



Contrastive Learning of Cough Descriptors for Automatic COVID-19 Preliminary Diagnosis

Swapnil Bhosale[§], Upasana Tiwari[§], Rupayan Chakraborty, Sunil Kumar Kopparapu

TCS Research, Tata Consultancy Services Limited, Mumbai, India

Abstract

Cough sounds as a descriptor have been used for detecting various respiratory ailments based on its intensity, duration of intermediate phase between two cough sounds, repetitions, dryness etc. However, COVID-19 diagnosis using only cough sounds is challenging because of cough being a common symptom among many non COVID-19 health diseases and inherent data imbalance within the available datasets. As one of the approach in this direction, we explore the robustness of multi-domain representation by performing the early fusion over a wide set of temporal, spectral and tempo-spectral handcrafted features, followed by training a Support Vector Machine (SVM) classifier. In our second approach, using a contrastive loss function we learn a latent space from Mel Filter Cepstral Coefficients (MFCCs) where representations belonging to samples having similar cough characteristics are closer. This helps learn representations for the highly varied COVID-negative class (healthy and symptomatic COVID-negative), by learning multiple smaller clusters. Using only the DiCOVA data, multi-domain features yields an absolute improvement of 0.74% and 1.07%, whereas our second approach shows an improvement of 2.09% and 3.98%, over the blind test and validation set, respectively, when compared with challenge baseline.

Index Terms: COVID-19 diagnosis, machine learning, contrastive loss, multi-domain cough descriptors.

1. Introduction

As of March 2021, the outbreak of COVID-19 has proliferated to over 219 countries estimating a loss of about 20.5 million of life worldwide [1]. Sensitivity and specificity metric based study for RT-PCR (Reverse Transcription Polymerase Chain Reaction) tests to estimate clinical effectiveness for COVID-19 diagnosis have already been conducted [2]. It is evident that extensive testing is a key measure in preventing the spread of COVID-19 and identifying possible hotspots. However, many countries are facing major challenges in scaling up testing capacity while simultaneously ensuring quality-assured testing and preventing the scale back of essential diagnostic services for other diseases [3][4]. Alternative capabilities in screening for early signs can help mitigate this substantially. For instance, in recent times the use of cough sounds as a potential biomarker for pre-screening as well as possible diagnosis has been widely studied [5],[6],[7]. Major approaches have been combining the knowledge of audio signal processing and deep models that could potentially extract relevant representations for distinguishing COVID-positive cough sounds from a COVID-negative cough. In our work, besides temporal and spectral dynamics, we also incorporate the temporal variation in the spectral dynamics of a cough audio signal, that com-

plements the discriminatory characteristics between COVID-positive and COVID-negative samples.

Segregating COVID-positive cough samples from symptomatic COVID-negative cough samples is still an open problem to solve. Moreover, cross-dataset evaluation is extremely difficult mainly due to (a) less number of annotated samples in databases, (b) crowd-sourced nature of the available databases, (c) limited recording environments, symptomatic COVID-negative samples and expert annotations [8]. It is evident for such systems to overfit, by learning inherent bias within the data such as recording environment rather than the robust discriminatory characteristics. To cater for the class imbalance, in this paper, we transform the task as a 2-way few-shot learning problem, where 2-way represents the two classes i.e., COVID-positive and COVID-negative. During training, a triplet loss is used to learn the latent representations from Mel-Frequency Cepstral Coefficients (MFCC) based features such that representations of audio samples having similar audio characteristics are closer. Post-training the few-shot model, we employ a logistic regression classifier with the learned embeddings as inputs. We also investigate the usefulness of multi-domain (spectral, temporal, and tempo-spectral) audio descriptors in COVID-19 cough detection task. To summarize, the main contributions of this work are (a) we devise a cough event detector based on simple filtering strategies applied on the raw time endpoints generated from a pre-trained Audio Event Detection (AED) model trained on generic sound events; (b) we propose a multi-domain representation of handcrafted features which is capable of incorporating spectral, temporal, and tempo-spectral characteristics of a cough audio signal; (c) we propose a few-shot learning model based on episodic triplets to extract the inter-cluster information between COVID-positive and COVID-negative samples, as well as the intra-cluster characteristics within the highly varied COVID-negative cough samples. We observe that it helps in treating outlier samples based on the duration and the cough-related audio content; something quite prevalent in crowdsourced data. Our experiments show that both (b) and (c) yield a higher performance compared to baseline Artificial Neural Network (ANN) classifier with MFCCs as features.

The remainder of this paper is organized, as follows. In section 2, we provide a brief literature review of the related work. In section 3, we explain in detail the system design and the proposed approach with the experimental details. The results are presented in section 4, followed by conclusion in section 5.

2. Related Work

Cough sounds being one of the prominent markers in detecting COVID-19 infection, has gained a rapid attention among researchers in order to evaluate its potential in COVID-19 diagnosis. In literature, there are several works [9, 7, 10, 8, 11, 12, 13, 14, 6] that mainly concentrated on cough recordings for COVID-19 detection. In [9], the authors use curated set

[§] equal contribution

(141 COVID patients; cough and breathing sounds) of the large crowdsourced data (10,000 samples from about 7000 unique participants) longitudinally collected as of May, 2020. In addition to using temporal and spectral handcrafted features, they extracted the embeddings from VGGish based transfer learning framework. The system report an area-under-the-curve (AUC) of 80% with a logistic regression classifier. The study in [7], uses a CNN based cough detector trained on spectrograms followed by ensemble of three classifiers which are transfer learning based multi-class and binary classifier; and a classical machine learning based multi-class classifier. The system reports a sensitivity of 94% and 91% of specificity. Unlike most of COVID-19 data collection which are crowdsourced, highly skewed and have self-reported labels by the participants, dataset introduced by [10] consists of 3621 individuals with 2001 COVID-positive labels directly received from healthcare authorities. The work uses a curated set which is restricted to only cough sounds (376 COVID-positives). They pre-trained a CNN architecture with a initial ResNet18 block on the larger dataset of cough samples collected from various public datasets to enhance the end-to-end learning. Moreover, the authors also use data augmentation to boost the performance and label smoothing for enhancing the model calibration. Their system achieved an AUC of 72%. [11] uses a medical dataset containing 328 cough recordings from 150 individuals recorded in controlled setting. They use an ensemble of Deep Neural Network (DNN) model trained on handcrafted cough features and transformer-based network called TabNet trained to learn the symptom embedding (using demographic and symptoms: fever, headache, aches, sore throat, etc). Their system achieved an accuracy of 90.08% using only cough data, 91.1% using only symptoms data and 96.5% when both data are used.

3. System Description

Unlike detection of other respiratory infections, for COVID-19 it is necessary to identify markers that discriminate beyond forced cough (simulated) and ailment cough. The large variations within the COVID-negative (healthy and diseases other than COVID-19) samples, make it difficult to learn representations into two non-overlapping clusters. Moreover, most COVID-19 datasets include high imbalance among the COVID-positive and healthy subjects. Fig. 1 represents the framework of our proposed approach.

3.1. Methodology

In order to address the above mentioned challenges, we propose a few-shot based approach such that the classification task is now converted to a 2-way N -shot learning problem where the ways represent 2 classes, i.e. COVID-positive and COVID-negative, and N shots represents N audio samples each from both the classes, provided as reference samples. Each query sample is then inferred depending on the distance between the learned representations of query and the provided reference samples. Interestingly, these representations are learned using an episodic triplet loss, which enables the model to learn the intra-clusters among the highly varied COVID-negative class. Moreover, during training, the loss in each episode is computed w.r.t to a fixed number of shots for each class, thereby maintaining a balance between the number of samples for both the classes [24] [25]. It is important to note that, in a supervised approach each training sample is equivalent to a single training configuration, i.e. the same sample shown multiple times to a model (during training) will make the model overfit (memorize it), however,

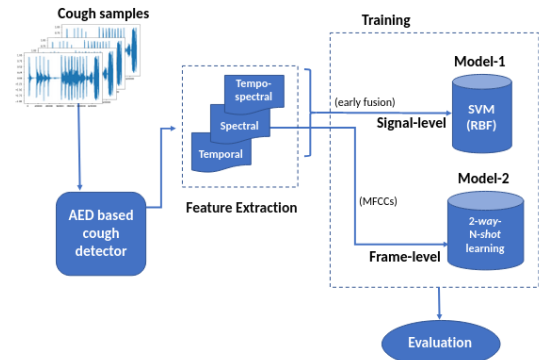


Figure 1: Framework

in this case a single sample may yield different loss values (and hence different possibilities of attaining convergence), depending on the reference samples (also called as support set) provided with it in each training iteration (also called as episode). In addition to the above approach, we also followed a feature-based machine learning technique where we explored the robustness of multi-domain cough biomarkers with a simple Support Vector Machine (SVM) classifier. Here, we experimented with a wide set of handcrafted features capturing the temporal, spectral and tempo-spectral descriptors (discussed in detail in Section 3.3) to learn the cough patterns that can discriminate between COVID-positive and COVID-negative samples.

3.2. Localizing cough-relevant regions using a pre-trained AED model

Analysing the spectrograms of audio waveforms, we observe that majority of the samples show high spectral spread under 8 kHz. Thus, we resampled each audio at 16 kHz using librosa audio processing library [26]. Because of its crowdsourced nature, the provided data consists of outliers in terms of both, the duration of the samples (see Fig.2) and the amount of undesirable audio content other than cough-related sounds (i.e. background noise, human speech etc) present. Towards this direction, we utilized a pre-trained AED model, YAMNet [27], trained on generic audio events (including cough sounds), to localize two types of regions, namely,

- **COUGH SOUNDS** events in the same subtree of actual 'cough' event from knowledge graph¹, i.e. semantically similar, example, throat clearing and other events (not in the subtree) but acoustically similar, example, plop.
- **OTHER** all audio events other than those mentioned above.

Once individual regions are obtained, a smoothing operation is applied to remove discontinuous regions (regions lesser than 500 ms). We specify two criteria based on the number of detected regions and their corresponding time durations, (a) no **COUGH SOUNDS** region detected throughout the duration of the audio sample, and (b) detected **OTHER** region is more than 4 secs. Audio files satisfying these criteria, are further split into smaller audio samples (to be used as individual training samples) based on detected regions. It is important to note that this even filters out smaller duration audio files with undesirable audio content. This gives a total of 1073 samples, fairly consistent across their duration (see Fig. 2b) as well as inclusion of audio content relevant to cough sounds.

¹please refer to Audioset ontology at <https://research.google.com/audioset/ontology/index.html>.

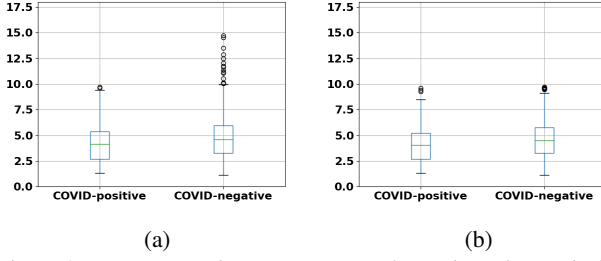


Figure 2: Variation in duration (y-axis) of samples. The marked circles depict the outliers (on the basis of duration). (a) Original (1040 samples), (b) Pre-processed (1073 samples)

3.3. Cough Sound Descriptors

In this section, we elaborate the tempo-spectral, spectral, and temporal feature extraction process in the context of a cough audio signal.

- **Discrete Wavelet Transformation (DWT) Based Descriptors:** DWT is a signal decomposition technique that represents the temporal changes in spectral dynamics of a signal. Cough sounds, being non-stationary in nature, can be better represented by tempo-spectral structure that captures the dynamic changes in signal. Here, we use *db3* as a mother wavelet due to their efficient time frequency localization properties in cough sound analysis [15, 16]. The choice for level of decomposition is based on prominent frequency components of the signal. Since cough sounds are known to have more energy in lower frequencies (varying from 20 Hz - 50 Hz in different studies) [28, 22, 7], we decomposed the signal till 10th level using PyWavelets tool [29]. Thereafter, we extract a set of six features for each frequency band, namely, Energy, Entropy, Root Mean Square (RMS), Recursing Energy Efficiency (REE), Logarithmic REE (LREE), Absolute Logarithmic REE (ALREE) [30]. Thus, we get a 60-dimensional feature vector. We refer this feature set as F_{ts} .
- **Statistical Descriptors:** In order to capture the spectral properties in a cough sound we first compute 64 low level descriptors (LLDs) with 20 ms of window length and 10 ms of overlap, followed by computing the delta and delta-delta coefficients in order to extract the temporal mutual information among the adjacent frames. A complete list of LLDs consists: 13 MFCC components, Zero-crossing, Spectral Centroid, Roll-off Frequency, Spectral Flux, Spectral entropy, Spectral Spread, 12 Chroma components, Jitter (ratio, percentage, factor), 26 Mel Spectrograms, Fundamental Frequency (F_0), Log energy, Entropy. The usage of these features are well studied in previous works on cough sound detection [20, 9, 18, 31, 32]. All LLDs are smoothed over time with a symmetric moving average filter with length of 3 frames. To capture the distributions beyond the mean, we further computed high-level descriptors (HLDs), statistical features over LLDs. A complete list of 25 HLDs is: mean, median, range, maximum, minimum, position of minimum and maximum, percentile, 1st, 2nd and 3rd quartile, interquartile range, standard deviation, variance, skewness, kurtosis, linear and quadratic regression coefficients. In total, we get a 4800-dimensional feature vector. We use pyAudioAnalysis library to implement this feature set [33], and is referred as F_s .
- **Temporal Descriptors:** To capture the temporal characteristics of a cough signal, we extract the hjorth parameters (mobility and complexity) which have been proved to be a powerful biomarker in respiratory sounds [17, 19]. Motivated by

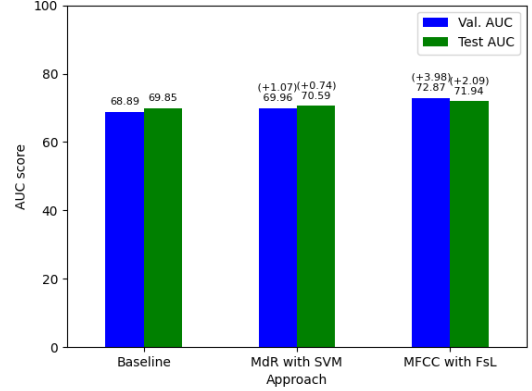


Figure 3: Performance of our proposed approaches on validation set and blind test set.

the wide usage of fractal dimension in lung sound analysis [21, 22], we compute the Petrosian Fractal Dimension (PFD) and Higuchi Fractal Dimension (HFD) of cough signals. Furthermore, we use Detrended Fluctuation Analysis (DFA) that captures the self-similarity within a time-series over a longer period, whose application has been explored in respiratory sounds [23]. We use an open source python module, PyEEG [34], to compute this 5-dimensional feature vector. We refer this feature set as F_t .

3.4. Methods

- **Using a Multi-domain Representation (*MdR*):** In this approach, we perform an early fusion over a wide set of hand-crafted features (as discussed in Section 3.3) to get a multi-domain representation of a cough signal. This results into 4865-dimensional ($5+4800+60$) feature vector. We refer this feature set as F_{t+s+ts} . Further, we train a SVM with RBF kernel on the resultant feature vector (as shown in Fig. 1). To prevent biased learning as a result of high imbalance in the data (COVID-positive=50, COVID-negative=772), we assign a higher weight to the minority class and lower to the majority class within the cost function. For this, we use an utility provided by python's *sklearn* which automatically adjusts weights inversely proportional to class frequencies in the data.
- **Using a Few-shot Learning (*FsL*) approach :** In this approach we use features similar to the baseline approach [35], i.e. 39 dimensional MFCCs combined with the delta and delta-delta features. Additionally, for an audio sample x with frame-level feature vectors, $x_i \in \mathcal{R}^{117}$, features from w succeeding and w preceding frames are concatenated. In our experiments, we use $w=3$ and thus, our input to the model is, $\bar{x}_i \in \mathcal{R}^{(7 \times \{3+1+3\})}$ ($7 \times \{3+1+3\}$) vector. The embedding block (referred as $f_\phi(\cdot)$ in [25]) consists of a stack of 2 dense layers, with 800, 512 hidden units each with ReLU activation. During training, in each episode a triplet loss is computed using the output from the last dense layer as the latent representation. Post-training, the output of last dense layer is used as embeddings to train a logistic regression classifier. Similar

** Specificity at 80% sensitivity.

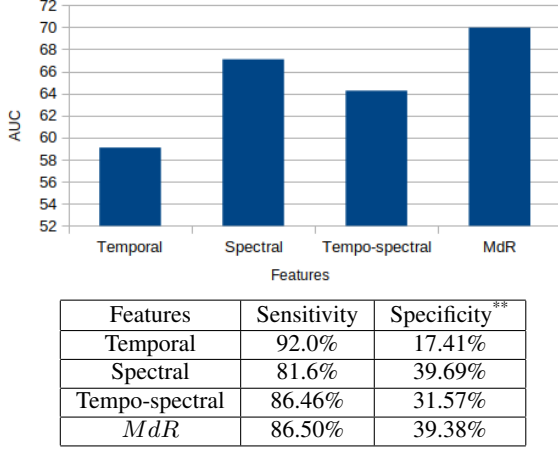


Figure 4: Performance of different handcrafted feature sets vs *MdR* (early fusion) on validation set.

to the baseline approach, our model provides the frame-level scores which are then averaged to obtain a prediction for the entire audio sample. We use number of shot as 5 (i.e. each training episode consists of 5 reference samples from each class) and margin = 0.5, where margin defines the maximum difference between the euclidean distances between anchor and positive sample, and between anchor and negative sample (as discussed in [25]).

4. Results

We compare our approaches with the provided baseline (replicated at our end) with area under the Receiver Operating Characteristic (ROC) curve as the metric (computed using probability score of a sample being COVID-positive). Further, specificity at 80% sensitivity [35] are also reported. The scores on blind test set as shown in Fig.3 have been provided by the challenge organizers. Our both approaches surpass the baseline when evaluated over the validation set as well as the blind test set. We hypothesize, this is primarily because of two reasons, firstly, robustness of handcrafted multi-domain representation in our first approach and the balanced episodes generated as a part of the few-shot pre-training accompanied by the triplet loss function which enables the intra class learning within the highly varied COVID-19 negative class in our second approach. Secondly, we use a balanced subset of the provided validation folds (since the provided validation folds are skewed), to decide on an early stopping criteria that guides the model convergence.

In order to analyse the independent performance of each feature set (i.e. F_t , F_s , F_{ts}) vs early fusion (F_{t+s+ts}), we evaluate them individually using the same classifier i.e. SVM (RBF). Fig 4 represents the average performance of each feature set on a 5-fold validation set. The 4800-dimensional F_s performs best among the three sets, whereas 5-dimensional F_t alone is unable to discriminate between the two classes. Despite the higher sensitivity obtained using F_t , its lower specificity compared to F_s can be accounted due to the lack of high number of statistical measures (refer to the HLD list, discussed in Section 3.3) computed over wider set of LLDs that are present in F_s . Interestingly, F_{ts} performs competitively w.r.t. to F_s , as a result of its capability to efficiently capture temporal variation in spectral dynamics of the cough sound (using DWT). Furthermore, early fusion (F_{t+s+ts}) boosts the model performance by complementing the discriminatory cough characteristics extracted from each domain.

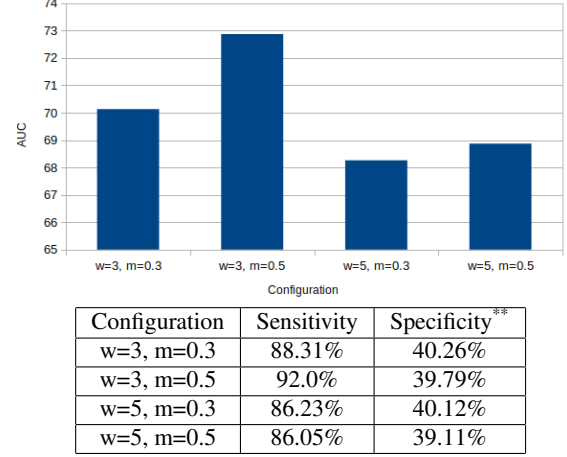


Figure 5: Performance of *F sL* with different configurations on validation set. { w : length of context window, m : margin}

Fig. 5 depicts the average performance on 5-fold validation set using the contrastive approach for different values of w and m . It can be clearly seen, that a shorter context (i.e. $w = 3$) with a larger margin performs better. The value of w decides the amount of temporal context associated with each individual sample, however, a larger context increases the feature dimensionality as well as the amount of overlapping information between two consecutive (training) samples. The margin value on the other hand, decides the difference between the distances of a sample with its positive and negative counterparts (i.e. other samples from the support set with the same or different label from that of the sample, respectively) in the latent space learned by the few-shot embedding model (output of $f_\phi(\cdot)$ [25]).

It is important to note that in both our approaches we use the samples provided as a part of the DiCOVA challenge, and do not use any external data. Compared to the challenge baseline that used an ANN model with MFCC features, our use of multi-domain features yields an absolute improvement of 0.74% and 1.07%, whereas our few-shot based model with MFCCs as inputs shows an improvement of 2.09% and 3.98%, over the blind test and validation set, respectively. Furthermore, *F sL* trained with *MdR* as input features, did not yield any further improvement. We believe the performance of our approaches could be further improved with access to larger databases with a fairly equal class distribution.

5. Conclusion

In this work, we proposed two novel approaches for diagnosing COVID-19 using cough sounds on the DiCOVA challenge dataset. Unlike most of the existing literature, which uses limited handcrafted features that mainly include either temporal or spectral features, a multi-domain fused feature set including tempo-spectral features shows a considerable improvement in terms of both specificity and AUC score. Our proposed few-shot learning based approach applied over simple MFCCs without explicit information regarding the cough domain, outperforms the baseline approach using the same features. Additionally, despite the high variation among the COVID-negative samples, our use of contrastive loss is able to learn the intra-cluster information within COVID-negative (healthy and symptomatic COVID-negative). We believe, this is a promising step in the direction of non-invasive real-time diagnosis of COVID-19.

6. References

- [1] H. P. i Arolas, E. Acosta, G. López-Casasnovas, A. Lo, C. Nicodemo, T. Riffe, and M. Myrskylä, “Years of life lost to covid-19 in 81 countries,” *Scientific reports*, vol. 11, 2021.
- [2] D. Jarrom, L. Elston, J. Washington, M. Prettyjohns, K. Cann, S. Myles, and P. Groves, “Effectiveness of tests to detect the presence of sars-cov-2 virus, and antibodies to sars-cov-2, to inform covid-19 diagnosis: a rapid systematic review,” *BMJ Evidence-Based Medicine*, 2020.
- [3] P. Ondoa, Y. Kebede, M. M. Loembe, J. N. Bhiman, S. K. Tessema, A. Sow, J. Nkengasong *et al.*, “Covid-19 testing in africa: lessons learnt,” *The Lancet Microbe*, vol. 1, no. 3, 2020.
- [4] R. W. Peeling, P. L. Olliaro, D. I. Boeras, and N. Fongwen, “Scaling up covid-19 rapid antigen tests: promises and challenges,” *The Lancet Infectious Diseases*, 2021.
- [5] J. Laguarda, F. Hueto, and B. Subirana, “Covid-19 artificial intelligence diagnosis using only cough recordings,” *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, 2020.
- [6] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, P. K. Ghosh, S. Ganapathy *et al.*, “Coswara—a database of breathing, cough, and voice sounds for covid-19 diagnosis,” *arXiv preprint arXiv:2005.10548*, 2020.
- [7] 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, 2020.
- [8] L. Orlandic, T. Teixeira, and D. Atienza, “The coughvid crowd-sourcing dataset: A corpus for the study of large-scale cough analysis algorithms,” *arXiv preprint arXiv:2009.11644*, 2020.
- [9] 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 *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.
- [10] 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.
- [11] A. Pal and M. Sankarasubbu, “Pay attention to the cough: Early diagnosis of covid-19 using interpretable symptoms embeddings with cough sound signal processing,” *arXiv preprint arXiv:2010.02417*, 2020.
- [12] “NYU Breathing Sounds for COVID-19,” <https://breatheforscience.com/>, 2020, [Online; accessed 07-Aug-2020].
- [13] “Cambridge University, UK - COVID-19 Sounds App,” <https://covid-19-sounds.org/en/>, 2020, [Online; accessed 07-Aug-2020].
- [14] “Cough Against COVID - Wadhwani AI Institute,” <https://coughagainstcovid.org/>, 2020, [Online; accessed 07-Aug-2020].
- [15] K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, “Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis,” *IEEE Transactions on Biomedical Engineering*, vol. 62, 2014.
- [16] T. H. Pingale and H. Patil, “Analysis of cough sound for pneumonia detection using wavelet transform and statistical parameters,” in *IEEE International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 2017.
- [17] A. Rizal, R. Hidayat, and H. A. Nugroho, “Lung sound classification using hjorth descriptor measurement on wavelet sub-bands,” *Journal of Information Processing Systems*, vol. 15, no. 5, pp. 1068–1081, 2019.
- [18] S. Z. H. Naqvi and M. A. Choudhry, “An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis,” *Sensors*, vol. 20, 2020.
- [19] A. Rizal, R. Hidayat, and H. A. Nugroho, “Determining lung sound characterization using hjorth descriptor,” in *2015 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC)*. IEEE, 2015, pp. 54–57.
- [20] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, “A cough-based algorithm for automatic diagnosis of pertussis,” *PloS one*, vol. 11, 2016.
- [21] A. Rizal, H. A. Nugroho, and R. Hidayat, “Fractal dimension for lung sound classification in multiscale scheme,” *Journal of Computer Science*, vol. 14, 2018.
- [22] S. Reichert, R. Gass, C. Brandt, and E. Andrès, “Analysis of respiratory sounds: state of the art,” *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, vol. 2, 2008.
- [23] G. C. Donaldson, T. A. Seemungal, J. R. Hurst, and J. A. Wedzicha, “Detrended fluctuation analysis of peak expiratory flow and exacerbation frequency in COPD,” *European Respiratory Journal*, vol. 40, 2012.
- [24] C. Huang, Y. Li, C. C. Loy, and X. Tang, “Deep imbalanced learning for face recognition and attribute prediction,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, 2019.
- [25] S. Bhosale, R. Chakraborty, and S. K. Kopparapu, “Semi supervised learning for few-shot audio classification by episodic triplet mining,” *arXiv preprint arXiv:2102.08074*, 2021.
- [26] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “librosa: Audio and music signal analysis in python,” in *Proceedings of the 14th Python in Science Conference*, vol. 8. Citeseer, 2015.
- [27] Tensorflow, “Yamnet,” <https://github.com/tensorflow/models/tree/master/research/audioset/yamnet>, 2020.
- [28] K. K. Lee, S. Matos, K. Ward, G. F. Rafferty, J. Moxham, D. H. Evans, and S. S. Birring, “Sound: a non-invasive measure of cough intensity,” *BMJ open respiratory research*, vol. 4, 2017.
- [29] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O’Leary, “Pywavelets: A python package for wavelet analysis,” *Journal of Open Source Software*, vol. 4, 2019.
- [30] M. Murugappan, N. Ramachandran, Y. Sazali *et al.*, “Classification of human emotion from eeg using discrete wavelet transform,” *Journal of Biomedical Science and Engineering*, vol. 3, no. 04, 2010.
- [31] T. Drugman, J. Urbain, N. Bauwens, R. Chessini, C. Valderrama, P. Lebecque, and T. Dutoit, “Objective study of sensor relevance for automatic cough detection,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, 2013.
- [32] M. J. Rahman, E. Nemati, M. Rahman, K. Vatanparvar, V. Nathan, and J. Kuang, “Efficient online cough detection with a minimal feature set using smartphones for automated assessment of pulmonary patients,” in *International Conference on Ambient Computing, Applications, Services and Technologies*, 2011.
- [33] T. Giannakopoulos, “pyaudioanalysis: An open-source python library for audio signal analysis,” *PloS one*, vol. 10, 2015.
- [34] F. S. Bao, X. Liu, and C. Zhang, “Pyeeeg: an open source python module for eeg/meg feature extraction,” *Computational Intelligence and Neuroscience*, vol. 2011, 2011.
- [35] A. Muguli, L. Pinto, N. Sharma, P. Krishnan, P. K. Ghosh, R. Kumar, S. Ramoji, S. Bhat, S. R. Chetupalli, S. Ganapathy *et al.*, “Dicova challenge: Dataset, task, and baseline system for covid-19 diagnosis using acoustics,” *arXiv preprint arXiv:2103.09148*, 2021.