.

[37] M. Pahar, T. Niesler, Machine Learning Based COVID-19 Detection from Smartphone Recordings: Cough, Breath and Speech, arXiv preprint arXiv: 2104.02477, 2021.

[38] L.L. Blagus, R, SMOTE for high-dimensional class-imbalanced data, BMC Bioinf. 14 (2013) 106, https://doi.org/10.1186/1471-2105-14-106.

[39] H. Han, W.-Y. Wang, B.-H. Mao, Borderline-Smote, A new over-sampling method in imbalanced data sets learning, in: International Conference on Intelligent Computing, Springer, 2005, pp. 878–887.

[40] H.M. Nguyen, E.W. Cooper, K. Kamei, Borderline over-sampling for imbalanced data classification, Int. J. Knowl. Eng. Soft Data Paradigms 3 (1) (2011) 4–21.

[41] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 1322–1328.

[42] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy, Kong-Pang Pun, An efficient MFCC extraction method in speech recognition, in: IEEE International Symposium on Circuits and Systems, 2006.

[43] M. Pahar, L.S. Smith, Coding and decoding speech using a biologically inspired coding system, in: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2020, 3025–3032, https://doi.org/10.1109/SSCI47803.2020.9308328.

[44] H. Chatrzarrin, A. Arcelus, R. Goubran, F. Knoefel, Feature extraction for the differentiation of dry and wet cough sounds. IEEE International Symposium on Medical Measurements and Applications, IEEE, 2011.

[45] M. Pahar, M. Klopper, B. Reeve, G. Theron, R. Warren, T. Niesler, Automatic Cough Classification for Tuberculosis Screening in a Real-World Environment, arXiv preprint arXiv:2103.13300, 2021.

[46] M.M. Azmy, Feature extraction of heart sounds using velocity and acceleration of MFCCs based on support vector machines, in: 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2017, pp. 1–4, https://doi.org/10.1109/AEECT.2017.8257736.

[47] S. Aydın, H.M. Saraoğlu, S. Kara, Log energy entropy-based EEG classification with multilayer neural networks in seizure, Ann. Biomed. Eng. 37 (12) (2009) 2626.

[48] R. Bachu, S. Kopparthi, B. Adapa, B.D. Barkana, Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy, in: Advanced Techniques in Computing Sciences and Software Engineering, Springer, 2010, pp. 279–282.

[49] L.T. DeCarlo, On the meaning and use of kurtosis, Psychol. Methods 2 (3) (1997) 292.

[50] E. Christodoulou, J. Ma, G.S. Collins, E.W. Steyerberg, J.Y. Verbakel, B. Van Calster, A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models, J. Clin. Epidemiol. 110 (2019) 12–22.

[51] S. Le Cessie, J.C. Van Houwelingen, Ridge estimators in logistic regression, J. Roy. Stat. Soc.: Series C (Applied Statistics) 41 (1) (1992) 191–201.

[52] Y. Tsuruoka, J. Tsujii, S. Ananiadou, Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty, in: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009, pp. 477–485.

[53] H. Yamashita, H. Yabe, An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization, SIAM J. Optim. 14 (2) (2003) 479–499.

[54] J. Monge-Álvarez, C. Hoyos-Barceló, P. Lesso, P. Casaseca-de-la Higuera, Robust detection of audio-cough events using local hu moments, IEEE Journal of Biomedical and Health Informatics 23 (1) (2018) 184–196.

[55] R.X.A. Pramono, S.A. Imtiaz, E. Rodriguez-Villegas, Automatic cough detection in acoustic signal using spectral features, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 7153–7156.

[56] S. Vhaduri, T. Van Kessel, B. Ko, D. Wood, S. Wang, T. Brunschwiler, Nocturnal cough and snore detection in noisy environments using smartphone-microphones, in: 2019 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2019, pp. 1–7.

[57] J.-C. Wang, J.-F. Wang, K.W. He, C.-S. Hsu, Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor, in: The 2006 IEEE International Joint Conference on Neural Network Proceedings, IEEE, 2006, pp. 1731–1735.

[58] V. Bhateja, A. Taquee, D.K. Sharma, Pre-processing and classification of cough sounds in noisy environment using SVM, in: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), IEEE, 2019, pp. 822–826.

[59] B.H. Tracey, G. Comina, S. Larson, M. Bravard, J.W. López, R.H. Gilman, Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 6017–6020.

[60] R.V. Sharan, U.R. Abeyratne, V.R. Swarnkar, P. Porter, Cough sound analysis for diagnosing croup in pediatric patients using biologically inspired features, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 4578–4581.

[61] H. Taud, J. Mas, Multilayer perceptron (MLP), in: Geomatic Approaches for Modeling Land Change Scenarios, Springer, 2018, pp. 451–455.

[62] L. Sarangi, M.N. Mohanty, S. Pattanayak, Design of MLP based model for analysis of patient suffering from influenza, Procedia Computer Science 92 (2016) 396–403.

[63] J.-M. Liu, M. You, Z. Wang, G.-Z. Li, X. Xu, Z. Qiu, Cough detection using deep neural networks, in: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2014, pp. 560–563.

[64] J. Amoh, K. Odame, DeepCough: a deep convolutional neural network in a wearable cough detection system, in: 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2015, pp. 1–4.

[65] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[66] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, Face recognition: a convolutional neural-network approach, IEEE Trans. Neural Network. 8 (1) (1997) 98–113.

[67] S. Albawi, T.A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in: 2017 International Conference on Engineering and Technology (ICET), IEEE, 2017, pp. 1–6.

[68] X. Qi, T. Wang, J. Liu, Comparison of support vector machine and softmax classifiers in computer vision, in: 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE), IEEE, 2017, pp. 151–155.

[69] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[70] I.D. Miranda, A.H. Diacon, T.R. Niesler, A comparative study of features for acoustic cough detection using deep architectures, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 2601–2605.

[71] E. Marchi, F. Vesperini, F. Weninger, F. Eyben, S. Squartini, B. Schuller, Non-linear prediction with LSTM recurrent neural networks for acoustic novelty detection, in: 2015 International Joint Conference on Neural Networks (IJCNN), IEEE, 2015, pp. 1–7.

[72] J. Amoh, K. Odame, Deep neural networks for identifying cough sounds, IEEE transactions on Biomedical Circuits and Systems 10 (5) (2016) 1003–1011.

[73] A. Sherstinsky, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, Phys. Nonlinear Phenom. 404 (2020) 132306.

[74] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[75] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: common objects in context, in: European Conference on Computer Vision, Springer, 2014, pp. 740–755.

[76] J. Laguarta, F. Hueto, P. Rajasekaran, S. Sarma, B. Subirana, Longitudinal Speech Biomarkers for Automated Alzheimer's Detection, 2020, https://doi.org/10.21203/rs.3.rs-56078/v1.

[77] S. Liu, Leave-*p*-out cross-validation test for uncertain verhulst-pearl model with imprecise observations, IEEE Access 7 (2019) 131705–131709.

[78] A. Rácz, D. Bajusz, K. Héberger, Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification, Molecules 26 (4) (2021) 1111, https://doi.org/10.3390/molecules26041111.

[79] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874.

[80] P.A. Devijver, J. Kittler, Pattern Recognition: A Statistical Approach, Prentice Hall, 1982.

Madhurananda Pahar received his BSc in Mathematics from the University of Calcutta, India, and his MSc in Computing for Financial Markets followed by his PhD in Computational Neuroscience from the University of Stirling, Scotland. Currently, he is working as a post-doctoral fellow at the University of Stellenbosch, South Africa. His research interests are in machine learning and signal processing for audio signals and smart sensors in bio-medicine such as the detection and classification of TB and COVID-19 coughs in a real-world environment.

Marisa Klopper is a researcher at the Division of Molecular Biology and Human Genetics of Stellenbosch University, South Africa. She holds a PhD in Molecular Biology from Stellenbosch University and her research interest is in TB and drug-resistant TB diagnosis, epidemiology and physiology. She has been involved in cough classification for the last 6 years, with application to TB and more recently COVID-19.

Robin Warren is the Unit Director of the South African Medical Research Council's Centre for Tuberculosis Research and Distinguished Professor at Stellenbosch University. He has a B2 rating by the National Research Council (NRF) and is a core member of the DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research and head the TB Genomics research thrust. He has published over 320 papers in the field of TB and have an average H-index (Scopus, Web of Science and Google Scholar) of 65.

Thomas Niesler obtained the B.Eng (1991) and M.Eng (1993) degrees in Electronic Engineering from the University of Stellenbosch, South Africa and a Ph.D. from the University of Cambridge, England, in 1998. He joined the Department of Engineering, University of Cambridge, as a lecturer in 1998 and subsequently the Department of Electrical and Electronic Engineering, University of Stellenbosch, in 2000, where he has been Professor since 2012. His research interests lie in the areas of signal processing, pattern recognition and machine learning.