INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Detecting COVID-19 from audio recording of coughs using Random Forests and Support Vector Machines

Isabella Södergren[1], *Maryam Pahlavan Nodeh*[2], *Prakash Chandra Chhipa*[2], *Konstantina Nikolaidou*[2], *György Kovács*[2]

[1]Digital Services and Systems, Luleå University of Technology, Luleå, Sweden
[2]EISLAB Machine Learning, Luleå University of Technology, Luleå, Sweden
[1]isasde-5@student.ltu.se, [2]firstname.[middlename].lastname@ltu.se

Abstract

The detection of COVID-19 is and will remain in the foreseeable future a crucial challenge, making the development of tools for the task important. One possible approach, on the confines of speech and audio processing, is detecting potential COVID-19 cases based on cough sounds. We propose a simple, yet robust method based on the well-known ComParE 2016 feature set, and two classical machine learning models, namely Random Forests, and Support Vector Machines (SVMs). Furthermore, we combine the two methods, by calculating the weighted average of their predictions. Our results in the DiCOVA challenge show that this simple approach leads to a robust solution while producing competitive results. Based on the Area Under the Receiver Operating Characteristic Curve (AUC ROC) score, both classical machine learning methods we applied markedly outperform the baseline provided by the challenge organisers. Moreover, their combination attains an AUC ROC score of 85.21, positioning us at fourth place on the leaderboard (where the second team attained a similar, 85.43 score). Here, we would describe this system in more detail, and analyse the resulting models, drawing conclusions, and determining future work directions.

Index Terms: COVID-19, acoustics, machine learning, respiratory diagnosis, random forest, SVM, OpenSmile

1. Introduction

The ongoing problem of the COVID-19 pandemic renders the task of cough classification, and thus challenges in the subject [1] topical. Because of this, we have dedicated significant effort to this task. One important lesson from earlier – similar – voice classification challenges was the benefit of combining various models; winners of many of these challenges using a method of ensembling [2, 3, 4, 5, 6], often times late fusion. Another guiding principle was the utility of classical machine learning models, particularly in scenarios where the amount of labeled training data is limited. Our efforts were also guided by the abundant related literature, including the already existing efforts for detecting COVID-19 from audio recordings, as well as earlier cough detection and classification efforts, and other works in audio signal classification. For this, in Section 1.1 we briefly introduce some of these efforts. The rest of the paper is organized as follows. In Section 2 we give a short introduction of the challenge and data set, followed by a discussion of our methodology is Section 3, then we present our results in Section 4, analyse our models in Section 5, and lastly, end our paper by presenting conclusions and future work in Section 6.

1.1. Related work

Researchers discovered human coronaviruses in the 1960s [7]. While some of its strains are responsible for the common cold, others cause severe diseases, like SARS, MERS, and lately COVID-19. In the case of patients affected by COVID-19, cough has shown to be one of the most prevalent symptoms in adults [8] as well as children and adolescents [9]. It is possible [10, 11] to diagnose several diseases like bronchitis, pertussis, and asthma by using cough sounds. And recent studies [12] show that cough- and respiratory sounds differ between COVID-19 patients and healthy patients as well. One reason for this is that the COVID-19 virus [10, 13] causes changes in the respiratory system, with distinct pathological signs even before covid symptoms appear. Therefore, if we manage to capture these signs in the audio signal, it should be possible to distinguish covid-coughs from other types of coughs.

This task would be a specific case of the general acoustic event recognition problem that has been researched extensively [14, 15, 16]. Several studies have been written on this specific problem of COVID-detection as well in the past year [17, 18]. An overview of some of these works [19] with results from nine articles (as well as our review of the literature) shows that most studies so far have used MFCC-features [17, 20], spectrograms, and a few [21] used handcrafted features. When it comes to algorithms, as SVM-models [19] are very efficient for two-class classification problems, five of these nine studies used SVM models. While four of the nine studies [19] used CNN models, and few studies used different models, like decision trees. A significant [19] issue in the research on COVID-19 cough is data scarcity. Another issue [19, 21] is that some research datasets are based on self-reported symptoms and not RT-PCR-tests and CT-scans.

2. Data

The DiCOVA Challenge contains two tracks, Track-1 for COVID-19 detection and Track-2 for COVID-19 analysis. We limit our submission to the first Track. The dataset for Track-1 consists of audio recordings of coughs produced by COVID-19 positive and COVID-19 negative people originated from the Coswara database [22]. The competition provides us 1040 audio files for training, as well as file lists for 5-fold cross-validation. Out of the 1040 recordings, only 75 come from positive COVID-19 participants, meaning that we have to deal with a highly imbalanced dataset. The audio files are in FLAC - Free Lossless Audio Codec format with a sampling rate of 44.1 kHz. Information about gender and nationality accompany the COVID-19 status class of the recordings. For further details, [1] describes the DiCOVA Challenge thoroughly.

http://dx.doi.org/10.21437/Interspeech.2021-2191

3. Methods

In our work, we have taken the simple approach of utilizing a well-known, easily applicable feature set (Compare 2016 [23]), that can be extracted from audio files using the OpenSmile toolkit [24]. Then, we trained classical machine learning models (namely, Random Forests and Support Vector Machines) on the resulting data with five-fold cross-validation, using the data partitioning provided by the organisers of the challenge. We optimized the meta-parameters of these models based on the validation sets of the five folds, selecting the parameter combination that provided the best average AUC ROC score on those validation sets. Then, for each method, we got the final probability predictions by averaging the predictions of the five models trained using the five different folds. Lastly, we combined our two selected models by taking the weighted average of their predictions, as a further layer of ensembling.

3.1. Feature Description

The open-source Speech and Music Interpretation by Large-space Extraction (openSMILE [24]) toolkit is a framework that automatically extracts sound descriptors (low-level features) such as frame energy, voice intensity/loudness, band spectra, loudness approximated from auditory spectra, fundamental frequency, spectral features, psychoacoustic sharpness, spectral harmonicity, and many more. OpenSMILE is a real-time (i.e. able to extract features in real-time) online and offline tool for processing large datasets. Its input can be from different audio formats (including the Free Lossless Audo Codec - flac - format used in the DiCOVA competition), and different operations (e.g. normalization, modification) can be performed in the processing stage. In our work we used the python implementation [25] of this tool to extract the 6373 features of the ComParE 2016 feature set [23] from each audio recording of coughs.

In ComParE (Computational Paranlinguistics challengE), organised yearly since 2009, there is a broad set of features defined. These features are obtained by applying a large set of statistical functions (e.g. quartiles, percentiles, standard deviation, amplitude range of peaks, percentage of non-zero frames) to acoustic Low-Level Descriptors (LLDs). The ComParE feature set contains 6373 static features from LLD contours which include a broad set of descriptors from different fields, such as speech processing, Music Information Retrieval, and general sound analysis. The feature set includes energy, spectral, and voicing related LLDs. Among energy related LLDs, loudness (sum of auditory spectrum) is one of the features which belongs to the prosodic group and is important in our task. The group of spectral LLDs include features such as MFCC (1–14), spectral variance, skewness, kurtosis, and many more. While jitter, and shimmer features, among others, are used in LLDs related to voicing. This baseline feature set has been used in all the Interspeech ComParE challenges. For more details about the feature set, we refer the reader to the journal paper of Weninger et al. [26].

3.2. Feature normalization

Z-score normalization is a method of normalizing data that avoids the outlier issue by controlling the data distribution using mean and standard deviation. It can be described by the following formula,

$$(x - \mu)/\sigma \tag{1}$$

where x is a feature, μ is the mean value, and σ is the standard deviation of the feature.

Z-score normalization serves as a metric for comparing quantitative features of different scales belonging to different distributions. This characteristic makes it a good fit as a normalization technique during feature development. By applying the z-score method on the CompParE 2016 features, the feature space was redefined with balanced representation and the scale of feature values was regularized, thus allowing models to learn quick and robust. It also prevented the feature domination during learning which improve model performance generalization with uniform feature dependencies. We applied z-score normalization on the feature set before using the CompParE 2016 features in Support Vector Machines, but not for Random Forests.

3.3. Classifier Description

For the classification task, we used two different classifiers implemented using the scikit-learn python library [27], namely Random Forests (RFs) and Support Vector Machines (SVMs). For both classifiers, we train all folds with different combinations of meta-parameters, and then select the parameter setting that provides the best AUC ROC score on average on the five validation sets. By using the meta-parameters that provide the best performance on average on the five validation sets, as opposed to using the meta-parameters that performed best on each individual set, we were hoping to avoid overfitting.

Random Forests are based on random decision trees and scikit's implementation combines classifiers by averaging their probabilistic prediction, while traditional random forests let each classifier vote for a single class. We optimize the number of trees, the split criterion, the maximum depth of the tree, the use of bootstraps when building trees, the use of out-of-bag samples, and the weights associated with the classes.

Support Vector Machines are commonly used in classification, regression, and outlier detection tasks. SVMs will classify data points by finding a hyperplane in n-dimensional space (where n is the number of features), that separates the classes. For SVMs, we optimize the regularization parameter C, the kernel type, the degree of the polynomial kernel function when polynomial kernel is chosen, the kernel coefficient gamma, and the multipliers/weights of parameter C for each class. The optimized meta-parameters for both the random forest, and the SVM model are presented in Table 1.

Table 1: *Meta-parameters resulting from the parameter search for the different methods*

Method	Parameters
RandomForest	bootstrap: False class_weight: None criterion: entropy max_depth: 24 n_estimators: 257 oob_score: False
SVM	C: 2.6499182736174296 class_weight: None degree: 2 gamma: scale kernel: rbf

Table 2: *ROC AUC scores attained using the various machine learning models on the validation and test set*

Method	ROC AUC score	
	Validation	Test
RandomForest	71.73	82.15
SVM	73.29	85.05
Baseline	68.54	69.85

4. Results and Discussion

Results of our experiments are listed in Table 2. As can be seen in Table 2, both the SVM and the Random Forest produced markedly higher scores than the baseline provided by the task organizers [1]. Moreover, we can say that our methods seem to show a good generalization ability as well as attain competitive scores on the test set. One possible explanation (as we will later see - particularly in the case of Random Forests) can be that the different models trained on different folds complement each other, thus their ensemble would perform better than the individual models.

4.1. Ensembling

Lastly, we combined our RandomForest and SVM models by taking the weighted average of their predicted probabilities for the COVID-19 class. Here, to avoid the overfitting of the weight parameters on the validation set , we examined only five weighting schemes. Results of these experiments on the validation set are shown in Figure 1. Based on these results, our final combination took the predictions of the RandomForest into account with a weight of 0.25, (and the predictions of the SVM model with a weight of 0.75). Using this combination, we managed to marginally improve our results on the test set from 85.05 to 85.21.

5. Analysis

Using our models, we have attained a competitive performance on the DiCOVA test set, ranking 4th on the leaderboard (outperformed only marginally by the second and third runner-ups). One important aspect of working with the task, however (beyond producing various test scores) is analysing and understanding the results. In the absence of available test labels, we do so by analysing the trained models.

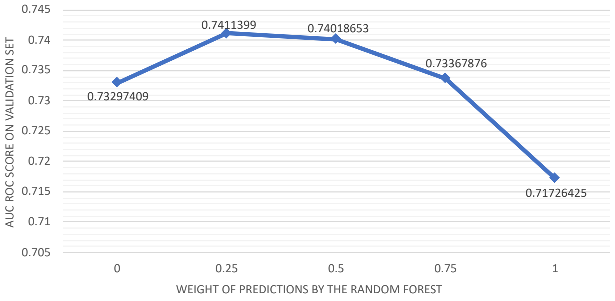

Figure 1: *Weights for the predictions originating from the Random Forest (sums to one with the weight for the predictions originating from the SVM model) plotted against the average AUC ROC score on the validation set*

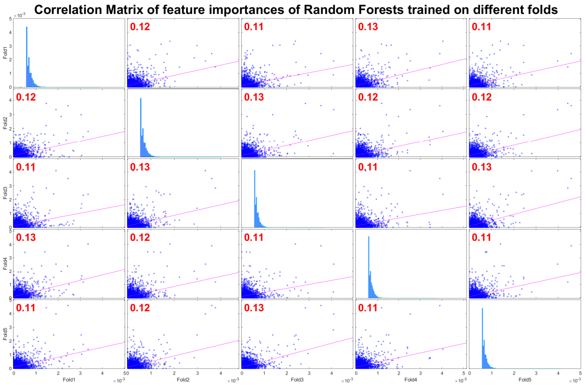

Figure 2: *Pair-wise correlation of feature importance scores of the five Random Forest models trained on the five different folds, calculated using Kendall's tau*

One opportunity to better understand the Random Forest models of scikit-learn is obtaining the feature importance scores (using the TreeInterpreter tool [28]) each model assigns to the different features. Given that all folds represent the same data, one can reasonably expect that on all folds, the same (or similar) features would be important. This is, however, not what we observe when examining the feature importance scores of models trained on different folds. While each model assigns an importance of higher than 0.0 to approximately four thousand features (4336, 4372, 4430, 4267, and 4324 respectively), only 1271 features have an importance for all of them. In other words, only 1271 features got a higher than 0.0 importance score assigned by all models. To gain further insight into this phenomenon, we calculated the pair-wise correlation between feature importance scores produced by different random forests using Kendall's tau [29], and represent them in Fig. 2. As can be seen in Fig. 2, while all correlations are significantly different from zero, they are all weak correlations. One question here is whether this is due to the features in the different folds not adequately representing the same task, or a consequence of our model selection. To examine this, we trained five different Random Forest models on fold 1, using the same meta-parameters. These models all attained a similar AUC ROC score on the validation set (between 75.19 and 75.65), but when looking at the correlation of their feature importance scores (see Fig. 3), we find once again weak to moderate correlations [30].

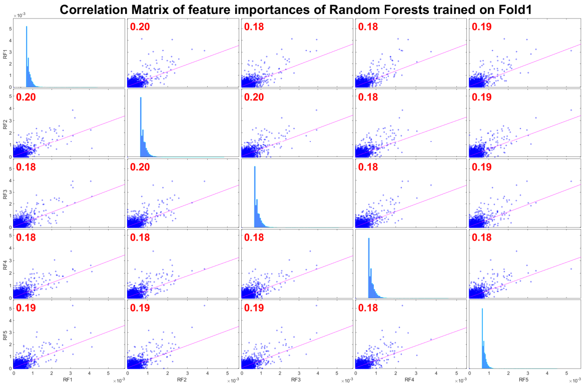

Figure 3: *Pair-wise correlation of feature importance scores of the five Random Forest models trained on the first fold, calculated using Kendall's tau*

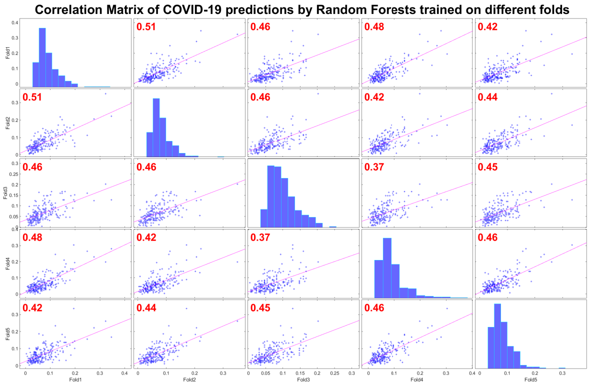

Figure 4: *Pair-wise correlation of test predictions of the five Random Forest models trained on the five different folds, calculated using Kendall's tau*

Moreover, we have also examined the correlations between the predictions of the different models trained on different folds (see Fig. 4), and the same fold (see Fig. 5). Here, we found that the various models arrived to similar predictions using different (but overlapping) feature sets, and different feature importance scores. This, along with our earlier observation suggest that the Random Forest method can converge towards high prediction performance with consistent prediction correlation, although it randomly initiates feature selection. It further shows the ability of the Random Forest method to generalize well on unseen data from the same distribution, as well as the possibility for the method to generalize well on unseen data from different sources and distribution.

To analyse the Support Vector Machine models, we attempt to visualize the feature-space, with the data points of the training set, including the support vectors. During the analysis of data points on the training set, and SVM model performance, we use t-Stochastic Neighbor Embedding (t-SNE) [31] to visualize the input data space of different folds, as shown in Figs. 6 and 7. Here, we also project the support vectors for each fold of the training set. As can be seen in the figures, no dot appears in the plots that represents a COVID-19 class, and is not a support vector, meaning that all COVID-19 data examples become part of the support vector space for each fold (note, that to adhere to page limits here, only visualizations of the first and last fold are presented, but the same property holds for all other folds too). This signifies that SVM model attempts to learn around every COVID-19 example; thus, COVID-19 class is completely covered by support vectors. This strongly suggest that the learning of the SVM model skewed towards the COVID-19 class, and is over-fitted on the available data.

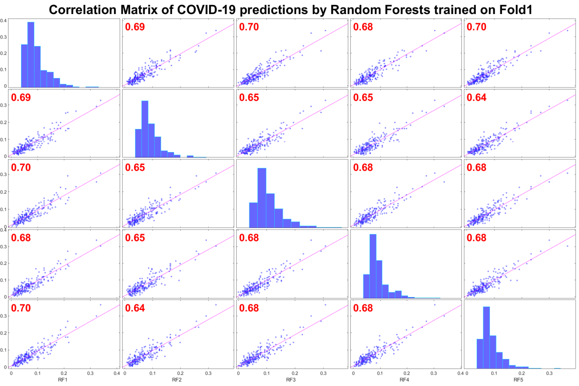

Figure 5: *Pair-wise correlation of test predictions of the five Random Forest models trained on the first fold, calculated using Kendall's tau*

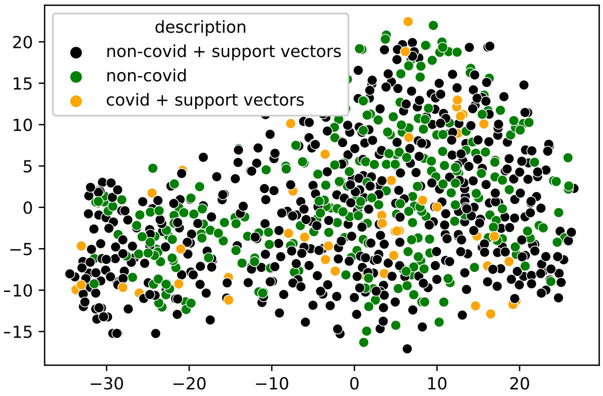

Figure 6: *t-SNE visualization on training set Fold 1*

A comparison of the Random Forest and SVM methods unveils that although both methods performed well on the given validation and test sets, the possibility of performance generalization for the Random Forests is higher on unseen data from a new data collection than for the SVMs.

6. Conclusions and Future Work

In this paper, we have examined the combination of the ComParE 2016 features and two classical machine learning methods for the task of COVID-19 detection on the DiCOVA dataset. We have found that both methods lead to robust models that attain a competitive performance on the test set. Moreover, these methods do not have large computational, or time complexities.

For future work, there are many interesting directions, including the use of CNNs, coupled with additional data, or data augmentation techniques. Moreover, as the data is highly imbalanced, one can experiment with various sampling methods. Lastly, as additional information (e.g. gender, origin of the subject) is available for the coughs, it could also be beneficial to examine adversarial multi-task training [32] techniques.

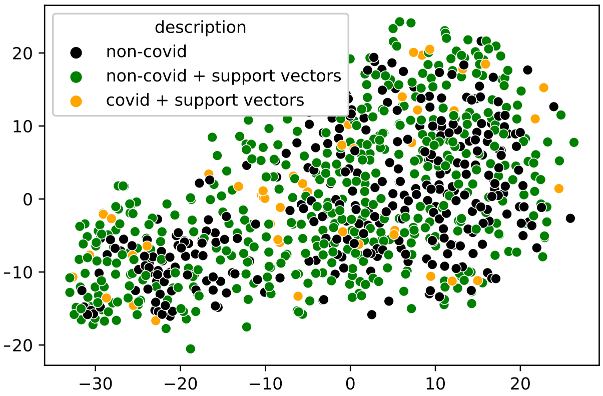

Figure 7: *t-SNE visualization on training set Fold 5*

7. References

[1] A. Muguli, L. Pinto, N. R., N. Sharma, P. Krishnan, P. K. Ghosh, R. Kumar, S. Ramoji, S. Bhat, S. R. Chetupalli, S. Ganapathy, and V. Nanda, "DiCOVA challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics," 2021.

[2] T. Grósz, R. Busa-Fekete, G. Gosztolya, and L. Tóth, "Assessing the degree of nativeness and Parkinson's condition using gaussian processes and Deep Rectifier Neural Networks," in *INTERSPEECH*, 2015, pp. 919–923.

[3] G. Gosztolya, R. Busa-Fekete, T. Grósz, and L. Tóth, "DNN-based feature extraction and classifier combination for child-directed speech, cold and snoring identification," in *INTERSPEECH*, 2017, pp. 3522–3526.

[4] G. Gosztolya, T. Grósz, and L. Tóth, "General utterance-level feature extraction for classifying crying sounds, atypical & self-assessed affect and heart beats," in *INTERSPEECH*, 2018, pp. 531–535.

[5] G. Gosztolya, "Using fisher vector and bag-of-audio-words representations to identify styrian dialects, sleepiness, baby & orca sounds," in *INTERSPEECH*, 2019, pp. 2413–2417.

[6] M. Markitantov, D. Dresvyanskiy, D. Mamontov, H. Kaya, W. Minker, and A. Karpov, "Ensembling End-to-End Deep Models for Computational Paralinguistics Tasks: ComParE 2020 Mask and Breathing Sub-Challenges," in *INTERSPEECH*, 2020, pp. 2072–2076.

[7] D. A. J. Tyrrell and M. L. Bynoe, "Cultivation of a novel type of common-cold virus in organ cultures," *BMJ*, vol. 1, no. 5448, pp. 1467–1470, 1965. [Online]. Available: https://www.bmj.com/content/1/5448/1467

[8] M. Grant, L. Geoghegan, M. Arbyn, M. Yousfi, L. McGuinness, E. Clarke, and R. Wade, "The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (sars-cov-2; covid-19): A systematic review and meta-analysis of 148 studies from 9 countries," *PLoS ONE*, vol. 15, no. 6, Jun. 2020.

[9] R. Viner, J. Ward, L. Hudson, M. Ashe, S. Patel, D. Hargreaves, and E. Whittaker, "Systematic review of reviews of symptoms and signs of covid-19 in children and adolescents," *Archives of Disease in Childhood*, Jan. 2020.

[10] A. Imran, I. Posokhova, H. Qureshi, U. Masood, M. Riaz, K. Ali, C. John, 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, 06 2020.

[11] D. S. Vijayakumar and M. Sneha, "Low cost COVID-19 preliminary diagnosis utilizing cough samples and keenly intellective deep learning approaches," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 549–557, 2021.

[12] K. Qian, B. W. Schuller, and Y. Yamamoto, "Recent advances in computer audition for diagnosing COVID-19: An overview," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, 2021, pp. 181–182.

[13] J. Laguarta, F. Hueto Puig, and B. Subirana, "COVID-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 09 2020.

[14] J. Schröder, J. Anemüller, and S. Goetze, "Performance comparison of GMM, HMM and DNN based approaches for acoustic event detection within task 3 of the DCASE 2016 challenge," in *Proc. Workshop Detect. Classification Acoust. Scenes Events*, 2016, pp. 80–84.

[15] S. Mun, S. Shon, W. Kim, and H. Ko, "Deep neural network bottleneck features for acoustic event recognition." in *Interspeech*, 2016, pp. 2954–2957.

[16] P. Dighe, G. Luyet, A. Asaei, and H. Bourlard, "Exploiting low-dimensional structures to enhance DNN based acoustic modeling in speech recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5690–5694.

[17] V. Bansal, G. Pahwa, and N. Kannan, "Cough classification for COVID-19 based on audio MFCC features using convolutional neural networks," in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, 2020, pp. 604–608.

[18] H. Coppock, A. Gaskell, P. Tzirakis, A. Baird, L. Jones, and B. W. Schuller, "End-2-End COVID-19 detection from breath & cough audio," 2021. [Online]. Available: https://arxiv.org/abs/2102.08359

[19] N. Mashouri, "Identifying COVID-19 by using spectral analysis of cough recordings: A distinctive classification," https://www.preprints.org/manuscript/202101.0239/v1, 2021.

[20] R. Dunne, T. Morris, and S. Harper, "High accuracy classification of covid-19 coughs using mel-frequency cepstral coefficients and a convolutional neural network with a use case for smart home devices," Aug. 2020, preprint at Research Square https://www.researchsquare.com/article/rs-63796/v1.

[21] 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. ACM SIGKDD KDD*, 2020, p. 3474–3484.

[22] 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 2020*, 2020, pp. 4811–4815. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2020-2768

[23] B. W. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. C. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The INTERSPEECH 2016 computational paralinguistics challenge: Deception, sincerity & native language," in *INTERSPEECH*, N. Morgan, Ed., 2016, pp. 2001–2005.

[24] F. Eyben, M. Wöllmer, and B. W. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor." in *ACM Multimedia*, A. D. Bimbo, S.-F. Chang, and A. W. M. Smeulders, Eds. ACM, 2010, pp. 1459–1462. [Online]. Available: http://dblp.uni-trier.de/db/conf/mm/mm2010.html#EybenWS10

[25] J. Wagner, C. Hausner, and H. Wierstorf, "Python wrapper for common opensmile feature sets," https://pypi.org/project/opensmile/, 2021.

[26] F. Weninger, F. Eyben, B. Schuller, M. Mortillaro, and K. Scherer, "On the acoustics of emotion in audio: What speech, music, and sound have in common," *Frontiers in Psychology*, vol. 4, p. 292, 2013. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00292

[27] 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.

[28] A. Saabas, "TreeInterpreter," 2021. [Online]. Available: https://github.com/andosa/treeinterpreter

[29] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938. [Online]. Available: http://www.jstor.org/stable/2332226

[30] R. Botsch, "Chapter 12. significance and measures of association," http://polisci.usca.edu/apls301/Text/, accessed on: 2021-04-02.

[31] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[32] L. Tóth and G. Gosztolya, "Reducing the inter-speaker variance of CNN acoustic models using unsupervised adversarial multi-task training," in *Proc. SPECOM*, 2019, pp. 481–490.