



Analyzing the impact of SARS-CoV-2 variants on respiratory sound signals

Debarpan Bhattacharya¹, Debottam Dutta¹, Neeraj Kumar Sharma¹, Srikanth Raj Chetupalli¹,
Pravin Mote¹, Sriram Ganapathy¹, Chandrakiran C², Sahiti Nori², Suhail K K², Sadhana
Gonuguntla³, Murali Alagesan⁴

¹LEAP lab, Indian Institute of Science, Bangalore, India, ²Ramaiah Medical College Hospital, Bangalore, India, ³General Hospital, Hoskote, Bangalore, India, ⁴PSG Institute of Medical Sciences and Research, India
sriramg@iisc.ac.in

Abstract

The COVID-19 outbreak resulted in multiple waves of infections that have been associated with different SARS-CoV-2 variants. Studies have reported differential impact of the variants on respiratory health of patients. We explore whether acoustic signals, collected from COVID-19 subjects, show computationally distinguishable acoustic patterns suggesting a possibility to predict the underlying virus variant. We analyze the Coswara dataset which is collected from three subject pools, namely, i) healthy, ii) COVID-19 subjects recorded during the delta variant dominant period, and iii) data from COVID-19 subjects recorded during the omicron surge. Our findings suggest that multiple sound categories, such as cough, breathing, and speech, indicate significant acoustic feature differences when comparing COVID-19 subjects with omicron and delta variants. The classification areas-under-the-curve are significantly above chance for differentiating subjects infected by omicron from those infected by delta. Using a score fusion from multiple sound categories, we obtained an area-under-the-curve of 89% and 52.4% sensitivity at 95% specificity. Additionally, a hierarchical three class approach was used to classify the acoustic data into healthy and COVID-19 positive, and further COVID-19 subjects into delta and omicron variants providing high level of 3-class classification accuracy. These results suggest new ways for designing sound based COVID-19 diagnosis approaches.

Index Terms: COVID-19, SARS-CoV-2 variants, Omicron, cough, breathing, vowel, counting, and speech.

1. Introduction

Since the global outbreak of COVID-19 in March 2020, the world has been responding to disruptions across the health, education, and economic sectors. Among various measures to contain the pandemic, rapid diagnosis and isolation of COVID-19 patients helped to control the scale of SARS-CoV-2 infections [1]. Over the time scale of more than 24 months, the SARS-CoV-2 virus has continued to make mutations in its gene structure. Studies such as He et. al. [2] suggest the possibility of numerous mutations in future, and highlight the need for adopting diagnosis and treatment strategies to combat them in the shortest possible time. Such mutated variants of the virus are identified as variants of concern (VoC).

Two such VoCs are the delta (B.1.617.1) and the omicron (B.1.1.529) variants, which have caused significant rise in the COVID-19 cases in the Indian subcontinent [3] in two non-overlapping time spans. The delta variant related COVID-19 outbreak happened in April-June 2021 and the omicron variant

related outbreak happened in Jan-Feb 2022 [4, 5]. Interestingly, a recent study by Hui et. al. [6] showed that omicron variants had a higher replication rate in bronchus and lesser replication rate in the lungs, as compared to the delta variant. Drawn by this, we hypothesize that the respiratory sounds of COVID-19 patients might encode a variant dependent attribute. That is, we are interested in exploring if the respiratory sound samples of a COVID-19 patient can help identify the category of the variant causing the infection. To validate this hypothesis, this paper presents a diverse sets of experiments with sound samples collected from healthy and COVID-19 patients across multiple waves of SARS-CoV-2 infection in the Indian subcontinent [7, 8].

To our knowledge, none of the COVID-19 audio datasets have recorded the SARS-CoV-2 variant information for the COVID-19 patients. However, the Coswara dataset [9], an open-access COVID-19 audio dataset, provides the timestamp corresponding to the date each COVID-19 patient contributed their audio data to the dataset. As a work around to obtain information related to variant identity, we match the variant identity to the one which was widespread in the locality on the date of audio data contribution. Based on this, a majority of the COVID-19 patients contributed data to the Coswara dataset during the outbreak of the delta and omicron variants in India. The Coswara dataset contains audio signals from nine sound categories, collected via crowdsourcing and majorly from India. This study aims to classify delta and omicron variants based on the claim, set forth by the medical community, that they are lower and upper respiratory diseases respectively. It presents discriminability analysis, both statistically and using LSTM classifiers, with the 9 types of respiratory sounds present in the Coswara dataset.

We observe that multiple sound categories, such as cough, breathing, and speech, indicate significant acoustic feature difference when comparing COVID-19 subjects with omicron variant versus delta variant. Further, the classification AUCs are significantly above chance for a omicron versus delta variant detection using sound samples only. Using fusion strategy on predictions obtained from multiple sound categories, we obtain AUC of 89% and sensitivity of 52.4% at 95% specificity.

2. Materials

A subset of Coswara dataset based on selecting subjects with: all audio files being of good quality, the country location being India, age between 15-90 years, and belonging to either healthy or COVID-19 positive category, is drawn. This results in a dataset consisting of 1169 healthy, and 560 COVID-19 positive subjects. Based on the timestamp associated with the date of au-

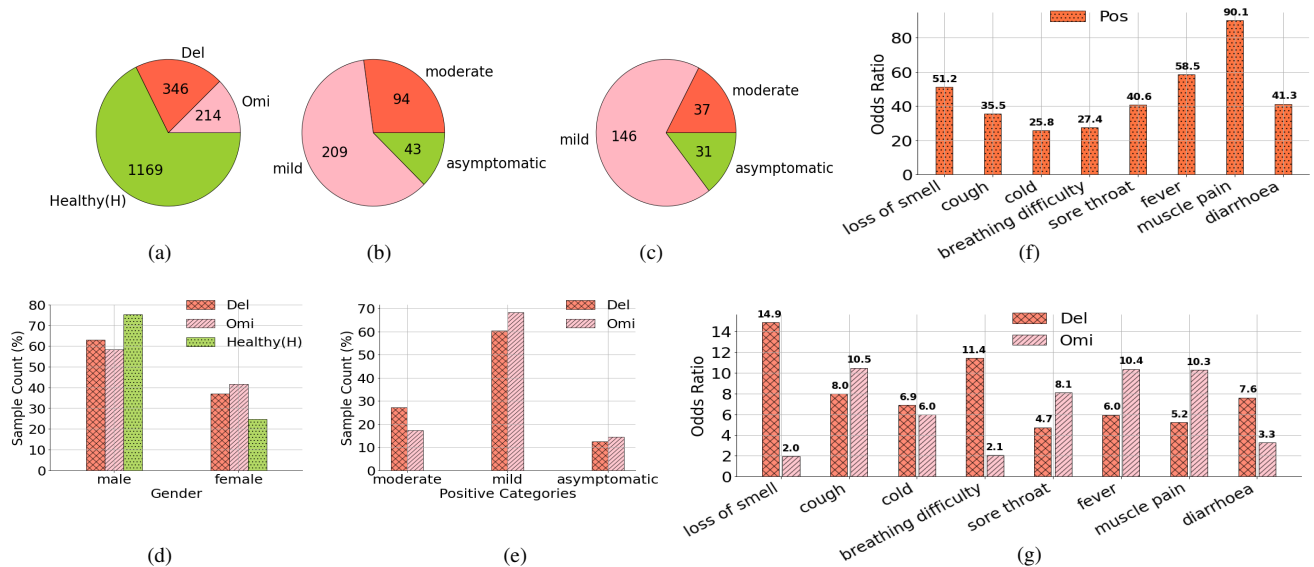


Fig. 1: (a) Pie plot representing number of samples for different subject categories as in Table 1; (b),(c) pie plot representing number of samples for different COVID severity conditions for delta and omicron positive samples respectively; (d) gender metadata plot; (e) COVID severity plot for positive samples in percentage; (f) odds ratio of different symptoms for positive samples; (g) comparison of odds ratio for different symptoms for omicron and delta positive samples.

Subject Category	Description
Healthy (<i>H</i>)	Healthy subjects, not infected with SARS-CoV-2
Delta (<i>Del</i>)	Infected with SARS-CoV-2 with a strain earlier to Omicron outbreak
Omicron (<i>Omi</i>)	Infected with SARS-CoV-2 with a strain during Omicron outbreak
Positive (<i>Pos</i>)	Del and Omi subjects pooled
Sound Category	Description
Breathing-shallow/deep	Few cycles of inspiration/expiration: with shallow/deep exertion on lungs
Cough-shallow/heavy	Few bouts of cough: with shallow/heavy exertion on lungs
Vowel-[u]/[i]/[æ]	Sustained phonation of three vowels: [u] in boot, [i] in beet, [æ] in bat
Counting-normal/fast	Counting numbers from 1 to 20 at: normal/fast pace

Table 1: Subject and sound categories analyzed in this study

dio data contribution to the dataset, the 560 COVID-19 subjects were further divided into two categories, namely, delta (346) and omicron (214). Each subject contributed nine audio signals to this dataset. The sound sample categories are described in Table 1. In the past, several studies have shown evidence for diagnosing a multitude of pulmonary diseases via computational analysis of these sound categories, example see [10, 11, 12]. Apart from the sound recordings, a diverse set of metadata information was also collected consisting of age/gender, broad geographic location, current health status w.r.t. symptoms associated with COVID-19, as well as the COVID-19 testing status. In addition, the COVID-19 positive samples are classified as one of the categories: asymptomatic, mild, and moderate based on the severity of the symptoms in the subject. The samples were not collected from severe patients considering their health condition. The metadata statistics for the subset of the data used in this study in terms of gender, COVID-19 severity and odds-ratio

of symptoms, are reported in Figure 1. The odds-ratio is defined as the ratio of the subjects exhibiting a particular symptom and being positive with particular variant to the ratio of subjects who have the same symptoms but are not positive with the particular variant. Odds ratio of more than 1 denotes positive correlation between the symptom and the disease. As seen in figure 1, the odds-ratio of positive subjects is highest for the condition of muscle pain. Further, among the variants, delta variant had higher odds-ratio for 'loss-of-smell', 'breathing-difficulty' and 'diarrhea', while omicron variant had higher odds-ratio for 'cough', 'sore-throat', 'fever' and 'muscle-pain'.

3. Methodology

3.1. Feature extraction

All the audio files are passed through a sound activity detector (SAD) after normalizing the sample range to ± 1 . Subsequently, any audio sample residing below a threshold of 0.01 is discarded. The log mel-spectrogram features are used for the analysis in the paper. Majority of the audio recordings have sampling rate of 48kHz. All the audio recordings are first re-sampled to 44.1 kHz. The log mel-spectrogram features are computed using a window of 1024 samples, a stride of 441 samples and with 64 mel-filters. The delta and delta-delta features, which are first and second derivatives of mel-spectrogram along the time axis respectively, are appended to the mel-spectrogram. Thus, the features are of size $192 \times N_k$, where N_k denotes number of short-time features obtained from k -th audio file.

3.2. Statistical analysis

We pursue a statistical analysis comparing the acoustic features derived from subjects belonging to a pair of conditions. For this analysis, the mel-spec feature matrix computed for each audio file is averaged across the N_k frames to obtain a 192×1 dimensional average mel-spec feature vector. Subsequently, we obtain two populations of such feature vectors, one corresponding to each subject category. We do a dimension-wise (across

192 dimensions) Mann-Whitney U test [13] to statistically compare the feature values between two populations. The Mann-Whitney U test is a non-parametric test of the null hypothesis that, for randomly selected values of X and Y from two populations, the probability of X being greater than Y is equal to the probability of Y being greater than X . For summarizing the statistical significance across all the feature dimensions, we compute the harmonic mean of p-values (HMP) from the 192 p-values computed over each dimension of the average mel-spec feature vector. The HMP has been shown to be a robust measure for p-value summarization in high dimensional data analysis [14].

3.3. Classification

Using the acoustic features computed in Section 3.1, we pursue the classifications tasks of i) *Omi-vs-Del*, ii) *Omi-vs-H*, iii) *Del-vs-H*, iv) *Pos-vs-H*, and v) *Omi-vs-Del-vs-H*. Here, *Omi* denotes the positive samples from the omicron variant, *Del* denotes the positive samples from the delta variant and *H* denotes the healthy subject samples.

For the three class hierarchical classification problem among *Omi*, *Del* and *H* samples, the given audio sample is first processed with the *Pos-vs-H* classifier. If the sample is classified as positive, it is processed through the *Omi-vs-Del* classifier.

3.4. Dataset division

We divide the pool of four subject categories, *Omi*, *Del*, *Pos* and *H*, into train, validation and test sets by drawing a random subset of 65% of the subjects for train, 15% for validation, and 20% subjects for testing. The split of the omicron and delta positive samples is done in such a way that COVID-19 severity groups (asymptomatic, mild and moderate) are stratified. For each of the classification tasks, the train and val splits are randomly sampled using 10 different seeds, without changing the test split.

3.5. Classifier model

We use the bidirectional long-short term memory (BLSTM) based neural network architecture as the classification model. This model served as a baseline in the Second DiCOVA Challenge [15]. The architecture is composed of two BLSTM layers having 128 cells, followed by a feed-forward network having 64 neurons and a $\tanh(\cdot)$ non-linearity. The final output is a single neuron having sigmoid activation and the output represents the probability of positive class.

Training: The model is trained using audio segments, as described in [15]. Each log mel-spectrogram is chunked into segments having contiguous frames, using 51 frames window size with a stride of 10 frames. All the segments generated from an audio recording are assigned the same class. During training, the minority class segments are over-sampled to reduce the impact of the class imbalance. The model is trained using the binary cross entropy (BCE) loss.

Inference: The test audio file is chunked into multiple segments of 51 frames. The output probabilities are calculated for all segments and their mean is computed to obtain the score for the audio file.

Performance Evaluation: We use the area-under-the-curve (AUC) measure of the receiver operating characteristic curve (ROC) [16] for quantifying binary classifier performance. The AUC is computed using the trapezoidal rule [17]. An AUC of

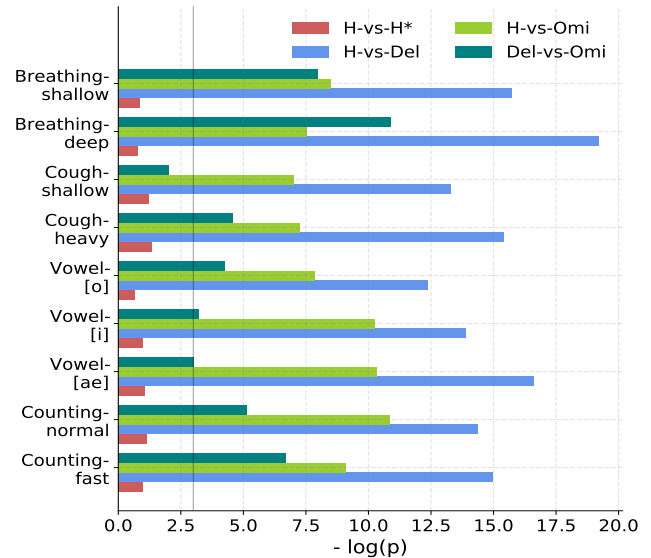


Fig. 2: Illustration of statistical analysis using Mann-Whitney U test comparing acoustic features of different subject populations. The (gray) line indicates statistical significance corresponding to $p = 0.001$. For this comparison, a random subset of 200 subjects is used from each subject category. H and H^* are two different random subset from the healthy subject pool.

0.5 and 1 indicate the chance and best performance, respectively.

4. Results

4.1. Statistical Analysis

The statistical significance obtained following the approach described in Section 3.2 is shown in Figure 2. A higher value for $-\log_{10}(p)$, which is > 3 , indicates that the acoustic features compared between the two subject populations are statistically significantly different. For a fair comparison in terms of population size, a random subset of 200 subjects is considered for each subject category. As expected, there is no significant

difference between two different subsets of the healthy pool (denoted as H and H^*). The significance increases for *Omi-vs-H* and *Del-vs-H* subject categories. This suggests that a binary classifier can be designed to separate healthy subjects from COVID-19 positive subjects. Interestingly, for a majority of sound modalities, there is also a significant difference between *Del* and the *Omi* subject categories, indicating that there are significant acoustic differences for the different variants of the SARS-CoV-2.

4.2. Classification Results

Four categories of two class classification problems are set up namely, *Omi-vs-Del*, *Omi-vs-H*, *Del-vs-H* and *Pos-vs-H* as described in section 3.4. The AUC distribution for the 9 sound modalities and the 4 classification tasks using the 10 validation folds is shown in Figure 3. As seen in the figure, the cough sound modality does not provide a good separation in the *Omi-vs-Del* classification. The speech counting is the best performing modality for *Omi-vs-Del* classification task. The performance of the vowel sound based classifiers is inferior to the other modalities of breathing sounds and speech-counting modalities.

Further, the test AUC values are computed, as given in Ta-

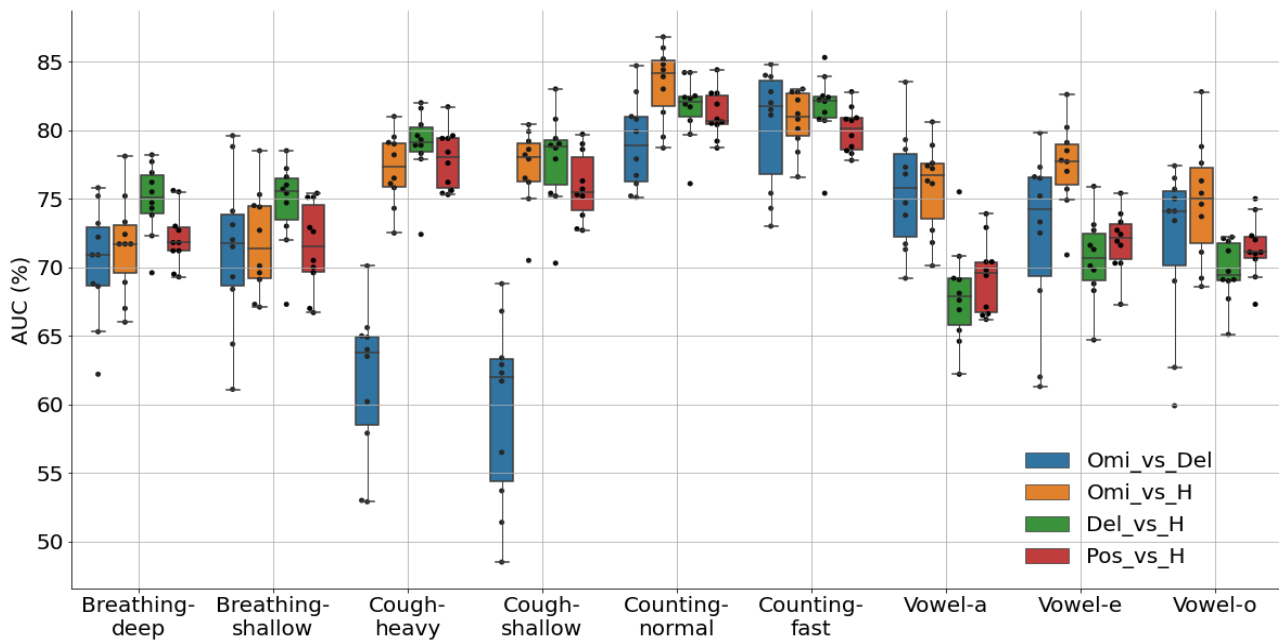


Fig. 3: 10 fold validation AUCs for 9 categories of sounds and for the 4 classification tasks.

Task	Breathing-deep	Breathing-shallow	Cough-heavy	Cough-shallow	Counting-fast	Counting-normal	Vowel- $[\text{æ}]$	Vowel- $[\text{e}]$	Vowel- $[\text{o}]$	Fusion (Sensitivity at 95% Specificity)
Omi-Del	79.9	73.4	65.2	65.7	81.4	79.8	73.0	75.3	77.9	89.0 (52.4)
Del-H	73.3	70.1	72.2	75.0	81.3	80.5	70.9	70.3	71.1	88.3 (68.6)
Omi-H	80.2	76.9	81.2	83.7	84.0	82.1	75.8	79.5	74.9	93.1 (71.4)
Pos-H	76.1	74.4	76.7	80.1	83.4	82.0	75.4	76.6	75.0	90.5 (71.4)

Table 2: Depiction of AUC for different classification tasks on the held out test set.

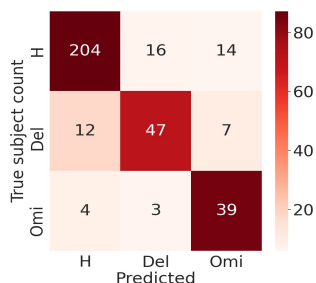


Fig. 4: Confusion matrix for hierarchical 3 class classification

ble 2. Among all the sound modalities, the counting-fast is the category that provides more than 80% ROC-AUC for all the 4 binary classifications tasks. Further, a simple fusion of the acoustic scores from the different modalities (average of probability scores from each of the classifiers trained on the individual modalities), achieves an AUC of 93.1% for classifying *Omi-vs-H* and 90.5% AUC for the *Pos-vs-H* classification. A hierarchical classifier, designed using a serial combination of *Pos-vs-H* followed by the *Omi-vs-Del* classifier yields a diagonally dominant confusion matrix on the test data, as shown in Figure 4. The confusion matrix shows that the proposed modeling and classification approaches allow the separation of healthy and positive samples, while further indicating the possibility of variant identification within the positive samples.

5. Conclusion

Analyzing the acoustic features of healthy and SARS-CoV-2 infected subjects in India, we find a statistically significant difference when comparing a pair of subject categories, namely, healthy versus COVID-19 positive (both for delta and omicron variants). Further, there is also a significant difference in acoustic features for audio collected from COVID-19 positive subjects with delta variant versus subjects with omicron variant. The binary classifiers designed to distinguish healthy from COVID-19 positive subjects performed significantly above chance across the different pairs of subject categories. Our findings suggest that the impact on the respiratory health of the COVID-19 subject with the omicron variant is different from that of the patients infected by the delta variant. While this study is not based on genomic confirmation of the SARS-CoV-2 variants on the subject pools, we still hypothesize that the insights developed in this paper encourage a more detailed study to allow the identification of SARS-Cov-2 variants from audio samples.

6. Acknowledgement

The authors thank the Department of Science and Technology, Govt. of India, for funding the Coswara Project through the RAKSHAK program. The authors express their gratitude to all the individuals who voluntarily contributed to the dataset, Anand Mohan for designing the web-based audio data collection platform, and Aditya Jana, Drishti Gupta, Priyanka Ramesh and Riya Bhattacharya for coordinating the data collection.

7. References

- [1] C. Pramesh, G. R. Babu, J. Basu, I. Bhushan, C. M. Booth, G. Chinnaswamy, R. Guleria, S. Kalantri, G. Kang, P. Mohan *et al.*, “Choosing wisely for COVID-19: ten evidence-based recommendations for patients and physicians,” *Nature Medicine*, vol. 27, no. 8, pp. 1324–1327, 2021.
- [2] X. He, W. Hong, X. Pan, G. Lu, and X. Wei, “Sars-cov-2 omicron variant: characteristics and prevention,” *MedComm*, 2021.
- [3] R. K. Mohapatra, R. Tiwari, A. K. Sarangi, S. K. Sharma, R. Khandia, G. Saikumar, and K. Dhama, “Twin combination of omicron and delta variant triggering a tsunami wave of ever high surges in covid-19 cases: a challenging global threat with a special focus on Indian sub-continent,” *Journal of medical virology*, 2022.
- [4] R. Gopal, V. Chandrasekar, and M. Lakshmanan, “Analysis of the second wave of COVID-19 in India based on SEIR model,” *The European Physical Journal Special Topics*, pp. 1–8, 2022.
- [5] R. Ranjan, “Omicron impact in india: An early analysis of the ongoing COVID-19 third wave,” *medRxiv*, 2022.
- [6] K. P. Hui, J. C. Ho, M.-c. Cheung, K.-c. Ng, R. H. Ching, K.-l. Lai, T. T. Kam, H. Gu, K.-Y. Sit, M. K. Hsin *et al.*, “Sars-cov-2 omicron variant replication in human bronchus and lung ex vivo,” *Nature*, pp. 1–5, 2022.
- [7] A. Muguli, L. Pinto, N. Sharma, P. Krishnan, P. K. Ghosh, R. Kumar, S. Bhat, S. R. Chetupalli, S. Ganapathy, S. Ramoji *et al.*, “DiCOVA challenge: Dataset, task, and baseline system for covid-19 diagnosis using acoustics,” *arXiv preprint arXiv:2103.09148*, 2021.
- [8] N. K. Sharma, A. Muguli, P. Krishnan, R. Kumar, S. R. Chetupalli, and S. Ganapathy, “Towards sound based testing of covid-19—summary of the first diagnostics of covid-19 using acoustics (dicova) challenge,” *Computer Speech & Language*, vol. 73, p. 101320, 2022.
- [9] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, R. Nirmala, P. K. Ghosh, and S. Ganapathy, “Coswara – a database of breathing, cough, and voice sounds for COVID-19 diagnosis,” in *Proc. Interspeech*, 2020, pp. 4811–4815.
- [10] D. G. Tinkelman, C. Lutz, and B. Conner, “Analysis of breath sounds in normal and asthmatic children and adults using computer digitized airway phonopneumography (cdap),” *Respiratory medicine*, vol. 85, no. 2, pp. 125–131, 1991.
- [11] M. Al-khassaweneh, S. B. Mustafa, and F. Abu-Ekteish, “Asthma attack monitoring and diagnosis: A proposed system,” in *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*. IEEE, 2012, pp. 763–767.
- [12] V. Nathan, K. Vatanparvar, M. M. Rahman, E. Nemati, and J. Kuang, “Assessment of chronic pulmonary disease patients using biomarkers from natural speech recorded by mobile devices,” in *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2019, pp. 1–4.
- [13] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” *The Annals of mathematical statistics*, pp. 50–60, 1947.
- [14] D. J. Wilson, “The harmonic mean p-value for combining dependent tests,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 4, pp. 1195–1200, 2019.
- [15] N. K. Sharma, S. R. Chetupalli, D. Bhattacharya, D. Dutta, P. Mote, and S. Ganapathy, “The Second DiCOVA Challenge: Dataset and performance analysis for COVID-19 diagnosis using acoustics,” *accepted in IEEE ICASSP*, 2022.
- [16] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, rOC Analysis in Pattern Recognition.
- [17] 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.