

# A novel cough audio segmentation framework for COVID-19 detection

Alice E. Ashby  
University of Brighton  
East Sussex, BN2 4GJ  
United Kingdom  
A.Ashby1@uni.brighton.ac.uk

Julia A. Meister  
University of Brighton  
East Sussex, BN2 4GJ  
United Kingdom  
J.Meister@brighton.ac.uk

Goran Soldar  
University of Brighton  
East Sussex, BN2 4GJ  
United Kingdom  
G.Soldar@brighton.ac.uk

Khuong An Nguyen  
University of Brighton  
East Sussex, BN2 4GJ  
United Kingdom  
khuong@cantab.net

**Abstract:** Despite its potential, Machine Learning has played little role in the present pandemic, due to the lack of data (i.e., there were not many COVID-19 samples in the early stage). Thus, this paper proposes a novel cough audio segmentation framework that may be applied on top of existing COVID-19 cough datasets to increase the number of samples, as well as filtering out noises and uninformative data. We demonstrate the efficiency of our framework on two popular open datasets.

*Keywords:* audio pre-processing, audio noise filtering, cough segmentation, COVID-19 open datasets

## 1. INTRODUCTION

One of the biggest roadblocks to scientific progress is the restriction of access to data. In the context of Machine Learning (ML) and Deep Learning (DL) methodologies, research teams and scientists alike may not develop training models without data.

When the COVID-19 pandemic took the world by storm in 2020, researchers worldwide raced to develop Artificial Intelligence (AI) solutions to accompany existing detection methods including antigen, molecular, and serological testing (such as PCR, Lateral Flow Tests), and medical imaging (such as X-Ray chest scan, lung ultrasound). Studies have shown that coughs contain unique characteristics, making it possible for diagnosis of COVID-19 via analysis of coughs (Han et al., 2021; Pahar et al., 2021). However, existing respiratory datasets were not adequate enough to meet the requirements needed for the feasibility of AI solutions in a clinical setting to be examined for three reasons:

- Existing audio datasets containing cough sounds, such as the Google Audio Set (Gemmeke et al., 2017), do not specify the diagnoses or pathologies of the coughs; the lack of clinical metadata therefore making them inappropriate.
- For datasets that do, often they have a sample size that is too small, especially for data hungry ML architectures. For example, NoCoCoDa is a database of reflex cough sounds selected

from media interviews of COVID-19 patients, but only contains 10 unique subjects (Cohen-McFarlane et al., 2020).

- Most existing COVID-19 cough datasets contain noisy samples (i.e., numerous coughs in the same recording, ambient noises, etc.) making COVID-19 prediction challenging.

As a consequence, two and a half years in the pandemic, there still has yet been any cough-based COVID-19 detection tool available for the wider public, at the time of writing.

### 1.1. Paper's contributions

This paper makes the following contributions:

- We assess the current challenges of open cough COVID-19 datasets.
- We propose a novel cough audio segmentation framework that may be applied on top of existing cough datasets to increase the number of samples.
- We demonstrate the efficiency of our proposed framework.

The remainder of the paper is organised as follows. Section 2 discusses the inadequacy of existing cough sound datasets, the collection of crowdsourced cough sound datasets, and open sourcing some of these datasets following open data initiatives and standardisation to promote open

research, respectively. Sections 3 and 4 discuss our contribution of a proof-of-concept (POC) novel audio pre-processing and segmentation tool, and experimentation and evaluation of said tool on two open-source COVID-19 cough sound datasets. Lastly, Section 5 and 6 discuss related work in the field of cough segmentation, and outline conclusions and future works.

## 2. DATASETS OF COUGH SOUNDS

After 2 years since the pandemic, a number of crowdsourced audio datasets containing cough, breath, and speech sounds were collected by various research teams via smartphone and web applications (see Table 1).

University of Cambridge amassed 53,449 audio samples, of which 2,106 are COVID-19 positive, from 36,116 participants via their smartphone app<sup>1</sup> in their dataset named COVID-19 Sounds (Xia et al., 2021). This dataset is only available to academic institutions due to its sensitive nature.

Another curated dataset by the Wadhvani Institute of Artificial Intelligence, Mumbai, consists of 3,117 coughs from 1,039 individuals (Bagad et al., 2020). This dataset was intended to be publicly accessible, but unfortunately was not due to legal constraints.

The Government of Buenos Aires, Argentina, collected a clinically validated dataset called IATos v1 of 5,884 coughs from 2,821 individuals via WhatsApp used in a clinical setting (Pizzo et al., 2021). This dataset was one of the few that are balanced, with 1,409 individuals testing positive for COVID-19 and 1,412 testing negative. In 12/2021, they collected an additional 140,530 coughs in a second dataset called IATos v2, this time imbalanced with only 18,271 positive COVID-19 recordings.

The Indian Institute of Science (IISc), Bangalore, began project Coswara, curating a dataset of 2,030 audio samples of deep and shallow cough sounds, fast and slow breath sounds, and phonation of sustained vowels, of which 343 are COVID-19 positive, through their web application<sup>2</sup> (Sharma et al., 2020) (see Figure 1).

The Embedded Systems Laboratory (ESL) at École polytechnique fédérale de Lausanne (EPFL) collected 27,550 cough recordings, of which 1,156 are COVID-19 positive, through their web application<sup>3</sup>, codenaming their curated, partially clinically validated dataset COUGHVID (Orlandic et al., 2021). A cough detection algorithm was run on all audio files

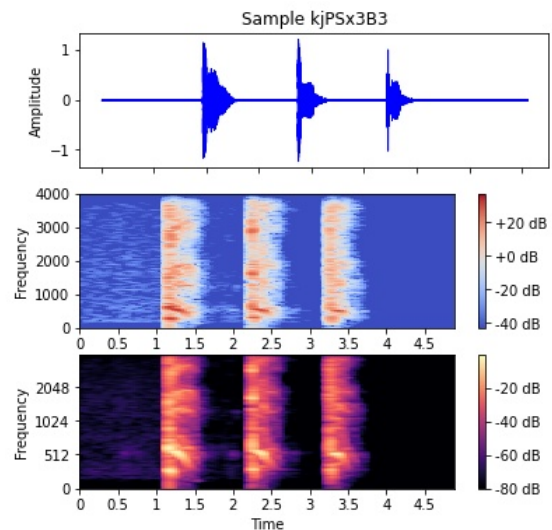
<sup>1</sup><https://covid-19-sounds.org/en/> - last accessed in 5/2022.

<sup>2</sup><https://record.coswara.iisc.ac.in> - last accessed in 5/2022.

<sup>3</sup><https://coughvid.epfl.ch> - last accessed in 5/2022.

in the dataset, producing a probability of likelihood that a cough is present (1.0 being likely, 0.0 being unlikely) (Orlandic et al., 2021). This can be used to filter data during pre-processing.

The Virufy AI Research Group are currently collecting cough sounds from multiple avenues, including via a smartphone app<sup>4</sup> and clinical collection in hospitals (Chaudhari et al., 2020; ?). This dataset was unreleased (save for a small sample of clinically validated coughs; 121 coughs from 16 participants, in which 48 coughs are COVID-19 positive (Chaudhari et al., 2020)), and the current number of samples kept private, though the team has stated that they intend to open source their datasets eventually. Additionally, the Virufy Common Data Format (CDF) was developed to standardise open-source datasets, i.e., to have the same column names (see Table 2) with an emphasis on clinical data, such as actual RT-PCR test results (or inferred where data is not/partially clinically validated) (Chaudhari et al., 2020).



**Figure 1:** An oscillogram, spectrogram, and Mel-spectrogram of an audio sample containing three 2-phase coughs from the Coswara dataset.

## 3. AUDIO SEGMENTATION

Suppose we have an algorithm that isolates each cough in an audio sample into its own audio file. An original sample with  $n=10$  coughs is split into 10 cough samples, increasing the sample count 10-fold. Necessary Digital Signal Processing (DSP) techniques, such as normalisation and silence removal, may be performed to remove uninformative data. This is the basis of our contribution. We believe that increasing the number of samples in a dataset may theoretically improve baseline

<sup>4</sup><https://virufy.org/study/welcome> - last accessed in 5/2022.

**Table 1:** Summary of some existing COVID-19 cough datasets.

Dataset Name	Institution	Samples	COVID-19	Healthy	Participants
COVID-19 Sounds	University of Cambridge, UK	53,449	2,106	51,343	36,116
Cough Against Covid	Wadhvani Institute of AI, Mumbai	11,724	3,867	7,857	Unknown
IATos v1	Government of Buenos Aires	16,626	8,268	8,358	2,771
IATos v2	Government of Buenos Aires	140,530	18,271	122,259	Unknown
Coswara	Indian Institute of Science (IISc)	2,030	343	1,687	941
COUGHVID	ESL, EPFL, Switzerland	27,550	1,156	26,394	Unknown
Virufy	Virufy AI Research Group	121	48	73	16

**Table 2:** The Virufy Common Data Format (CDF) defines a column structure for dataset metadata. Benefits of standardisation include faster and more efficient pre-processing of data, data filtering via `cough_detected`, and more emphasis on clinical data such as `pcr_test_result` for clinically validated and `pcr_test_result_inferred` for partially clinically validated data.

Column	Description
<code>row</code>	Row number of data.
<code>source</code>	Source of cough data.
<code>patient_id</code>	Unique identifier for patient.
<code>cough_detected</code>	Probability that the audio file contains an actual cough submission.
<code>audio_path</code>	File path to the audio file containing patient’s cough submission.
<code>audio_type</code>	Either cough or speech.
<code>age</code>	Age of patient.
<code>biological_sex</code>	Sex at birth of patient. The can be <code>male</code> , <code>female</code> , or <code>NaN</code> .
<code>reported_gender</code>	Reported gender of patient.
<code>submission_date</code>	Date cough was submitted by patient.
<code>pcr_test_date</code>	Date PCR test for presence of COVID-19 was taken.
<code>pcr_result_date</code>	Date test result from the PCR test for presence of COVID-19 was received.
<code>respiratory_condition</code>	Boolean indicator of whether patient suffers from a respiratory condition.
<code>fever_or_muscle_pain</code>	Boolean indicator of whether patient was suffering from a fever or muscle pain.
<code>pcr_test_result</code>	Result of patient’s PCR test for presence of COVID-19. This can be <code>positive</code> , <code>negative</code> , <code>recovered</code> , or <code>untested</code> .
<code>pcr_test_result_inferred</code>	Best guess of a patient’s COVID-19 diagnosis based on information specific to the dataset source. Can be <code>positive</code> , <code>negative</code> , <code>recovered</code> , or <code>untested</code> .
<code>covid_symptoms</code>	Boolean indicator of whether patient was experiencing COVID-19 symptoms.

accuracies (following the Law of Large Numbers) for data hungry DL models such as neural networks, but also general ML models if the data is of good quality.

Ensuring that data is clean and of a high quality is a ubiquitous ML problem. This is because it has been shown that noise can significantly skew prediction performance by obscuring class-differentiating information and therefore, reduce classification accuracy. We have chosen to demonstrate our POC cough segmentation tool on the open-source COVID-19 cough datasets Coswara (heavy recordings) and COUGHVID, standardised by Virufy’s CDF. The Coswara audio samples were encoded as WAV files at 48kHz/16 bit PCM (Pulse-Code Modulation) and the COUGHVID audio samples as WEBM files at 48kHz/16 bit PCM, though we converted these to WAV files for consistency, using the `ffmpeg` package. Both datasets have a large variation in sample duration (0.2-10 seconds), number of coughs per sample (1-15 individual coughs), and background noise levels. Thus, the proposed algorithm intends

to isolate individual coughs from each sample, regardless of the sample duration, in turn removing any background noise or silence present in the sample.

### 3.1. Cough event detection

To segment a time-series in a non-overlapping manner, the starting point of the event to isolate needs to be identified. This is known as onset detection, where extreme changes in an audio signal are located. The event here is a cough, the most common form physiologically characterised by three phases. During the first inspiratory phase, air is drawn into the lungs. This is followed by the compressive phase characterised by forced expiratory effort against the closed glottis. Finally, during the expiratory or expulsive phase, there is an opening of the glottis and rapid outflow of air. This sudden release of turbulent expiratory airflow is the signature sound that the human ear would recognise as a cough.

However, the difficulty in cough segmentation comes with the subtle differences between various cough types. A type of cough known as a peal cough, for example, is a cough that often occurs in epochs, or bouts, defined as a cluster of two or more cough sounds, separated from the next by an interval of no more than 2 seconds. Therefore, an epoch is where the initial inspiratory phase is followed by a series of further compressive phases associated with glottal closure, sometimes with additional inspirations. Additionally, another type of cough does not have a third phase, meaning that every energy peak is an individual cough (Cohen-McFarlane et al., 2020). Therefore, the difficulty in cough detection is developing an algorithm that can identify and accurately segment all of these cough types in their entirety, rather than the individual phases (Cohen-McFarlane et al., 2020). An analysis and summary of how our algorithm handles this variation are given in Sections 4 and 6 respectively.

### 3.2. Proposed segmentation framework

For cough onset detection, we select Root Mean Square Energy (RMSE) as the underlying feature as it may be interpreted as the signal’s amplitude or loudness (Meister et al., 2021). RMSE is computed from the Short-Term Energy (STE), which is the energy of a signal corresponding to the total magnitude of the signal. Loudness is an intuitive indicator for a cough event with relatively low background noise levels. The audio feature in Equation 1 describes a signal’s energy as a function of its amplitude  $x$  in frame  $n$ .

$$RMSE_n = \sqrt{x_n^2} \quad (1)$$

Given RMSE, the Python audio-processing package `librosa` identifies any extreme changes in the signal’s energy levels. We set a threshold whereby the peak amplitude of a cough event may be detected, but not the peak amplitudes of background noise events, such as doors closing, vehicular traffic, and conversation. However, one challenge with this approach is the possibility that the beginning of some cough events may be excluded. One attempt to mitigate this is to use backtracking to move the onset index back by a minuscule number of seconds i.e., 10 milliseconds. For a trade-off between code complexity and performance, we treat the second cough’s onset as the first cough’s offset. We also avoid detecting onsets in the last few frames of a sample. The intuition is that some audio files have a percussive peak at the end of the recording (i.e., button press of the recording device). The algorithm has a number of parameters that can be configured manually, divided between constants and variables. These parameters are described in Table 3.

The cough samples have differences in amplitude, thus, we normalise each sample between the range  $(-1, 1)$  to compensate. A cough has an average duration of 350ms with an initial peak frequency of 400Hz, a secondary peak of highest continuous frequency 4kHz, and frequency components of sound spread up to 20kHz (Murata et al., 1998). To properly capture characteristic information of the coughs, many studies include re-sampling the audio samples to 8kHz, as it has been observed that most of the frequency information of a cough sound is between the range of 350Hz and 4kHz (Shin et al., 2008), but this must be doubled according to the Nyquist sampling theorem to avoid anti-aliasing. However, whilst 8kHz may be adequate for feature extraction and the generation of spectrograms for classification, we wish to create a new dataset of audible isolated coughs. Therefore, we select a higher sampling rate of 22.05kHz.

## 4. EXPERIMENTAL RESULTS

This section demonstrates the performance of our proposed cough segmentation framework on two open cough COVID-19 datasets.

### 4.1. Evaluation metrics

According to Murata et al., (1998), cough events are between 0.2 and 1 seconds. Therefore, we will analyse how many isolated coughs fall within this range; for those outside the range, we assume that these have either not been correctly segmented, or that background noise or silence was not adequately removed. Additionally, we will use the Signal-to-Noise Ratio (SNR) to assess the reduction of background noise which we consider uninformative data. SNR in Equation 2 measures the average power of the background noise and the average power of the foreground sound, in this case a cough event. It is measured on a logarithmic decibel scale, where the higher the SNR score, the higher the quality of the sample.

$$SNR_{dB} = 20 \cdot \log_{10} \left( \frac{\text{amplitude}_{signal}}{\text{amplitude}_{noise}} \right)^2 \quad (2)$$

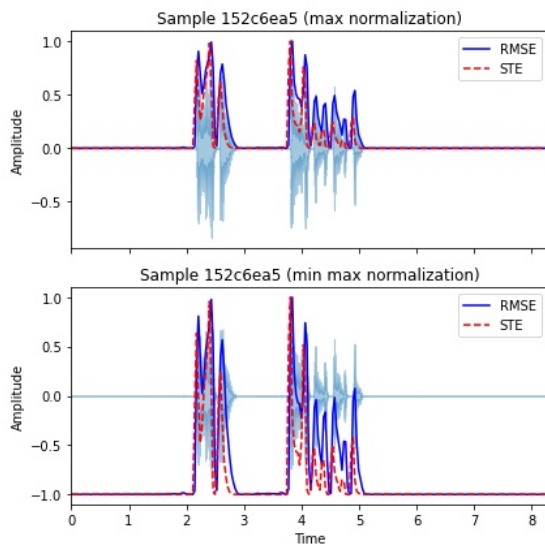
### 4.2. Experiments using COVID-19 open data

Prior to running the segmentation procedure, we perform some diagnostic procedures to analyse the performance of the onset detection algorithm. Firstly, we demonstrate the difference between using Max Normalisation, where the sample is normalised between the range  $(0, 1)$ , and Min Max Normalisation, where the sample is normalised between the range  $(-1, 1)$ , with the Root Mean

**Table 3:** A list of configurable constants and parameters used by the cough segmentation algorithm. Note that the constants can only be changed in the source code, whereas the parameters can be set at runtime with each instance of the cough segmentation tool created by the user.

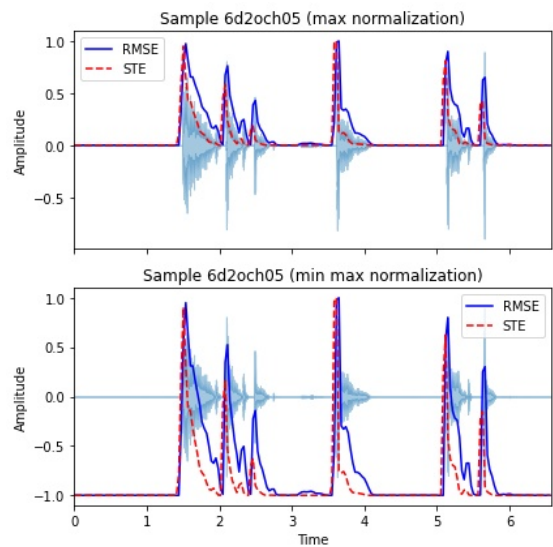
Parameter	Datatype	Intuition
Sample Rate	int	The number of samples per second in kHz.
Frame Length	int	The number of samples in a frame.
Hop Length	int	The number of samples between frames.
Silence Threshold	int	The threshold in dB below reference to consider as silence.
RMSE Threshold	int	The threshold of Root Mean Square Energy from which onsets are detected.
Minimum Distance	int	The minimum distance between detected high RMSE frames.
Backtrack	float	The number of seconds to subtract from onset times.
Trim Frames	bool	A toggle to turn on functionality for trimming the last onset frames.
End Frames	int	The number of onset frames to trim from the end of the audio file.

Square Energy (RMSE) and Short-Term Energy (STE) of the signal graphed over the oscillogram of the signal. Max Normalisation is a common choice in the literature, but we proposed using Min Max Normalisation as it performs better when the sample contains any of the three cough types; 2-phase, 3-phase, or peal. It worked particularly well with peal coughs as demonstrated in Figure 2 with coughs from the COUGHVID dataset, and Figure 3 with coughs from the Coswara (heavy recordings) dataset, respectively. When Max Normalisation was used, the third phase of a 2-phase cough, or the epochs of a peal cough, was incorrectly identified as individual coughs. This does not occur when Min Max Normalisation was used.



**Figure 2:** RMSE and STE graphed over an oscillogram of a 3-phase and peal cough from the COUGHVID dataset.

Next, we set the constants and tune the parameters of the onset detection algorithm (see Table 3. Table 4 outlines the constants used for all datasets. For the COUGHVID and Coswara datasets, Table 5 lists the optimal parameters for onset detection.



**Figure 3:** RMSE and STE graphed over an oscillogram of 2-phase, 3-phase, and peal coughs from the Coswara (heavy recordings) dataset.

**Table 4:** The selected values for the constants of the onset detection algorithm.

Constant	Selected Value
Sample Rate	22,050
Frame Length	512
Hop Length	256
Silence Threshold	10

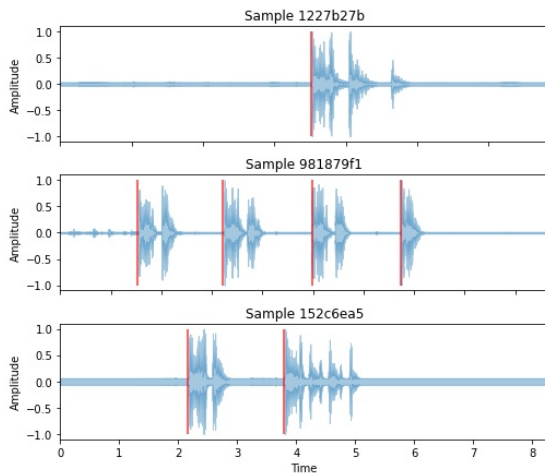
**Table 5:** The optimal parameters for the onset detection algorithm with the 2 datasets.

Parameter	COUGHVID	Coswara
RMSE Threshold	0.6	0.5
Minimum Distance	50	60
Backtrack	0.01	0.01
Trim Frames	True	True
End Frames	1	1



With parameter tuning done the onset detection algorithm is run on 100 samples of the COUGHVID and Coswara (heavy recordings) datasets and the resulting oscillograms of 10 coughs from COUGHVID and 10 from Coswara are shown in Figures 4 and 5. Detected onsets are displayed on the oscillograms as red vertical lines. We observed that the algorithm worked well with both 2-phase and 3-phase coughs from both datasets, but could be hypersensitive<sup>5</sup> to peal coughs, though this occurs more in Coswara than in COUGHVID. Whilst this is not a huge issue as the cough epoch still contains informative data, some data is lost from not having the full peal cough isolated in its entirety.

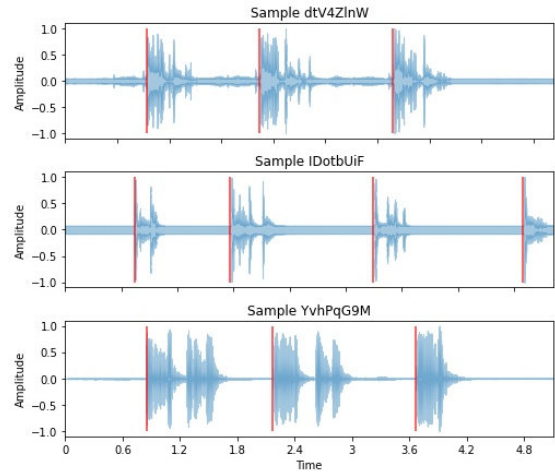
We found that our algorithm handles moderate noisy conditions well. Most background noise events were not classed as a cough, though with some exceptionally loud noises, mis-classification did occur. We go into more detail about how the effects of this can be mitigated in Section 6.



**Figure 4:** Detected cough onsets displayed on oscillograms of 2-phase, 3-phase, and peal coughs from the COUGHVID dataset.

The next step is to test the full segmentation pipeline comprising of onset detection, segmentation, pre-processing, and exportation procedures on the COUGHVID and Coswara datasets. We used a subset of 100 samples from each dataset in our segmentation experiment. The isolated audio samples were exported to WAV format with a sampling rate of 22.05kHz with a bit-depth of 16 bits. Our segmentation algorithm performed well, with almost 90% of isolated samples being of a high-quality, and under 25% of isolated samples being discarded, for the COUGHVID dataset; and with over 80% of isolated samples being of a high-quality, and under 30% of isolated samples being discarded, for the Coswara dataset.

<sup>5</sup>By hypersensitive, we mean to say that the epochs of a peal cough can be occasionally misidentified as a standalone cough.



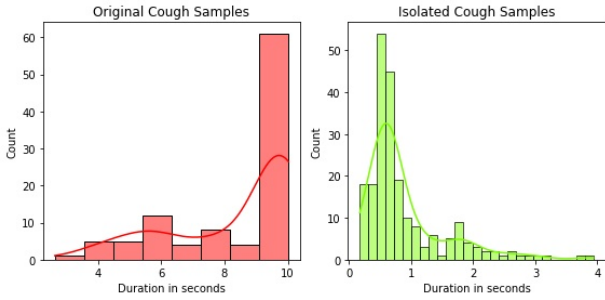
**Figure 5:** Detected cough onsets displayed on oscillograms of 2-phase, 3-phase, and peal coughs from the Coswara (heavy recordings) dataset.

### 4.3. Summary of results

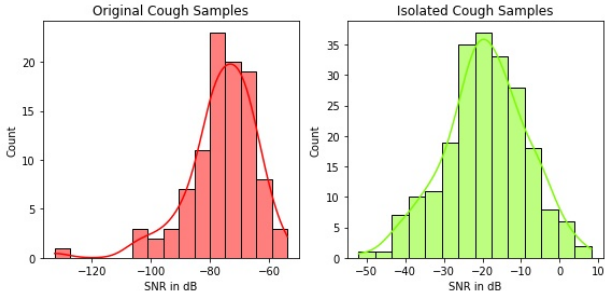
To summarise, we evaluate the datasets produced from COUGHVID and Coswara using the metrics outlined above graphed as a comparison between the original and isolated datasets.

Firstly, looking at the COUGHVID dataset, from our subset of 100 samples, the mean average duration is 8.4 seconds, decreasing to 0.9 seconds for the isolated cough dataset. For Coswara, from our subset of 100 samples, the mean average duration is 6 seconds, decreasing to 0.8 seconds for the isolated cough dataset. Therefore, we hypothesise that the majority of isolated cough samples were correctly segmented with both datasets, as the average cough duration's are both within the range of 0.2 to 1 seconds assumption. This is further supported by Figures 6 and 8. As for the cough SNRs, the mean average SNR from our subset of 100 COUGHVID samples is -76dB, compared with -19dB from the isolated cough dataset. The mean average SNR from our subset of 100 Coswara samples is -74dB, compared with -19dB from the isolated cough dataset. For both datasets the latter is significantly higher, with Figure 7 and Figure 9 showing some isolated cough samples even reaching an approximate SNR of 10dB (occured more-so in Coswara). Therefore, we hypothesise that the signal quality of the isolated cough samples is better than that of the original samples which include uninformative data i.e., background noise.

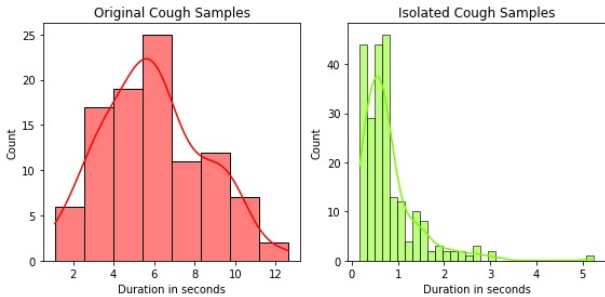
Lastly, we analyse the the class distributions of both the COUGHVID and Coswara subsets of 100 samples, as well as the isolated cough datasets produced from them. Here, we may analyse the class distribution to see if the segmentation procedure



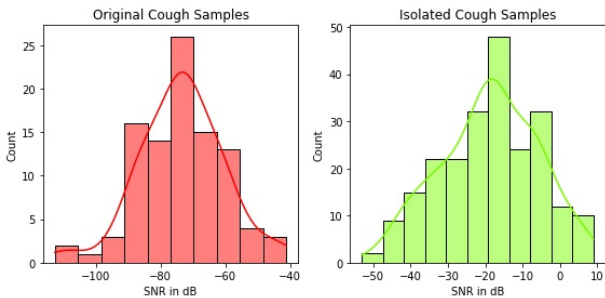
**Figure 6:** Comparison of cough duration's from COUGHVID and the isolated cough dataset.



**Figure 7:** Comparison of cough SNR's from COUGHVID and the isolated cough dataset.



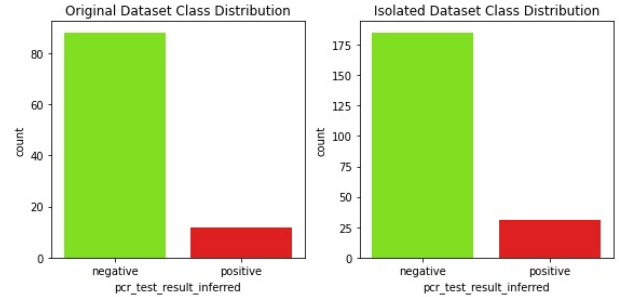
**Figure 8:** Comparison of cough duration's from Coswara heavy and the isolated cough dataset.



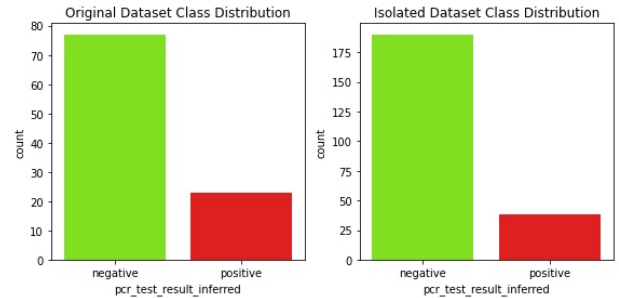
**Figure 9:** Comparison of cough SNR's from Coswara heavy and the isolated cough dataset.

has had an effect on the class imbalance (as both datasets are imbalanced). From Figures 10 and 11, we see that, the class imbalance has improved by 2.35% and 6.33% respectively. Our understanding is that the segmentation procedure may positively

impact the class imbalance (by increasing the frequency of samples in the positive class) by segmenting an original positive sample into multiple positive samples, and decreasing the proportion of the negative class via quality filtering (negative class samples are predominant and more likely to be discarded for low-quality).



**Figure 10:** Class distributions of COUGHVID and isolated cough datasets.



**Figure 11:** Class distributions of Coswara (heavy recordings) and isolated cough datasets.

## 5. RELATED WORK

Cough segmentation to support COVID-19 detection is still in its infancy. Nevertheless, there are relevant work in other respiratory research, as follows.

Amrulloh et al., (2015) propose a complex automated cough extraction methodology. The process begins with noise reduction using a High Pass Filter (HPF) and Power Subtraction Filter (PSF). Next, the signal is processed into  $n$  sub-blocks and extract the features Mel-Frequency Cepstral Coefficients (MFCCs), Formant frequency, Zero Crossing Rate (ZCR), Non-Gaussianity Score (NGS), and Shannon entropy. The feature matrix is then input into a trained Artificial Neural Network classifier to differentiate sub-blocks into cough and non-cough classes. Smoothing using a moving average filter is followed by thresholding and cough event identification. Infante et al., (2017) propose a method where the data is first smoothed by applying local regression using a  $2^{nd}$  degree polynomial model. A peak detection algorithm is applied to the smoothed signal, and then each peak is analysed

individually. The zero-crossing of the first derivative is used to determine the cough onset, and the slope of the trailing edge of the cough for the offset.

## 6. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a novel audio segmentation framework for cough-based COVID-19 detection. In optimal conditions, where cough events are evenly spaced with low background noise, our framework worked well on both 3-phase coughs and 2-phase coughs, correctly isolating for example a sample containing  $n=3$  coughs into three distinct audio files containing a single cough, with normalised amplitudes and no uninformative data.

However, our algorithm still has limitations in this POC stage. It is hypersensitive to some peal coughs characterised as epochs, or clusters of two or more cough sounds, identifying each as an individual cough between 0.1 and 0.5 seconds. This was because some frames within the epoch have a high enough RMSE value to exceed the threshold, whereby being classified as a standalone cough. Increasing the threshold addressed this problem, at the cost of some incorrectly identified onsets. Adapting the algorithm to identify cough clusters as opposed to individual cough sounds could mitigate the limitation of hypersensitive onset detection, but is outside the scope of this paper, and is considered as a potential future work.

## REFERENCES

- Han, J., Brown, C., Chauhan, J., Grammenos, A., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P. and Mascolo, C., 2021. Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data. In *ICASSP 2021 International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8328-8332). IEEE.
- Pahar, M., Klopper, M., Warren, R. and Niesler, T., 2021. COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*, 135, p.104572.
- Gemmeke, J.F., Ellis, D.P., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M. and Ritter, M., 2017. Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 776-780). IEEE.
- Cohen-McFarlane, M., Goubran, R. and Knoefel, F., 2020. Novel coronavirus cough database: Nococoda. *Ieee Access*, 8, pp.154087-154094.
- Xia, T., Spathis, D., Ch, J., Grammenos, A., Han, J., Hasthanasombat, A., Bondareva, E., Dang, T., Floto, A., Cicuta, P. and Mascolo, C., 2021, August. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.
- Bagad, P., Dalmia, A., Doshi, J., Nagrani, A., Bhamare, P., Mahale, A., Rane, S., Agarwal, N. and Panicker, R., 2020. Cough against covid: Evidence of covid-19 signature in cough sounds. *arXiv preprint arXiv:2009.08790*.
- Pizzo, D.T., Esteban, S. and Scetta, M., 2021. IATos: AI-powered pre-screening tool for COVID-19 from cough audio samples. *arXiv preprint arXiv:2104.13247*.
- Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S.R., Ghosh, P.K. and Ganapathy, S., 2020. Coswara—a database of breathing, cough, and voice sounds for COVID-19 diagnosis. *arXiv preprint arXiv:2005.10548*.
- Orlandic, L., Teijeiro, T. and Atienza, D., 2021. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific Data*, 8(1), pp.1-10.
- Chaudhari, G., Jiang, X., Fakhry, A., Han, A., Xiao, J., Shen, S. and Khanzada, A., 2020. Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. *arXiv preprint arXiv:2011.13320*.
- Meister, J.A., Nguyen, K.A. and Luo, Z., 2021. Audio feature ranking for sound-based COVID-19 patient detection. *arXiv preprint arXiv:2104.07128*.
- Murata, A., Taniguchi, Y., Hashimoto, Y., Kaneko, Y., Takasaki, Y. and Kudoh, S., 1998. Discrimination of productive and non-productive cough by sound analysis. *Internal Medicine*, 37(9), pp.732-735.
- Shin, S.H., Hashimoto, T. and Hatano, S., 2008. Automatic detection system for cough sounds as a symptom of abnormal health condition. *IEEE Transactions on Information Technology in Biomedicine*, 13(4), pp.486-493.
- Amrulloh, Y.A., Abeyratne, U.R., Swarnkar, V., Triasih, R. and Setyati, A., 2015. Automatic cough segmentation from non-contact sound recordings in pediatric wards. *Biomedical Signal Processing and Control*, 21, pp.126-136.
- Infante, C., Chamberlain, D., Fletcher, R., Thorat, Y. and Kodgule, R., 2017, October. Use of cough sounds for diagnosis and screening of pulmonary disease. In *2017 IEEE Global Humanitarian Technology Conference (GHTC)* (pp. 1-10). IEEE.