COVID-19 Cough Classification using Machine Learning and Global Smartphone Recordings

Madhurananda Pahar, Marisa Klopper, Robin Warren, and Thomas Niesler

Abstract—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-1-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.

Index Terms—COVID-19, cough classification, machine learning, logistic regression (LR), k-nearest neighbour (KNN), support vector machine (SVM), multilayer perceptron (MLP), convolutional neural network (CNN), long short-term memory (LSTM), Resnet50

1 INTRODUCTION

COVID19 (COronavirus Disease of 2019), 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 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 that this makes it possible to distinguish between coughs due to TB [14][15], asthma [16], bronchitis and pertussis (whooping cough) [17][18][19][20].

Respiratory data such as breathing, sneezing, speech, eating behaviour and coughing can be processed by machine learning algorithms to diagnose respiratory illness [21][22][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 seconds of cough audio which is analysed automatically to provide...
an indication of COVID-19 status within 2 minutes [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 [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 [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 usable ‘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 (SARS-COV-19 South 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 Figure 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 ± 0.08 when trained and evaluated on the Coswara dataset, outperforming the baseline results presented in [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 ± 0.04 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.

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.isc.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 Figure 2. In this study, we have made use of the raw audio recordings and applied pre-processing as described in Section 2.3.
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).

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](https://coughtest.online)) under the research project name ‘COVID-19 screening by cough sound analysis’. This dataset will be referred to as ‘Sarcos’ (SAR COVID-19 South Africa). 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 Figure 3. As there were no subjects from Africa in the Coswara dataset, together the Coswara and Sarcos dataset include subjects from all six continents.

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.

2.3 Data Pre-processing

The raw cough audio recordings from both datasets have the sampling rate ($\mu$) of 44.1 KHz and is subjected to some simple pre-processing steps, described below. We note, time-window length ($\lambda$) as 0.05 seconds and amplitude threshold value ($\Phi$) 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.

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

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

(1)

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

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

(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 $\lambda = 0.05$ seconds. $0 \leq j \leq \lfloor \frac{\Xi}{\mu \lambda} \rfloor$, where $\Xi$ is the length of signal $c_i(t)$. $C(t)$ is calculated by following Equation 3.

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

(3)

where, $\oplus$ means concatenation and, $C_I(t) \geq \Phi$, if $C_I(t) \geq \Phi$, where $\forall t \in I$.

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. Figure 4 shows an example of the original raw audio, as well as the pre-processed audio.

![Processed COVID-19 Cough Audio](image)

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.

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 under-represented in both datasets. To compensate for this imbalance, which can detrimentally affect machine learning, we have applied SMOTE data balancing to create equal number of COVID-19 positive coughs during training. This technique has previously been successfully applied to cough detection and classification based on audio recordings.

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 synthetic examples, instead of for example random oversampling, were generated. Further details of the pre-processing are described in the supplementary material.

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

### 3 Feature Extraction

The feature extraction process is illustrated in Figure 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. They have also been found to be useful in differentiating dry coughs from wet coughs and classifying tuberculosis coughs.

We have used the traditional MFCC extraction method considering higher resolution MFCCs along with the velocity (first-order difference, $\Delta$) and acceleration (second-order difference, $\Delta\Delta$) as they have shown classifier improvement in the past. Log frame energies can improve the performance in audio classification tasks. The ZCR is the number of times a signal changes sign within a frame, indicating the variability present in the signal. The kurtosis 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 $S$ 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 $(3M + 3, S)$ for $M$ number of MFCCs along with $M$ number of velocity ($\Delta$) and $M$ number of acceleration ($\Delta\Delta$), as illustrated in Figure 5. If $A$ is the number of samples in the cough audio, we can calculate the number of samples between consecutive frames $\delta$ using Equation 5:

$$\delta = \left\lceil \frac{A}{S} \right\rceil$$

Fig. 5. Feature Extraction: Pre-processed cough audio recordings, shown in Figure 4 are split into individual segments after which features such as MFCCs, MFCCs velocity ($\Delta$), MFCCs acceleration ($\Delta\Delta$), log frame energies, ZCR and kurtosis are extracted. So, for $M$ number of MFCCs and $S$ number of segments, the final feature matrix has $(3M + 3, S)$ dimensions.
TABLE 1
Summary of the Coswara and Sarcos Datasets: In the COSWARA dataset, there were 1171 subjects with usable ‘deep cough’ recordings, 92 of whom were COVID-19 positive while 1079 were healthy. This amounts to a total of 1.05 hours 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 minutes 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 mins   | 2.77 sec            | 1.62 sec          |
| Coswara   | Healthy                | 1079     | 0.98 hours  | 3.26 sec            | 1.66 sec          |
| Coswara   | Total                  | 1171     | 1.05 hours  | 3.22 sec            | 1.67 sec          |
| Sarcos    | COVID-19 Positive      | 18       | 0.87 mins   | 2.91 sec            | 2.23 sec          |
| Sarcos    | COVID-19 Negative      | 26       | 1.57 mins   | 3.63 sec            | 2.75 sec          |
| Sarcos    | Total                  | 44       | 2.45 mins   | 3.34 sec            | 2.53 sec          |

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 (l1 penalty) and ridge (l2 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, 55, 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 has for 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 (Figure 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 Figure 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 Figure 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].
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 were optimised by using a leave-\(p\)-out cross-validation scheme.

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

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 to 30, as shown in Figure 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 Figure 8. Since only the Coswara dataset was used for training and parameter optimisation, \(N = 1171\) in Figure 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 2. 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 Figure 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 (\(\gamma_{EE}\)) 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 COVID-19 positive subject by \(\hat{P}\):

\[
\hat{P} = \frac{1}{K} \sum_{i=1}^{K} P(Y = 1|X_i, \theta)
\]

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 \(\theta\) for the \(i^{th}\) frame. Now we define the indicator variable \(C\) as:

\[
C = \begin{cases} 1 & \text{if } \hat{P} \geq \gamma_{EE} \\ 0 & \text{otherwise} \end{cases}
\]

We then define two COVID-19 index scores (\(COVID_{I_1}\) and \(COVID_{I_2}\)) in Equations 8 and 9 respectively.

\[
COVID_{I_1} = \frac{N_i}{N_1} \sum_{i=1}^{N_1} C
\]
The COVID-19 index scores, given by Equations 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 Table 3 and 4.

### 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 0.956 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% 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 Figure 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.

### 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 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.

The feature selection hyperparameters in these experiments were 13 MFCCs extracted from 2048 samples (i.e. 0.46 sec) long frames while coughs were grouped into 70 segments. Thus, SFS could select from a total of 42 features: 39 MFCCs along with their velocity (Δ), log frame energies, ZCR and kurtosis (Equation 4). Peak performance is observed after selecting the best 13 features.
MFCCs ranging from 3 to 12 along with their velocity ($\Delta$) and acceleration ($\Delta\Delta$), 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 [52]. 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.

Acknowledgements

This project was funded by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Intramural Post-doctoral programme, the South African National Treasury, as well as an EDCTP2 programme supported by the European Union (TMA2017CDF-1885). We would like to thank the South African Centre for High Performance Computing (CHPC) for providing computational resources on their Lengau cluster for this research, and gratefully acknowledge the support of Telcom South Africa. We would also like to thank Jordan Govendar and Rafeeq du Toit for their invaluable assistance in the Sarcos data collection.

The content and findings reported are the sole deduction, view and responsibility of the researchers and do not reflect the official position and sentiments of the SAMRC, EDCTP2, European Union or the funders.

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 and Development, vol. 34, no. 2, pp. 124–127, 2012.
[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,” JAMA, vol. 323, no. 11, pp. 1061–1069, 2020.
[4] A. Carfi, R. Bernabei, F. Landi et al., “Persistent symptoms in patients after acute COVID-19,” JAMA, vol. 324, no. 6, pp. 603–605, 2020.
[5] John Hopkins University. (2020, Nov.) COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE). John Hopkins University. [Online]. Available: https://coronavirus.jhu.edu
[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, May 2020, available at SSRN: https://ssrn.com/abstract=3648863 or http://dx.doi.org/10.2139/ssrn.3648863.
G. Rudraraju, S. Palreddy, B. Mamidgi, N. R. Sripada, Y. P. Sai, N. K. Vodnala, and S. P. Haranath, “Cough sound analysis and objective correlation with spirometry and clinical diagnosis,” Informatics in Medicine Unlocked, p. 100319, 2020.

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

A. N. Belkacem, S. Oubbi, A. Lakas, E. Benkhelifa, and 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,” Frontiers in Medicine, vol. 8, p. 372, 2021.

B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, and 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.

C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthansombat, D. Spathis, T. Xia, P. Cicuta, and 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.

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

P. Bagad, A. Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A. Mahale, S. Rane, N. Agarwal, and R. Panicker, “Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds,” arXiv preprint arXiv:2009.08790, 2020.

J. Laguarta, F. Hueto, and B. Subirana, “COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings,” IEEE Open Journal of Engineering in Medicine and Biology, 2020.

J. Andreu-Perez, H. Perez-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, no. 01, pp. 1–1, 2021.

M. Cohen-McFarlane, R. Goubran, and F. Knoefel, “Novel Coronavirus Cough Database: NoCoCoDa,” IEEE Access, vol. 8, pp. 154 087–154 094, 2020.

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

A. Muguli, L. Pinto, N. Sharma, P. Krishnan, P. K. Ghosh, R. Kumar, S. Ramoji, S. Bhat, S. R. Chetupalli, S. Ganapathy et al., “Dicovas: Challenge Dataset, task, and baseline system for COVID-19 diagnosis using acoustics,” arXiv preprint arXiv:2103.09148, 2021.

J. Van Hulse, T. M. Khoshgoftaar, and A. Napolitano, “Experimental perspectives on learning from imbalanced data,” in Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 935–942.
S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, J. Laguarta, F. Hueto, P. Rajasekaran, S. Sarma, and T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual E. Marchi, F. Vesperini, F. Weninger, F. Eyben, S. Squar.

S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[109] P. A. Devijver and J. Kittler, Pattern recognition: A statistical approach. Prentice Hall, 1982.
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.