DiCOVA Challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics

Ananya Muguli[†], Lancelot Pinto[‡], Nirmala R.[†], Neeraj Sharma[†], Prashant Krishnan[†],
Prasanta Kumar Ghosh[†], Rohit Kumar[†], Shrirama Bhat[∓], Srikanth Raj Chetupalli[†],
Sriram Ganapathy[†], Shreyas Ramoji[†], Viral Nanda[‡]

[†]Indian Institute of Science, Bangalore, India
[‡]P. D. Hinduja Hospital, Mumbai, India
[∓]KMC Hospital, Mangalore, India

sriramg@iisc.ac.in

Abstract

The DiCOVA challenge aims at accelerating research in diagnosing COVID-19 using acoustics (DiCOVA), a topic at the intersection of speech and audio processing, respiratory health diagnosis, and machine learning. This challenge is an open call for researchers to analyze a dataset of sound recordings, collected from COVID-19 infected and non-COVID-19 individuals, for a two-class classification. These recordings were collected via crowdsourcing from multiple countries, through a website application. The challenge features two tracks, one focusing on cough sounds, and the other on using a collection of breath, sustained vowel phonation, and number counting speech recordings. In this paper, we introduce the challenge and provide a detailed description of the task, and present a baseline system for the task.

Index Terms: COVID-19, acoustics, machine learning, respiratory diagnosis, healthcare

1. Introduction

The COVID-19 pandemic has emerged as a significant health crisis. At the time of writing ($15-$June-2021), more than 175 million cases and more than 3.8 million casualties have been reported by the World Health Organization (WHO) from about 200 countries across the world [1]. Physical distancing and implementation of wide-scale population testing have served as key measures to contain the pandemic. The testing methods in use can be broadly divided into molecular and antibody testing. In molecular testing, chemical reagents are used to detect the constituents, like nucleic acids and proteins, of the SARS-CoV-2 virus in an individuals' throat or nasal swab sample. The reverse transcription polymerase chain reaction (RT-PCR) is one such testing method, and currently serves as a gold standard for COVID-19 testing. However, cost of machinery, time, and expertise have limited the scalability of this method. The rapid antigen test (RAT) is another molecular testing method which alleviates the time limitation of RT-PCR but has high false negatives (low specificity). The swab based tests and molecular tests also violate physical distancing between participant and the health worker, posing a serious practical challenge. In summary, there is a need to discover alternative methodologies to diagnose COVID-19 infection that are efficient in terms of time, cost, and ease, allowing scalability.

The WHO [1] has maintained dry cough, breathing difficulty, chest pain, and fatigue as symptoms of the infection, manifested between $2-14$ days after exposure to the virus. This was also validated by a modeling study that analyzed data
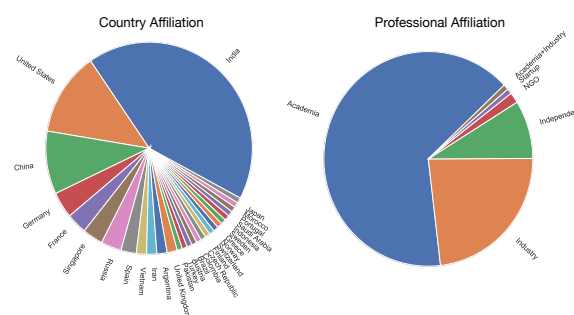

Figure 1: *Illustration of the distribution* 80 *plus challenge registrants (or teams).*

pertaining to the symptoms reported by 7178 COVID-19 positive individuals [2]. The chest X-ray (and CT) scans of many COVID-19 infected individuals have revealed infection in the lungs [3], and effort is being directed to evaluate the feasibility of early diagnosis using imaging techniques. Interestingly, respiratory medical literature suggests that sounds emanating through coordinated release of air pressure through the lungs, such as breathing, cough, and speech, are intricately tied to changes in the anatomy and the physiology of the respiratory system [4]. A lung infection can affect the inspiratory and expiratory capacity. This, in addition to the presence of cough, can result in difficulty in vocalizing sustained phonation and/or continuous speech [5, 6]. This has been the scientific principle based on which studies analyzing vocal sounds have shown some success in detecting respiratory ailments, such as pertussis [7], chronic obstructive pulmonary disease (COPD) [8], and tuberculosis [9].

Based on such biological plausibility, we hypothesize that the evaluation of the accuracy of detecting COVID-19 using the acoustics of respiratory sounds merits research. A success can provide an excellent point-of-care, quick, easy to use, and cost-effective tool to diagnose COVID-19 infection, and consequently contain COVID-19 spread. Altogether, it can supplement the molecular testing methods for COVID-19 detection or screening. The DiCOVA Challenge[1] is designed to accelerate research efforts along this direction by creation and release of an acoustic signal dataset, and inviting researchers to build detection models and report performance on a blind test set. Since its release on $04-$Feb-2021, the DiCOVA Challenge has cre-
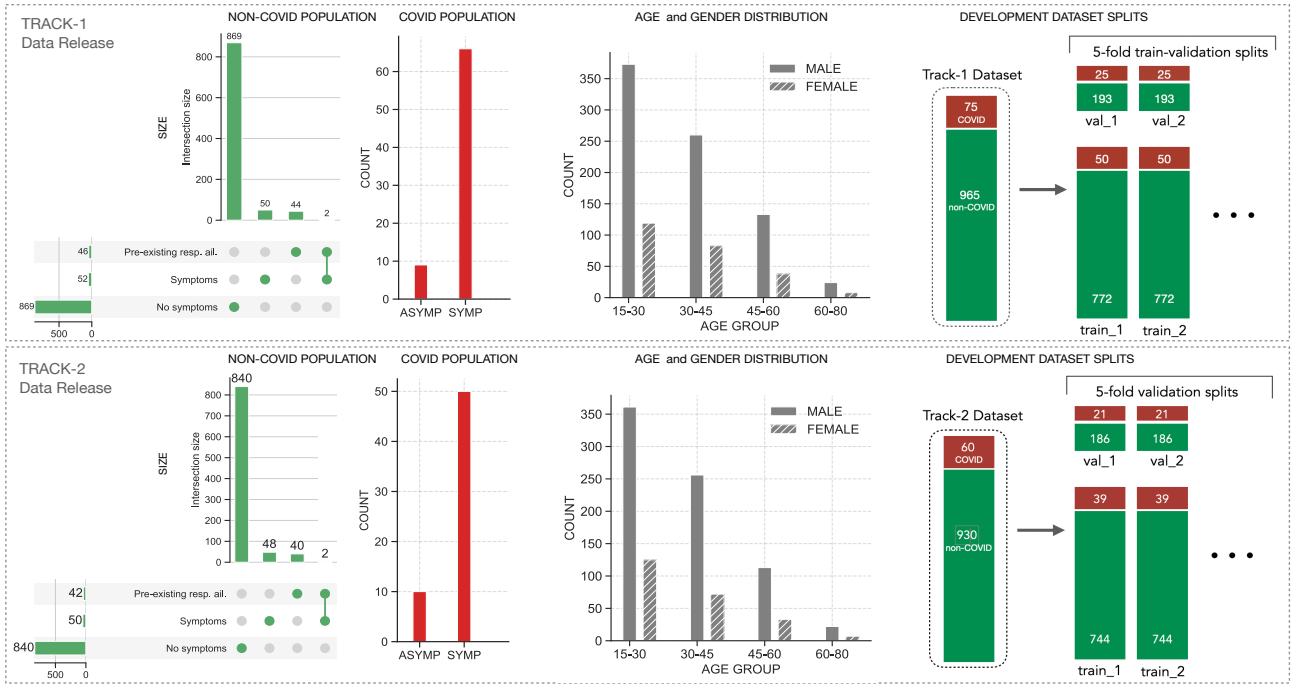
---

[1]http://dicova2021.github.io/

Figure 2: *In each track, the dataset is grouped non-COVID and COVID subjects. The non-COVID subjects are either healthy, have symptoms (cough/cold), or have pre-existing respiratory ailments (chronic lung disease, asthma, or pneumonia). The COVID subjects are either symptomatic or asymptomatic COVID positive. The distribution of age, gender, and the splits of the development dataset is also shown.*

ated a widespread interest amongst researchers. We have received registration from more than 80 teams. These come from various countries and professional affiliation (see Figure 1). In this paper, we present an overview of the topic, tasks in the challenge, and the baseline system.

2. Literature Review

Since the onset of the COVID-19 pandemic, several attempts are being made to evaluate the potential of sound based screening (and diagnosis). These attempts [10, 11, 12, 13, 14, 15, 16, 17] have primarily focused on cough sounds, and are work in progress. Brown et.al. [16] use cough and breathing sounds from 141 COVID-19 patients, extract a collection of short-time frame-level acoustic features and embeddings from a VGGish network, and pass these through a logistic regression classifier. An area-under-the-curve (AUC) 80% is reported. The study by Imran et al. [15] uses sound samples from 48 COVID-19 patients, and reports a sensitivity of 94% (and 91% specificity) using a convolutional neural network (CNN) architecture, fed with mel-spectrogram features as the input. The study by Bagad et.al. [17] uses cough samples from 376 COVID-19 patients, and a CNN architecture based on ResNet18 with short-time magnitude spectrogram as input, and reports an AUC of 72%. Altogether, these studies are encouraging. The limitations include: (i) a different COVID-19 patient population used in each study, (iii) varied evaluation methodology, (iii) small population size, and (iii) lack of insight on acoustic feature differences between healthy and COVID-19 individuals. The DiCOVA Challenge is aimed to encourage multiple research groups to analyze the same dataset, evaluate the system performance using fixed metrics, and facilitate obtaining benchmarks

for future system development.

3. Dataset

The DiCOVA Challenge dataset is derived from the Coswara dataset [18], a crowd-sourced dataset of sound recordings from COVID-19 positive and non-COVID-19 individuals. The Coswara data is collected using a web-application[2], launched in April-2020, accessible through the internet by anyone around the globe. The volunteering subjects are advised to record their respiratory sounds in a quiet environment. Each subject provides 9 audio recordings, namely, (a) shallow and deep breathing (2 nos.), (b) shallow and heavy cough (2 nos.), (c) sustained phonation of vowels [æ] (as in bat), [i] (as in beet), and [u] (as in boot) (3 nos.), and (d) fast and normal pace 1 to 20 number counting (2 nos.). The subjects also provided metadata corresponding to their current health status (includes COVID-19 status, any other respiratory ailments, and symptoms), demographic information like age and gender. From this Coswara dataset, we have created two datasets: (a) Track-1 dataset: composed of cough sound recordings, and (b) Track-2 dataset: composed of deep breathing, vowel [i], and number counting (normal pace) speech recordings.

3.1. Metadata

For the challenge, the subjects have been divided into two groups, namely,

- non-COVID: Subjects who are either healthy or have symptoms such as cold or cough, or have pre-existing respiratory ailments (asthma, pneumonia, chronic lung

---

disease), and confirm that they are not COVID-19 positive.

- COVID: Subjects who confirm as COVID-19 positive, either symptomatic and asymptomatic.

The Track-1 and Track-2 development datasets are composed of 1040 (965 non-COVID subjects) and 990 (930 non-COVID subjects), respectively. A breakdown of the subject population with respect to symptoms, age group, and gender is shown in Figure 2.

3.2. Audio

The Coswara data collection is via crowd-sourcing, which means the quality of the audio files has high variability and serves as a good representation of audio data collected in the wild. A majority of the audio files are clean as confirmed via informal listening. More than 90% of the collected files have a sampling rate of 48 kHz and stored as WAV files. For the challenge datasets, all audio recordings have been re-sampled to 44.1 kHz and compressed as FLAC format files. The Track-1 audio files correspond to cough sound signals. Each audio file is derived from one unique subject, and has one or more cough bouts. In total, there are 1040 recordings. The average duration of recordings across subjects is 4.72(standard error ± 0.07) sec. The Track-2 audio files correspond to one of the three different sound categories, namely, breathing, vowel [i], and 1 to 20 number counting. In total, there are 3(categories)$\times$990 (subjects) sound recordings in Track-2. The average duration of recordings across subjects is: breath 17.72(± 0.68) sec, vowel [i] 12.40(± 0.17) sec, and number counting speech 14.71(± 0.11) sec.

4. Challenge Tasks

The DiCOVA challenge features two tracks. Below we present the task and the instructions associated with each track. A participant can choose to participate in one or both the tracks.

4.1. Track-1

The goal is to use cough sound recordings from COVID-19 and non-COVID-19 individuals for the task of COVID-19 detection.

- The Track-1 development dataset is composed of cough audio data from 1040 subjects. The dataset also contains lists corresponding to a $5-$fold cross validation split. The distribution of COVID and non-COVID in these splits is shown in Figure 2(a). All participants are required to adhere to these lists and report the average performance over the 5 validation sets.

- A separate blind evaluation dataset is provided to all participants. The participants are required to report their COVID-19 detection scores as probabilities.

- This is the primary track for the challenge. A baseline system is provided, and an online leaderboard [3] is set up for all participants to report and compare their performance.

4.2. Track-2

The goal is to use breathing, sustained phonation, and speech sound recordings from COVID-19 and non-COVID-19 individuals for any kind of detailed analysis which can contribute towards COVID-19 detection.

- The Track-2 development dataset is composed of three sets of sound recordings, namely, breathing, vowel [i], and number counting, from 990 subjects.

- The dataset also contains 5 train-validation splits. The distribution of COVID and non-COVID in these splits is shown in Figure 2(b).

- The participants are encouraged to design COVID-19 detection systems using above splits.

- This track has no baseline system and leaderboard. A non-blind test set is provided to all participants.

Participants are free to use any other data except the publicly available Project Coswara dataset [4] for data augmentation, transfer learning, etc.

4.3. Performance Evaluation

Both Track-1 and Track-2 are binary classification tasks. With a focus on COVID detection, the performance is evaluated using the traditional detection metrics, namely, true positive (TP) and false positive (FP) rates, over a range of decision thresholds between $0 - 1$ with a step-size of 0.0001. For track-1, the participant is required to submit a COVID probability score for every audio file (corresponding to a subject) in the blind test set. In the evaluation, we use the probability scores to compute the receiver operating characteristic (ROC) curve, and use the area under the curve (AUC) to quantify the model performance. An AUC $> 50\%$ indicates a better than chance performance, and an AUC closer to 100% indicates the ideal model performance. We also compute the model specificity at 80% sensitivity.

5. Baseline System

5.1. Data preparation

The audio data is pre-processed by normalizing the amplitude range to ± 1. Subsequently, a simple sample level sound activity detection (SAD) is applied. This keeps any audio sample with absolute value greater than 0.01 (and a margin of ± 50 msec around it) and discards the rest of the audio samples. Further, the initial and the final 20 msec audio samples are also discarded to remove abrupt start and end burst due to device noise.

5.2. Feature Extraction

Here, 39 dimensional mel-frequency cepstral coefficients (MFCC) [19] and the delta and delta-delta coefficients are extracted with a window of size 1024 samples and a hop of size 441 samples. The `librosa` python library [20] is used for the computation.

5.3. Model Training

Three different classifier models are trained for the two class classification tasks of COVID versus non-COVID detection. The models are trained using the extracted features and a (class) balanced loss function, separately, for each of the five training data splits. The implementation uses the `scikit-learn` python library [21]. The classifier models include the following.

- Logistic regression (LR): A logistic regression classifier trained with an added ℓ_2 penalty, regularization strength of 0.01 and `liblinear` optimizer is used. The maximum number of iterations is chosen as 25.
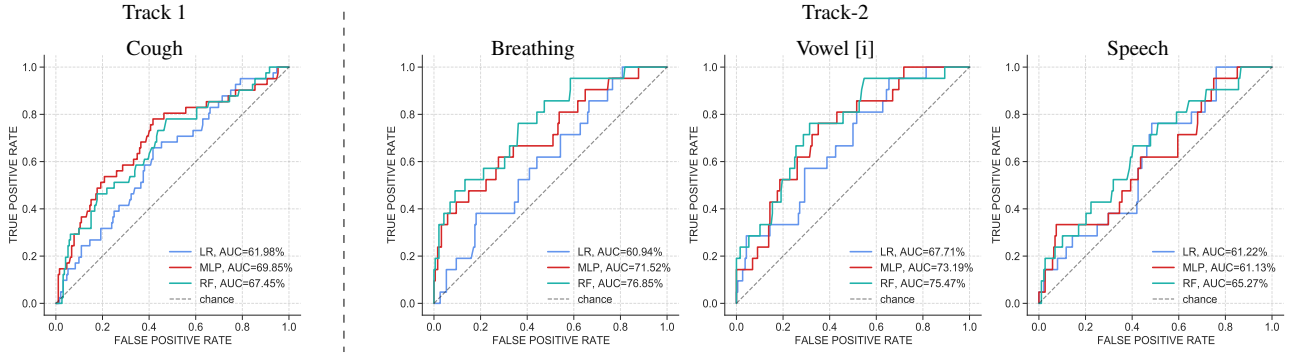

Figure 3: *Illustration of baseline systems ROCs obtained on Track-1 and Track-2 test datasets.*

- **Multi-layer perceptron (MLP):** A single layer perceptron with 25 hidden units, $\tanh()$ activation, and ℓ_2 regularization penalty with a weight of 0.001 is used. The loss function is optimised using the Adam optimizer with an initial learning rate of 0.001. To implement balanced loss for MLP, the COVID class samples are randomly oversampled in appropriate proportion to match the count of non-COVID class samples.

- **Random Forest (RF):** The random forest classifier is trained with 50 trees in the forest and *gini* impurity criterion to measure the split quality.

5.4. Model Inference and Decisions

To obtain a classification score for an audio file: (i) a pre-processing with amplitude normalization and SAD is done, (ii) frame-level MFCC features are extracted, (iii) frame-level probability scores are computed using the trained model, and (iv) all the frame scores are averaged to obtain a single COVID probability score for the audio file.

5.5. Results

Table 1 depicts the AUCs obtained on the validation folds. For each fold (shown in Fig. 2), the classifier is trained using the training data and evaluated on the validation data. The average validation AUC denotes the average over the AUCs for the five folds. For Track-1, RF gave the best average AUC, equating to 70.69%, and this was followed by MLP (at 68.80%) and LR (at 66.95%). For Track-2, RF gave superior performance on breathing sound (75.17% AUC). All models performed similar for vowel sound with an AUC close to 70%. The MLP gave a superior performance for speech sound, with 73.57% AUC.

For evaluation on the test dataset, the COVID probability score for each file was computed by taking the average over the score outputs from the five validation fold models. The Track-1 blind test dataset release contains 233 (41 COVID) cough audio files for classification into COVID/non-COVID. For Track-1, the LR, MLP, and RF gave 61.98%, 69.85%, and 67.45% AUCs, respectively. The corresponding ROCs are shown in Fig. 3.

The Track-2 test dataset release contains 209 (21 COVID) audio files for each of the three sound categories. Here, the RF model gave a better performance than other models in all the three sound categories. Its performance was best for breathing (76.85% AUC) and worst for speech (65.27% AUC).

Table 1: *The baseline system performance on the validation folds.*

Track	Sound	Model	Avg.Val (Std. Err.)
1	Cough	LR	66.95 (±1.74)
		MLP	68.54 (±1.65)
		RF	70.69 (±1.39)
2	Breathing	LR	60.95 (±2.17)
		MLP	72.47 (±1.96)
		RF	75.17 (±1.23)
	Vowel [i]	LR	71.48 (±0.55)
		MLP	70.39 (±1.84)
		RF	69.73 (±1.93)
	Speech	LR	68.93 (±1.09)
		MLP	73.57 (±0.71)
		RF	69.61 (±1.56)

6. Conclusion

The uniqueness of the dataset makes the DiCOVA challenge a first-of-its kind in the INTERSPEECH conference. The practical and timely relevance of the task encourages a focused effort from researchers across the globe, and from diverse fields such as respiratory sciences, speech and audio processing, and machine learning. Along with the dataset, we also provide the baseline system software to all the participants. We expect this will serve as an example data processing pipeline for the participants. Further, participants are encouraged to explore different kinds of features and models of their own choice to obtain significantly better performance compared to the baseline system.

7. Acknowledgement

8. References

[1] "WHO Coronavirus Disease (COVID-19) Dashboard," https://covid19.who.int/, 2020, [Online; accessed 10-Feb-2021].

[2] C. Menni, A. M. Valdes, M. B. Freidin, C. H. Sudre, L. H. Nguyen, D. A. Drew, S. Ganesh, T. Varsavsky, M. J. Cardoso,

J. S. El-Sayed Moustafa, A. Visconti, P. Hysi, R. C. E. Bowyer, M. Mangino, M. Falchi, J. Wolf, S. Ourselin, A. T. Chan, C. J. Steves, and T. D. Spector, "Real-time tracking of self-reported symptoms to predict potential COVID-19," *Nature Medicine*, 2020. [Online]. Available: https://doi.org/10.1038/s41591-020-0916-2

[3] N. Islam, S. Ebrahimzadeh, J.-P. Salameh, S. Kazi, N. Fabiano, L. Treanor, M. Absi, Z. Hallgrimson, M. Leeflang, L. Hooft, C. Pol, R. Prager, S. Hare, C. Dennie, R. Spijker, J. Deeks, J. Dinnes, K. Jenniskens, D. Korevaar, J. Cohen, J. Van den Bruel, Y. Takwoingi, J. de Wijgert, J. Damen, J. Wang, and M. McInnes, "Thoracic imaging tests for the diagnosis of COVID-19," *Cochrane Database of Systematic Reviews*, no. 3, 2021. [Online]. Available: https://doi.org/10.1002/14651858. CD013639.pub4

[4] J. E. Huber and E. T. Stathopoulos, *Speech Breathing Across the Life Span and in Disease*. John Wiley & Sons, Ltd, 2015, ch. 2, pp. 11–33.

[5] L. Lee, R. G. Loudon, B. H. Jacobson, and R. Stuebing, "Speech breathing in patients with lung disease," *American Review of Respiratory Disease*, vol. 147, pp. 1199–1199, 1993.

[6] A. Chang and M. P. Karnell, "Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue," *Journal of Voice*, vol. 18, no. 4, pp. 454–466, 2004.

[7] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "A cough-based algorithm for automatic diagnosis of pertussis," *PloS one*, vol. 11, no. 9, 2016.

[8] A. Windmon, M. Minakshi, P. Bharti, S. Chellappan, M. Johansson, B. A. Jenkins, and P. R. Athilingam, "Tussiswatch: A smartphone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and congestive heart failure," *IEEE J. Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1566–1573, 2018.

[9] G. Botha, G. Theron, R. Warren, M. Klopper, K. Dheda, P. Van Helden, and T. Niesler, "Detection of tuberculosis by automatic cough sound analysis," *Physiological Measurement*, vol. 39, no. 4, p. 045005, 2018.

[10] "Cambridge University, UK - COVID-19 Sounds App," https://covid-19-sounds.org/en/, 2020, [Online; accessed 07-Aug-2020].

[11] "Cough Against COVID - Wadhwani AI Institute," https://coughagainstcovid.org/, 2020, [Online; accessed 07-Aug-2020].

[12] "NYU Breathing Sounds for COVID-19," https://breatheforscience.com/, 2020, [Online; accessed 07-Aug-2020].

[13] "EPFL Cough for COVID-19 Detection," https://coughvid.epfl.ch/, 2020, [Online; accessed 07-Aug-2020].

[14] "CMU sounds for COVID Project," https://node.dev.cvd.lti.cmu.edu/, 2020, [Online; accessed 07-Aug-2020].

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

[16] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," in *Proc. 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: Association for Computing Machinery, 2020, p. 3474–3484. [Online]. Available: https://doi.org/10.1145/3394486.3412865

[17] 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.

[18] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, N. R., P. K. Ghosh, and S. Ganapathy, "Coswara – a database of breathing, cough, and voice sounds for COVID-19 diagnosis," in *Proc. INTERSPEECH, ISCA*, 2020.

[19] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[20] B. McFee, V. Lostanlen, A. Metsai, M. McVicar, S. Balke, C. Thomé, C. Raffel, F. Zalkow, A. Malek, Dana, K. Lee, O. Nieto, J. Mason, D. Ellis, E. Battenberg, S. Seyfarth, R. Yamamoto, K. Choi, viktorandreevichmorozov, J. Moore, R. Bittner, S. Hidaka, Z. Wei, nullmightybofo, D. Hereñú, F.-R. Stöter, P. Friesch, A. Weiss, M. Vollrath, and T. Kim, "librosa/librosa: 0.8.0," Jul. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3955228

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.