COVID-19 VOICE DIAGNOSIS

Sireesh reddy[1], Abhinav Stephen[2] ,Pranav Steyn[3],Devanjali Relan[4]
BML Mun jal University, Gurgaon- Haryana, India
E-mail: koduru.reddy.19cse@bmu.edu.in, mandala.stephen.19cse@bmu.edu.in ,
mandala.steyn.19cse@bmu.edu.in, devanjali.relan@bmu.edu.in

Abstract – **Healthcare professionals have regularly used audio signals generated by the human body (e.g., sighs, breathing, heart, digestion, vibration sounds) as indicators to diagnose disease or assess disease progression. Until recently, such signals were usually collected manually during scheduled visits. Digital technology is now being used in research to collect bodily sounds (e.g., from digital stethoscopes) for cardiovascular or respiratory examination, which can then be used for automatic analysis. Some preliminary research shows promise in detecting COVID-19 diagnostic signals in voice and coughs. We describe our data analysis over a large-scale crowdsourced dataset of respiratory sounds collected to aid in the diagnosis of COVID-19 in this paper. Coughs and breathing are used to determine how distinguishable COVID-19 sounds are from those in asthma or healthy controls. Our findings show that even a simple binary machine learning classifier can correctly classify healthy and COVID-19 sounds.**

Keywords – **COVID-19, Crowdsourcing Platform, AudioAnalysis, Coughing, Breathing.**

I. INTRODUCTION

Clinicians and clinical researchers have frequently used audio signals generated by the human body (e.g., sighs, breathing, heart, digestion, vibration sounds) in disease diagnosis and monitoring. Until recently, however, such signals were typically collected through manual auscultation during scheduled visits. Digital technology is now being used in research to collect bodily sounds (e.g., digital stethoscopes)and run automatic analysis on the data , for example, for wheeze detection in asthma . Researchers have also been testing the use of human voice to aid in the early diagnosis of a variety of illnesses: Parkinson's disease is associated with softness of speech (due to a lack of coordination of the vocal muscles), voice frequency with coronary arterydisease (hardening of the arteries, which may affect voice production) , and vocal tone, pitch, rhythm, rate, and volume with invisible illnesses such as post-traumatic stress disorder . The use of human-generated audio as a biomarker for various illnesses holds enormous promise for early diagnosis as well as affordable solutions that could be made available to the general public if embedded in commodity devices. This is especially true if such solutions can monitor people in their daily lives in an authentic manner. Recent research has begun to investigate how respiratory sounds (e.g., coughs, breathing, and voice) collected by devices from COVID-19 positive hospital patients differ from sounds from healthy people. In

symptoms

, digital stethoscope data from lung auscultation is used as a diagnostic signal for COVID-19; in, a study of COVID- 19 cough detection using phone data is presented using a cohort of 48 COVID-19 patients versus other pathological coughs on which an ensemble of models is trained. Speech recordings from hospital patients with COVID-19 are analyzed in to automatically categorize patients' health states. In our work, we investigate the use of human respiratory sounds as diagnostic markers for COVID-19 in crowdsourced, uncontrolled data. This paper,in particular, describes our preliminary findings from a subset of our dataset, which is currently being crowdsourced globally at www.covid-19- sounds.org. The data was gathered using an app (Android and Web) that asked volunteers to provide samples of their voice, coughs, and breathing, as well as their medical history and

The app also inquires as to whether the user has tested positive for COVID-19. To date, we have collected approximately 10,000 samples from approximately 7000 unique users. This is, to our knowledge, the largest uncontrolled, crowdsourced data collection of COVID-19 related sounds worldwide We present preliminary findings for COVID-19 on the discriminatory power of coughs and breath sounds. We develop three binary activities: one to differentiate COVID-19 positive users from healthy users; one to differentiate COVID-19 positive users with a cough from healthy users with a cough; and one to differentiate COVID-19 positive users with a cough from users with asthma who reporthaving a cough. The results show that performance for all activities remains above 80 % Area Under Curve (AUC). We are able to correctly classify healthy and COVID-19 sounds with an AUC of 80%. (Activity 1). When attempting to distinguish a user who tested positive for COVID-19 and has a cough from a healthy user with a cough ((Activity 2), our classifier achieves an AUC of 82%, whereas when attempting to distinguish users who tested positive for COVID-19 and have a cough from users with asthma and a cough ((Activity 3), we achieve an AUC of 80%.

- We demonstrate how we can use audio data augmentation to improve the recall performance of some of our activities with less data. We see a 5% and 8% improvement in performance for (Activities 2 and 3, respectively.
- Discussion of results and their potential, as well as illustration of several future directions for our analysis and sound-based diagnostics in

the context of COVID-19, which could lead to COVID-19 pre-screening and progressiondetection.

II. 1 MOTIVATION AND RELATED WORK

Over 521 million cases of the novel coronavirus disease have been reported since its outbreak, with 62 million deaths. Researchers and scientists have made significant progress in developing COVID-19 treatments and vaccines, and effective and easily accessible tests have been critical in quickly tracing infected people. The reverse transcription polymerase chain reaction (RT-PCR) assay to detect the presence of viral ribonucleic acid (RNA) from swab samples is currently the most commonly used and first-line diagnostic tool for COVID-19. RT-PCR tests are highly sensitive in the laboratory (over 95 % diagnostic sensitivity and specificity),but have been found to perform differently in commercial kits, with sensitivity ranging from 75 to 100 percent, and in the worst case reaching as low as 38 %. Furthermore, the sample analysis procedure is involved. time consuming, and only available in approved laboratories highly trained personnel, resulting in limited testing capacity and failure to keep up with the rapid increase in demand computer Tomography (CT). Scanners are becoming more popular for COVID-19 diagnostics in some areas like China.

However, This method, had not been tried because there are still a lot of doctors around is skeptical of the reported high sensitivity.

In addition, CT scanners are specialized and expensive equipment suitable only for medical centers with trained staff for its operation. For inpatients, on top of a high price tag for a single scan, patient transport to and from the scanner requires to break the isolation, which significantly increases the infection transmission risk. It is critical that the pandemic response overcomes the limitations of RT-PCR and CT to test on a large scale in a timely manner. This requires quick, low-cost, long-term, and effective testing methods that can be repeat over time to track progress. This would not only help to contain the current spread, but it would also suppress resurgence and reduce health risks.

AUTHOR	DATASET	TASKS	ACCURACY
Quian I. et al.,	COVID-19 Crowd-sourced Sounds Dataset	1. Sleep 2. Fatigue 3. Anxiety	55 42 49
Lara O. et al.,	COUGHVID Dataset	1.Wheezing 2.Audible Dyspnea 3.Stridor Sound 4.Nasal-Congestion 5. Choking 6.Labeled as COVID-19 mild	90 93 98 99 99 86
Ali Imran et al.,	Data collected through mobile app with name of COVID-19 Samples	Speech Cough Overall	92 92 88
Bader M. et al.,	Own data set collected from hospital with 14 patients	COVID-19 Negative Vs Positive COVID-19 COVID-19 with Cough Vs COVID019 without Cough	Cough 42 Breath 43 Voice 79 Cough 65 Breath 58
Jiang Z. et al.,	Data Collected from Ruijin Hospitals	COVID-19 Detection from Thermal videos and breathing Patterns.	83
Al Ismail M. et al.,	Data set collected from 521 individuals	Linear Regression for COVID detection	82
Shui-Hua Wang et al.,	CT Scan Image Dataset	COVID-19 Detection	97

Machine learning methods have been developed to recognize and diagnose respiratory diseases based on sounds, particularly coughs utilizes convolutional neural networks (CNN's) to detect cough in ambient audio and diagnose three potential illnesses (bronchitis, bronchiolitis, and pertussis) based on their distinct audio features. Several models for COVID-19 prediction using audio have been developed and published in the last year in this context. Machine learning advances have demonstrated the potential of automated auscultation of respiratory sounds and opened up new opportunities for fully automated COVID-19 screening.

III. METHODOLOGY

Feature-based machine learning and shallow classifiers were used due to the moderate size of the dataset chosen and the relevance of explainability given the public health implications of our research. we adopted standard data processing and modeling practices from the audio and sound processing literature for medical purposes. We detail the extracted features and the approach we used to train robust classification models in this part, taking into consideration theunique characteristics of our data (e.g., longitudinal mobile users and cross-validation). We looked at characteristics that were created and those that were learned through transfer learning. We put Logistic Regression (LR), Gradient Boosting Trees (GBTs), and Support Vector Machines (SVMs) to thetest; the results may be found in the results section. An SVM classifier using a Radial Basis Function (RBF) kernel was tested.

A. Dataset used for this analysis

We have obtained the dataset from the University of Cambridge using one-to-one legal agreements for research purposes, due to the sensitive nature (e.g. voice) of thedataset.The title of the paper should be as succinct as possible, stating the subject of the paper in a very clear manner. It should be centered at the top of the first page, in bold, type size 14 points, with the whole title in capital letters.

This dataset focused on a curated collection of the acquireddata for this research, guided mostly by the imbalance of COVID-19 tested participants in the dataset (until 22 May 2020). They also limited our work to coughs and breathing solely (and not the voice samples). One audio recording is represented by a sample. After filtering (silent and noisy samples), we present the number of samples used in our study. They extracted and carefully verified all samples from individuals who indicated they had tested positive for COVID- 19 (during the last 14 days or before), totaling 141 cough and breathing samples. There were 54 samples from users who had a dry or wet cough.

Three sets of users are used as a control

group in our study. Users from countries where the virus was not prevalent at the time of data collection (up to around 2000 cases) are classified as non-COVID users. Albania, Bulgaria, Cyprus, Greece, Jordan, Lebanon, Sri Lanka, Tunisia, and Vietnam were among the countries we chose. Non-COVID users are defined as people who have a clean medical history, have never smoked, have not tested positive for COVID19, and have not reported any symptoms. 298 samples were contributed by these users. The non-COVID with cough group consisted of individuals who met the same criteria as the non-COVID group but reported cough as a symptom; this group provided 32 samples. Finally, asthma with cough users had asthma, had not tested positive for COVID-19, and coughed; they providedus with 20 samples.

B. Pre-Processing

1) Feature extraction:

Handcrafted Features: The apps' raw sound wave forms are re-sampled to 22kHz, which is a common value for audio operations. Our audio processing library was librosa . At the frame and segment level, numerous handmade characteristics including frequency- based, structural, statistical, and temporal properties are retrieved from the re-sampled audio. A segment is the entiretyof a single audio recording, whereas a frame is a portion (subset) of the entire audio data included in a segment.

- **Duration**: the total duration of the recording after trimming leading and trailing silence.
- **Onset**: the number of pitch onsets (pseudo syllables) is computed from the signals, by identifying peaks from an onset strength envelope, which is obtained by summing each positive first-order difference across each Mel bandoccur at regular temporal intervals. In our context, it is used for its peak detection capabilities.
- **Spectral Centroid**: the mean (centroid) extracted per frame of the magnitude spectrogram.
- **Roll-off Frequency**: the center frequency for a spectrogram bin so that at least 85% of the energy of the spectrum in this frame is contained in this bin and the bins below.
- **MFCC**: Mel-Frequency from the short-term power spectrum, based on a linear cosine transform of the log power spectrum on a nonlinear Mel scale. MFCCs are amongst the most common features in audio processing.We use the first 13 components.
- **Period**: the main frequency of the envelope of the signal. We calculate the FFT on the envelope and identify the frequency with the highest amplitude from the 4th mode upwards (as the envelope has non-zero mean).
- **RMS Energy**: the root-mean-square of the magnitude of a short-time Fourier transform which provides the powerof the signal.
- **Δ-MFCC**: the temporal differential (delta) of the MFCC.

.

We extract many statistical characteristics for the features that generate time series (RMS Energy, Spectral Centroid, Roll-off Frequency, and all versions of MFCCs) to capture the distributions beyond the mean. Mean, median, root- mean-square, maximum, minimum, 1st and 3rd quartile, interquartile range, standard deviation, skewness, and kurtosis are all included in the list. The first four segment- level features, four frame-level features represented by their statistics, and three variations of MFCCs with each component represented by its statistics total 477 handcrafted features ($4 + 4 \times 11 + 3 \times 13 \times 11 = 477$).

C. Algorithm /Model Architecture

Features from Transfer Learning. We use VGGish to extract audio features automatically in addition to handcrafted features. VGGish is a type of convolutional neural network. Audio classification based on raw audio input was proposed. A large-scale YouTube dataset was used to train the VGGish model. The model parameters that were discovered were made available. It's something we use as a feature extractor for converting raw audio wave forms to embeddings (features) that are subsequently used to train a shallow neural network classifier. The VGGish pre-trained model separates data samples into 0.96-sec non-overlapping sub-samples and outputs a 128-dimensional feature vector for every 0.96 seconds. 16 kHz is the sampling rate. The final features are the mean and standard deviation across the entire segment, with dimension 256 (128x2). Because VGGish only acceptsa spectrogram as input, some key temporal properties may be ignored in the feature space, necessitating the employment of a combination of VGGish and handcrafted features. Section 5 demonstrates that this combination improves AUC when compared to utilizing either VGGish or handmade features.

We get a 477-dimensional handcrafted feature vector, a 256-dimensional VGGish-based feature vector, and many composite feature vectors totaling 733 dimensions for each modality (cough, breathing). Each combined feature vector is made up of a subset of the handmade feature sets and VGGish- based features concatenated together. Principal Components Analysis (PCA) further reduces these feature vectors while maintaining a portion of the initially explained variance.

IV. EVALUATION

The classification of audio samples as COVID-19 or healthy using the features given in Section 4 is now detailed. A subset of the originally gathered

dataset (detailed in Section 3.3) was used due to the high-class imbalance. We begin by explaining how data from various modalities was combined and the dataset was partitioned for the experiments. The experiments, as well as the Android and iOS app codebases, are freely available to encourage repeatability. The section concludes with a discussion of the findings and outcomes.

A. Experimental setup

1) Classification activities. : We focus on three clinically important binary classification activities based on the data collected (Section III-A)

Activity 1: Separate users who have declared they tested positive for COVID-19 (COVID-positive) from users who have not declared a positive test for COVID-19, have a clean medical history, have never smoked, have no symptoms, and were in countries where COVID-19 was not prevalent at the time, as described in Section 3. (non-COVID). While we cannot guarantee that they were not contaminated, the chances are slim.

Activity 2: Separate users who have declared they tested positive for COVID-19 and have listed cough as a symptom (a common symptom in persons with COVID, as shown in Figure 3), (COVID-positive cough), from users who have claimed that they have not tested positive for COVID-19, that they have a clean medical background, that they have never smoked, that they lived in places where COVID-19 was not prevalent at the time, and that they have a cough as a symptom (non-COVID with cough).

Activity 3: Separate users who have declared they tested positive for COVID-19 and have cough as a symptom (COVID-positive with cough) from users who have not declared they tested positive for COVID-19, are from countries where COVID-19 was not prevalent at the time, have asthma in their medical history, and have cough as a symptom (COVID-positive with cough) (non-COVID with cough).

2) Data exploration: We evaluate the differences between the distributions of the features derived from cough and breathing subdivided by respective class as a first step following feature extraction. We cannot provide all distributions due to the large dimensionality of the features, thus we focus only on the mean statistical characteristic of each feature family (e.g., Centroid is Centroid mean here). Coughs and breaths from COVID-positive users had longer total durations, more onsets, higher periods, and lower RMS, but their MFCC features [first component and deltas] have fewer outliers, as shown in Figure 5. COVID-positive users' samples concentrate more towards the mean of the distributions in both activities, whereas the general (healthy) population displays a wider range (interquartile range), The theory is that a (perhaps forced) healthy cough and breathing are extremely varied. This could also imply that coughs and breaths are effective sounds for identifying COVID and non-COVID users.

3) Feature ablation studies: We repeat our studies with three different audio inputs: only cough, only breathing, and mixed to see which audio modality (cough or breathing) contributes more to categorization performance. We run tests to discover the appropriate cut-off value for PCA to account for the rising dimensionality of the composite representation and to make a fair comparison (see results in next section). The percentages of explained variance are [70%, 80%, 90%, and 95 %]. This means that the classifiers will require fewer Input dimensions if the explained variance is lower, and vice versa. To avoid over fitting, a mixed representation may require a more compressed representation than a representation based solely on coughs or breaths.

4) Cross-validation with users: We make training and test sets from disjoint user divides, ensuring that no samples from the same user appear in both splits. It is important to note that approach does not produce precisely balanced class splits; however, when necessary, we down sampled the majority(non-COVID) class. The balance of the test set is maintained.

Even so, it's difficult to ensure that a split selects a representative test-set, therefore we used a 10-fold-like cross validation in the outer loop (80% /20% split) and a hyper-parameter search in the inner loop to discover the best parameters (using the 80% trainset in a 5-fold cross validation). This configuration is similar to nested cross-validation. We test 5400 models (3 Activities × 3 modalities × 10 user splits × 4 dimensionality reduction cut-offs × 3 feature types × 5 hyper-parameter cross- validation runs). We used the Receiver Operating Characteristic - Area Under Curve (ROC-AUC), Precision, and Recall as standard evaluation criteria. The standard deviation and average performance of the outer folds (10 user-splits) are reported. The performance of our three activities is reported in the next section.

5) Sensitivity to demographics: Including age and sex as one-hot-encoded features in our models (for example, age group: 40-49 years old) had no significant effect on the results($< \pm$ 2 AUC).

B. Distinguishing COVID-19 users from healthy users

Table 1 shows the classification results for the three activities mentioned earlier. We present the best results for each activity, which may have been acquired using a single modality (cough or breathing noises) or a combination of both.The first row shows the classification results for Activity 1: the binary

Task	Modality	Mean ± std			
		PCA	ROC-AUC	Precision	Accuracy
1. COVID-positive / non-COVID	Cough+Breath	**0.95**	**0.80(0.07)**	0.72(0.06)	0.7(0.06)
2. COVID-positive with **cough** / non-COVID with **cough**	Cough	**0.9**	**0.82(0.18)**	0.80(0.16)	0.77(0.16)
3. COVID-positive with **cough** / non-COVID **asthma cough**	Breath	**0.7**	**0.80(0.14)**	0.69(0.20)	0.70(0.21)

classification activity of distinguishing users who state that they have tested positive for COVID-19 (COVID- positive) from those who did not (non-COVID). According to the metrics, there appear to be some discriminatory signals in the data, implying that user coughs paired with breathing could be a good predictor when screening for COVID-19. The AUC for this job is at 80%, while precision and recall are around 70%. Activity 1 has the lowest standard deviations across the user-splits when compared to the other activities (Activity 2 and 3), owing to the higher data set. We used a simple classifier (Logistic Regression), and the data was perhaps too small to eliminate the noise and variability brought by our crowd sourced data collection (e.g., differences in microphones, surrounding noises, ways of inputting the sounds). Nonetheless, these findings provide us with confidence in the signal's strength. We also discovered that when handcrafted features are combined with features learned via VGGish, the results are better than when handcrafted or transfer learning features are used alone, demonstrating the value of applying transfer learning in our research.

C. Distinguishing COVID-19 coughs from other coughs

The binary classification of users who reported testing positive for COVID-19 and also declared a cough, as wellas a similar number of users who indicated they did not test positive for COVID-19 but declared a cough, is described in the second row of Table 1. (Activity 2). An AUC of 82 percentis the best result. This job has an accuracy of 80%, indicating that cough noises can distinguish COVID-19 positive users rather well. This model has a little lower recall (72 percent), indicating that it casts a good but somewhat specialized net:

it does not catch every COVID-19 cough, but it does detect a lot of them. Nonetheless, given the amount of the data and the relatively high standard deviations compared to Activity 1, renders this result preliminary. Users who stated they did not test positive for COVID-19 but had asthma and declared a cough were compared to users who said they did not test positive for COVID-19 but had asthma and declared a cough, as indicated above. Table 1's last row indicates an AUC of 80%. While recall is acceptable, precision is likewise high forthis assignment, as it is for the other two. Breathing noises appear to be more potent signals for discriminating users in this activity, which is intriguing. We investigated the use of data augmentation to increase performance on Activities 2 and3.

V. DISCUSSION AND CONCLUSIONS

We discussed an ongoing initiative to crowd-source respiratory sounds and investigate how such data could help diagnose COVID-19. These findings merely scrape the surface of this type of data's potential; while they are encouraging, they are not strong enough to be used as a single screening tool. To deal with the issue that the fraction of COVID-19 positive users is modest, we've confined ourselves to using a subset of the data acquired for the time being. We alsohad no ground truth on health status, thus we assumed users from countries where COVID-19 was not widespread at the time were healthy when self-reporting as such (however, this limited our dataset further).

We mentioned a current project to crowd source respiratory sounds and look into how this information could help identify COVID-19. These findings only scratch the surface of the potential of this type of data; while encouraging, they are insufficient to be employed as a single screening tool. To deal with the issue of a small percentage of COVID-19 positive users, we've limited ourselves to using a subset of the data for the time being. We also had no ground truthon health status, therefore we assumed users from countries where COVID-19 was not widely used at the time werehealthy when self-reporting (however, this limited our dataset further). While we only looked at the difference between cough sounds in COVID-19 and asthma, our dataset includes people with different respiratory diseases, and we plan to look into this further to see how distinguishable COVID-19 is in this regard. Because the mobile app encourages users to produce samples every couple of days, we have many users for whom we can examine the evolution of respiratorysounds over time. This is extremely important for COVID-19, and it's something we haven't investigated yet in our

present research.

Finally, our mobile app just gathers data and does not provide medical advice; while we believe that the models created from this data may be beneficial in illness screening, we are aware of the difficulties involved in providing medicaladvice to users and the arguments that this generally causes.

ACKNOWLEDGEMENTS

We thank Dr.Devanjali Relan for encouraging to choose this topic right from the beginning. Her teachings and advises were very helpful to us to complete our project.

REFERENCES

1 2019. librosa.feature.delta https://librosa.github.io/librosa/generated/librosa.feature.delta.html, note= Accessed: 2020-05-30.

2 2020. coughvid. https://coughvid.epfl.ch/about/, note= Accessed: 2020-05-30

3 2020. Detect Now https://detectnow.org/, note= Accessed: 2020-05-30

4 Charles Bales, Muhammad Nabeel, Charles N. John, Usama Masood, Haneya N. Qureshi, Hasan Farooq, Iryna Posokhova, and Ali Imran. 2020. Can machine learning be used to recognize and diagnose coughs? arXiv:2004.01495 [eess.AS] 10 pages.

5 Debrup Banerjee, Kazi Islam, Keyi Xue, Gang Mei, Lemin Xiao, Guangfan Zhang, Roger Xu, Cai Lei, Shuiwang Ji, and Jiang Li. 2019. A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. Knowledge and Information Systems 60, 3 (2019), 1693–1724.

6 L Brabenec, J Mekyska, Z Galaz, and Irena Rektorova. 2017. Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation. Journal of Neural Transmission 124, 3 (2017), 303–334.

7 Gavin C Cawley and Nicola LC Talbot. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research 11 (2010), 2079–2107.

8 Yohan Chon, Nicholas D. Lane, Fan Li, Hojung Cha, and Feng Zhao. 2012. Auto- matically characterizing places with opportunistic crowdsensing using smart- phones. In Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp). Pittsburgh, Pennsylvania, 481–490.

9 Steven Davis and Paul Mermelstein. 1980. Comparison of parametric representa- tions for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing 28, 4 (1980), 357–366.

10 Gauri Deshpande and Björn Schuller. 2020. An overview on audio, signal, speech, language processing for COVID-19. arXiv preprint arXiv:2005.08579 (2020).