INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

MMLung: Moving Closer to Practical Lung Health Estimation using Smartphones

Mohammed Mosuily, Lindsay Welch, Jagmohan Chauhan

University of Southampton, United Kingdom

mtm1g19@soton.ac.uk, lindsay.welch@wessexahsn.net, j.chauhan@soton.ac.uk

Abstract

Long-term respiratory illnesses like Chronic Obstructive Pulmonary Disease (COPD) and Asthma are commonly diagnosed with the gold standard spirometry, which is a lung health test that requires specialized equipment and trained healthcare experts, making it expensive and difficult to scale. Moreover, blowing into a spirometer can be quite hard for people suffering from pulmonary illnesses. To solve the aforementioned limitations, we introduce MMLung, an approach that leverages information obtained from multiple audio signals by combining multiple tasks and different modalities performed on the microphone of a smartphone to estimate lung function. Our proposed approach achieves the best mean absolute percentage error (MAPE) of 1.3% on a cohort of 40 participants. Compared to the reported performances (5%–10% MAPE) on lung health estimation using smartphones, MMLung shows that practical lung health estimation is viable by combining multiple tasks utilizing multiple modalities.

Index Terms: Spirometry, Smartphone, Audio, Speech, Cough

1. Introduction

Pulmonary diseases such as COPD and Asthma are considered to be the third leading cause of mortality in the world [1, 2]. According to the report by the Global Initiative for Chronic Obstructive Lung Disease (GOLD), [3] Spirometry test is the gold standard when it comes to measuring lung function using the ratio of Forced Expiratory Volume in 1 second (FEV_1) to the Forced Vital Capacity (FVC) to identifying respiratory diseases such as COPD [4]. However, it requires medical-grade equipment and trained medical professionals to operate the equipment. Hence, such methods are challenging to scale, expensive, and lack accessibility. For example, patients wait a longer time to receive general treatments and diagnostics with the National Health Services (NHS) in England [5]. Additionally, the wait time increased post-COVID-19 as spirometry stopped entirely for around 18 months due to infection control risks [6]. Hence, the need to explore the potential of using ubiquitous technologies to predict lung function estimation is becoming important.

Multiple studies proposed using smartphones for lung health assessment such as, SpiroSmart[7], SpiroCall[8], Mobispiro [9], SpiroConfidence[10], SpiroSonic[11], ExhaleSense [12], and Ubilung[13]. Thus, showing that it is possible to estimate lung function using multiple audio modalities (cough, speech, vowel, spirometry) collected from a smartphone. However, two major limitations came to light in the existing works: (a) the performance obtained is in the range of 5%–10% MAPE which is slightly higher than the recommended MAPE (less than 5%) for clinical use, and (b) they have not compared the performance of different modalities using same methods to evaluate their utility for lung health assessment.

In this work, we aim to improve the performance of lung health estimation so that it becomes practical to be used in real life and in clinical practice. To this end, we collected data from 40 participants on a smartphone and designed a machine-learning pipeline that utilizes and combine audio signals obtained from different modalities sourced through various tasks performed by the participants. The results show that our approach - MultiModal Lung (MMLung) is effective and obtains state-of-the-art results in estimating lung health functionality.

In summary, MMLung makes the following contributions:

- A first-of-its-kind benchmarking study to compare different audio modalities - cough, mobile spirometry, vowels, and speech using a single framework and investigate their performance for lung health estimation.
- We explore how and if multiple modalities and tasks should be combined to achieve the best performance - lower MAPE than the existing works.
- We show that the best-performing task (Long Sentence in one breath) using speech achieved a MAPE of 7.40%. While combining all 14 tasks resulted in a MAPE of 1.13% which comes with a trade-off between the number of tasks and reaching the best result. However, as performing all the tasks at once might not be feasible for a user our results also show that we can obtain a MAPE (<3%) with a combination of three to five small tasks without involving arduous mobile spirometry task.

2. Related Work

There have been multiple studies exploring the potential of various mobile sensors including microphones for lung health assessment. Larson et al. [7] proposed SpiroSmart which utilizes the smartphone's built-in microphone as an accurate spirometer measurement tool. They emphasize that this technology is not to replace spirometry tests used in hospitals. But, can be used as a convenient tool to monitor lung health in a non-clinical environment. SpiroSmart obtained a mean error of 5.1% compared to a clinical spirometer. Another system by Goel et al. [8] introduced SpiroCall, where they target tested lung function over a phone call service. They also proposed using a 3D-printed vortex whistle to test the airflow through the system. SpiroCall achieved a mean error of 6.2% for lung function estimation. Both studies suggest that it is possible to use the smartphone as a spirometry test estimation tool. However, these studies use 3D-printed whistles and rely on telecommunication services. Our approach addresses these limitations by combining different types of audio sounds to achieve state-of-the-art results without needing external hardware.

ExhalseSense [12] also enforces the idea that using smart-

10.21437/Interspeech.2023-721

phones for lung function estimation is achievable. Their audio signal processing and regression model produced a 7.57% mean absolute percentage error for lung obstruction approximation. Saleheen et al. [14] proposed the use of the vowel [a] in a monosyllabic voice to estimate lung functionality. Complementing their approach, we explore if other vowels [e, i, o, u] can be used to estimate respiratory health. We also combine the acoustic features from multiple vowels to reach better outcomes. Another approach by Chun et al. [15] demonstrates the possibility of predicting pulmonary obstruction using natural speech which was recorded by a smartphone. They stated that their study predicted the FEV_1/FVC ratio with a mean absolute percentage error of 8.6%. Moreover, recent research proposed the potential of combining cough and natural speech in estimating lung function which achieves a mean absolute percentage error of 7.2% [13]. However, they also used heart rate in the feature merging step which requires a wearable device that might not always be available to an end user.

Overall, the current literature supports the notion that audio sounds such as cough, vowels, speech, and mobile spirometry can be used to assess lung health on smartphones. However, simply using one type of modality provides limited utility as existing research predicts lung health with a mean absolute percentage error ranging from 5.1% to 8.6%. To be useful in reality we would want the mean absolute percentage error to reach zero [16]. With MMLung we took one step further in achieving this goal by combining audio signals obtained from different modalities sourced from multiple tasks. Furthermore, our proposed system only requires a smartphone which ensures that MMLung is reachable to the masses.

3. Methodology

3.1. Data Collection

The data was collected from 40 participants (20 male, 20 female) with an age range of 18-85 years old; English speaking from the UK. Among them, 12 were healthy participants, while the others consisted of seven self-reported COPD patients, seven self-reported asthma patients, and 14 people with other long-term conditions. Three devices were used to collect the data: Google Pixel 6 Smartphone with an app installed for the data collection, and an Easy on-PC ultrasonic spirometer by ndd Medical Technologies. We obtained ethics approval for our study and we are planning to release a copy of the data set in the future.

The audio data collection from smartphones was conducted in stereo mode at a sampling rate of 44100 Hz. The data was saved in the .wav format. The collection took place in a silent room conditions. The process consisted of collecting data for four audio modalities i.e. cough, vowels, mobile spirometry, and speech via a series of tasks from each participant in a single session. We asked the participants to perform the following tasks: (a) force cough up to 10 times [17]; (b) pronounce the vowels [a, e, i, o, u] in one breath like [aaaa...] in two-three iterations; (c) spirometry task on phone by taking a deep breath and blowing into the microphone until they have expired all their air; (d) read the Rainbow Passage [18]; (e) describe a picture to record spontaneous speech for two minutes; (f) read one short sentence within one breath; (g) read one long sentence within one breath; (h) describe the room that we are in during the recording session; (i) read one text full of action words; (j) read the text that does not contain action words. Note that the relation between the vowel and International Phonetic Alphabet (IPA) notation is as follows: a = [ɑ], e =[ɛ], i=[ɪ], o= [ɔ], u= [ʊ] [19]. Ground truth data were collected using a medical-grade spirometer by a healthcare professional as per European Respiratory Society (ATS/ERS) clinical standards [20]. We also, quality assessed the blows from the spirometer, rejecting blows that did not qualify either within or best of 150mls repeatability. However, with any objective measure that is reliant on individual effort, there may always be unforeseen errors (effort-dependent blows) [21].

3.2. System Design

Figure 1 displays our system design, we will provide a detailed description of each step below. We utilized various libraries and tools in our project[1], including Python, Pandas, NumPy, SciPy, Librosa, pyAudioAnalysis, and Scikit-learn [22].

3.2.1. *Pre-processing*

We started by cleaning the data initially using Audacity to ensure data quality. This includes removing noise, such as participant chatter occurring before and after each recording, to ensure that the recordings accurately represented the intended task.

In addition to the silence at the beginning and end of an audio file. Further work involved picking a single cough event from a series of coughs - one with the highest amplitude and longest duration. The same was done for the vowel sounds. Next, we prepare the audio signal for feature extraction. We start by applying min-max normalization on the signal. Due to varying phone positions and as tasks were recorded in stereo mode one of the channels usually has a higher amplitude and contains more information. We calculate the root mean square (RMS) to select the channel with the higher amplitude [23, 12].

3.2.2. *Feature Engineering Layer*

There are three blocks in the feature engineering layer as shown in Figure 1. Before we start extracting features we divide the audio into multiple frames of 100 ms with no overlap. We extracted multiple features from each frame such as Energy Entropy, Zero Crossing Rate, Energy, Spectral Flux, Spectral Spread, Spectral Entropy, Spectral Centroid, Chroma, Bark Frequency Cepstral Coefficients (BFCCs), Linear-frequency Cepstral coefficients (LFCCs), Linear Prediction Components (LPCs), Mel-Frequency Cepstral Coefficients (MFCCs), Magnitude-based Spectral Root Cepstral Coefficients (MSRCCs), Constant Q-transform Cepstral Coefficients (CQCCs), Gammatone Frequency Cepstral Coefficients (GFCCs), and Rasta Perceptual Linear Prediction Coefficients (RPLPs). Next, we calculate statistical features such as mean, standard deviation, skewness, and kurtosis by combining the features from each frame.

There are six levels for the following: Zero Crossing Rate, Energy, Energy Entropy, Spectral Centroid, Spectral Spread, Spectral Entropy, Spectral Flux, and Spectral Rolloff. Four statistical measures (mean, standard deviation, skewness, and kurtosis) are computed for all six levels amounting to 192 features. Similarly, 52 features are generated for Chroma (13 levels*4 statistical measures). Different coefficient features such as MFCC contribute 104 features (8*13). In addition, there are five features representing the whole audio signal. Finally, we ended up getting 353 features.

Note that each task represents one audio file. To improve

[1] `https://github.com/MohammedMosuily/mmlung`

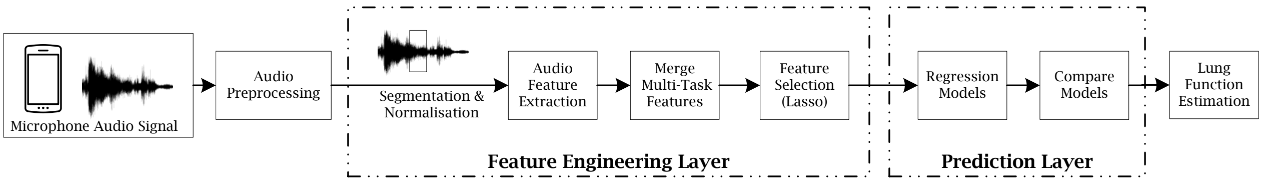

Figure 1: *Flowchart representing Audio signal analysis and processing pipeline. Consenting of the pre-processing stage, feature engineering layers, and prediction layer and ending with the lung function estimation.*

the performance, we explored if merging audio signal information from multiple modalities and tasks can help. To this end, we combined features from different tasks representing a modality (cough, speech, vowel, mobile spirometry). More than 25 different combinations were tested, such as, combining audio data (features) from all 14 tasks or merging data from all the vowels-based tasks to achieve the best performance. Table 3 shows the best 12 combinations with our approach.

After the merging block, we would have a large number of features depending on the chosen combination. To avoid overfitting we tested methods for feature selection such as Lasso Regression. Next, we used the ELI5 library to create an explainable weight that will be used to rank the selected features [24]. An example of the selected features is shown in Table 1 which presents a snapshot of the top 15 from the 35 features and their prediction explainable weights that were used in obtaining a MAPE of 1.13% (shown in bold) in table 3.

3.2.3. Prediction Layer

In the last layer of our pipeline, we evaluated our data with multiple machine learning methods such as Linear Regression, Ridge, Stochastic Gradient Descent, Support Vector, Nu Support Vector, K-nearest Neighbours, Decision Tree, AdaBoost, Gradient Boosting, and XGBoost regression methods. We used Scikit-learn to create the regression models [22]. We used grid search and explored multiple hyperparameters to obtain the best results. We did leave one out cross-validation (LOOCV) [25] per subject to check the generalization of our model and used the mean absolute percentage error of the FEV_1/FVC ratio as a metric to compare the performance of different machine learning methods in estimating lung health function.

Table 1: *A snapshot of the top 15 from the 35 features after merging all the tasks which resulted in MAPE of 1.13%*

#	Feature	Weight
1	Long_MSRCC_10	0.1999
2	Long_Chroma_13_kurtosis	0.1180
3	Non_Action_MFCC_10	0.1149
4	[o]_Single_Level_1_Spectral_Entropy_mean	0.1106
5	[i]_Single_Level_2_Zero_Crossing_Rate_kurtosis	0.1070
6	Long_MFCC_2	0.1034
7	[i]_Single_selected_channel	0.0747
8	Long_Level_5_Spectral_Spread_skew	0.0671
9	[e]_Single_Level_5_Spectral_Flux_mean	0.0644
10	Long_Chroma_8_std	0.0629
11	Describe_Sth_Chroma_13_kurtosis	0.0624
12	Action_LFCC_8	0.0619
13	Cough_MSRCC_13	0.0577
14	[o]_Single_Zero_Crossing_Rate_std	0.0512
15	[u]_Single_Level_3_Spectral_Rolloff_kurtosis	0.0510

4. Results

Existing studies focus on using a single task/modality to do the lung function estimation such as using the vowel [a] [14], or spirometry test by blowing into a smartphone [12] and usually obtaining a MAPE between 5% and 8.5%. To complement existing studies, we started by experimenting with single tasks. The results are shown in Table 2. The best-performing task is reading a long sentence - a MAPE of 7.49% with 28 features on a Linear Regression model. Participants read the following long sentence, "The weather today is sunny with lots of wind and lots of clouds in the sky, which will bring heavy rain and thunderstorms in the afternoon". MAPE across all tasks varied from 7.49 % to 12.35%. The speech modality was followed by mobile spirometry and a vowel in terms of performance while cough is the least significant modality. This result provides the first comparison of different modalities using a baseline approach and further bolsters the claim that speech can be the most effective modality to estimate lung health.

The MAPE equation used in this and previous studies is:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \qquad (1)$$

Here, A_t is the actual value for time period t, F_t is the forecasted value for time period t, and n is the total number of time periods. The MAPE measures the average absolute percentage difference between actual and forecasted values. The closer the MAPE is to zero, the better the forecast.

Our approach to combining signal information from multiple modalities and tasks is shown in Table 3. The specific combination of tasks provides promising results - MAPE between 1.13% and 6.23% on models trained on Linear, Ridge, and Nu Support Vector regression methods while the number of features varied from 35 to 39. The best result is obtained by merging all 14 tasks together - a MAPE of 1.13% with 35 features using Linear Regression. In fact, the first 11 combinations achieved a MAPE of less than 5% which shows the effectiveness of our approach and moves the performance of smartphone-based lung health assessment closer to practice in the real world.

Additionally, we also observe that some combinations of tasks especially #2, #4, and #6 are some of the most promising task combinations for lung health assessment. Primarily, this shows that a small number of tasks when combined can achieve very good performance (<3%) compared to doing all the tasks

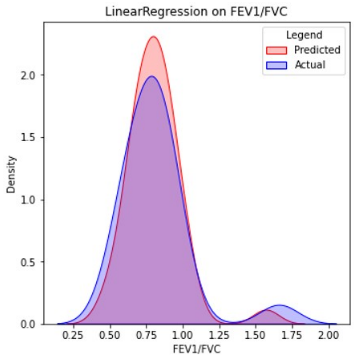

Figure 2: *Predication performance of FEV_1/FVC. Using Linear Regression achieved 1.13% MAPE when Combining all tasks.*

which can be time-demanding and exhausting for a person. For example, performing all the vowel-related tasks #2 will be easier than tasks described in #1. Also, combinations #2, #4, and #6 do not require users to perform mobile spirometry which is highly usable/desirable as people suffering from pulmonary illnesses and the elderly population finds it difficult to perform full-blown spirometry.

Our results also signify the power of combining modalities via tasks. For example, cough alone is not a powerful modality (12.35% MAPE) but a combination of cough with the vowel sound [i], and reading a long sentence provides a MAPE of 1.55%. Figure 2 shows the predicted performance when combining all the tasks and using the Linear Regression model achieving 1.13% MAPE. It also shows how close is the prediction to the actual ground truth values.

5. Discussion and Future Work

MMLung achieved excellent results which can help push the boundaries to make lung function estimation more accurate and ubiquitous. However, there are various avenues which can be explored in the future. Firstly, our user study consists of 40 participants. In the future, we are planning on collecting more data from a diverse population. Secondly, we are planning to focus on exploring the performance of machine learning models for other metrics such as FEV1 and FVC. Thirdly, we will like to collect users' subjective opinions about how they feel about doing different tasks and which type of tasks they would prefer for doing lung health assessment on smartphones in terms of user comfort and usability. Finally, our study is currently focused on the English language and incorporates regional accents from the UK. In the future, we plan to expand the scope of our research by testing our methods in additional languages.

Table 2: *Single-task results show MAPE and mean squared error, the regression model used, number of features and the task. The results are ordered according to MAPE.*

#	Task	feat.	Model	MSE	MAPE
1	**Long Sentence**	**28**	**Linear**	**0.52%**	**7.49%**
2	Spirometry	29	NuSVR	0.74%	8.68%
3	Vowel [i]	29	Ridge	1.74%	9.22%
4	Vowel [o]	31	Ridge	2.22%	9.99%
5	No-Action Txt	22	Ridge	1.52%	10.21%
6	Describe Room	29	Ridge	1.27%	10.76%
7	Vowel [u]	31	SGD	2.71%	11.12%
8	Desc. Picture	26	SGD	1.43%	11.24%
9	Short Sentence	23	Ridge	1.36%	11.52%
10	Vowel [e]	32	SGD	2.41%	11.54%
11	Action Text	27	Linear	1.34%	11.55%
12	Vowel [a]	33	Ridge	1.69%	11.61%
13	Rainbow Txt	28	SGD	1.66%	11.64%
14	Cough	26	NuSVR	1.57%	12.35%

Table 3: *Multi-task combinations results show MAPE and mean squared error, the regression model used, number of features and the task. The tasks are ordered according to MAPE.*

#	Task	feat.	Model	MSE	MAPE
1	**All Tasks**	**35**	**Linear**	**0.01%**	**1.13%**
2	All Vowels	39	NuSVR	0.02%	1.15%
3	Cough,Spiro., Long,[i]	39	Linear	0.02%	1.22%
4	Cough,[i],Long	39	NuSVR	0.03%	1.55%
5	All (No Spiro.)	39	NuSVR	0.03%	1.68%
6	Cough,Long,Short	39	NuSVR	0.10%	2.18%
7	Spiro.,Long,[i]	39	NuSVR	0.05%	2.48%
8	All Speech	39	Ridge	0.07%	3.03%
9	All (No Speech)	39	Ridge	0.14%	3.04%
10	Cough,Spiro.,Long	38	Ridge	0.16%	3.72%
11	Cough, Spiro, [i]	37	Ridge	0.20%	3.87%
12	Long, [i]	38	Ridge	0.34%	5.42%

6. Conclusion

Many studies suggest that smartphone spirometry could not yet replace current clinical spirometry devices. However, this study has demonstrated that personal lung monitoring is a possibility, at this stage as a screening tool, to enable public awareness of changes or detection in the early deterioration of their lung health, an example of this is (KardiaMobile Cardiac Screening) [26]. The best case scenario will be to have continuous monitoring and reporting to the medical care experts, which could be effective for an earlier treatment if needed. In this direction, this paper proposed MMLung an approach to merge audio signal information from multiple modalities via multiple tasks to achieve impressive performance for lung function estimation.

We obtained the optimum result with a MAPE of 1.13% by combining all the tasks. However, as performing all the tasks at once might not be feasible for a user our results also show that we can obtain an MAPE (<3%) with a combination of three to five small tasks without involving mobile spirometry. Our work is also the first one to compare all modalities - cough, speech, vowels, and mobile spirometry using a single benchmark setting and show that speech is the best modality to be deployed for testing lung functionality with smartphones. Overall, MMLung takes one step further to make smartphone-based lung health assessment more accurate and practical.

7. Acknowledgement

This work was supported by Eupnoos and the Institute for Life Sciences (IfLS) HEIF Research Stimulus Fund. Participants are associated with the University of Southampton and Care Ashore (retirement accommodation).

8. References

[1] J. B. Soriano and et al, "Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," *The Lancet Respiratory Medicine*, vol. 5, no. 9, pp. 691–706, 9 2017.

[2] X. Li, X. Cao, M. Guo, M. Xie, and X. Liu, "Trends and risk factors of mortality and disability adjusted life years for chronic respiratory diseases from 1990 to 2017: systematic analysis for the Global Burden of Disease Study 2017," *BMJ*, vol. 368, 2020. [Online]. Available: https://www.bmj.com/content/368/bmj.m237

[3] GOLD, "Global Strategy for Prevention, Diagnosis And Management Of COPD," Global Initiative for Chronic Obstructive Lung Disease, Tech. Rep., 2022. [Online]. Available: https://goldcopd.org/2022-gold-reports-2/

[4] D. M. Halpin, G. J. Criner, A. Papi, D. Singh, A. Anzueto, F. J. Martinez, A. A. Agusti, and C. F. Vogelmeier, "Global Initiative for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease," *American Journal of Respiratory and Critical Care Medicine*, vol. 203, no. 1, pp. 24–36, 1 2021.

[5] J. Shand, "Backlogs and Bulges: Can Value Chains Fix the National Health Service Waiting List?" 2022. [Online]. Available: https://catalyst.nejm.org/doi/full/10.1056/CAT.22.0252

[6] A. Kouri, S. Gupta, A. Yadollahi, C. M. Ryan, A. S. Gershon, T. To, S. M. Tarlo, R. S. Goldstein, K. R. Chapman, and C. W. Chow, "Addressing Reduced Laboratory-Based Pulmonary Function Testing During a Pandemic," pp. 2502–2510, 12 2020.

[7] E. C. Larson, M. Goel, G. Boriello, S. Heltshe, M. Rosenfeld, and S. N. Patel, "SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 280–289. [Online]. Available: https://doi.org/10.1145/2370216.2370261

[8] M. Goel, E. Saba, M. Stiber, E. Whitmire, J. Fromm, E. C. Larson, G. Borriello, and S. N. Patel, "SpiroCall: Measuring Lung Function over a Phone Call," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 5675–5685. [Online]. Available: https://doi.org/10.1145/2858036.2858401

[9] E. J. Sakka, P. Aggelidis, and M. Psimarnou, "Mobispiro: A Novel Spirometer," in *XII Mediterranean Conference on Medical and Biological Engineering and Computing*. Berlin: Springer Berlin Heidelberg, 2010, pp. 498–501.

[10] V. Viswanath, J. Garrison, and S. Patel, "SpiroConfidence: Determining the Validity of Smartphone Based Spirometry Using Machine Learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5499–5502. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8513516

[11] X. Song, B. Yang, G. Yang, R. Chen, E. Forno, W. Chen, and W. Gao, "SpiroSonic: Monitoring Human Lung Function via Acoustic Sensing on Commodity Smartphones," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3372224.3419209

[12] M. M. Rahman, T. Ahmed, E. Nemati, V. Nathan, K. Vatanparvar, E. Blackstock, and J. Kuang, "ExhaleSense: Detecting High Fidelity Forced Exhalations to Estimate Lung Obstruction on Smartphones," in *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2020, pp. 1–10.

[13] E. Nemati, X. Xu, V. Nathan, K. Vatanparvar, T. Ahmed, M. M. Rahman, D. McCaffrey, J. Kuang, and A. Gao, "Ubilung: Multi-Modal Passive-Based Lung Health Assessment," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 551–555.

[14] N. Saleheen, T. Ahmed, M. M. Rahman, E. Nemati, V. Nathan, K. Vatanparvar, E. Blackstock, and J. Kuang, "Lung Function Estimation from a Monosyllabic Voice Segment Captured Using Smartphones," in *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3379503.3403543

[15] K. S. Chun, V. Nathan, K. Vatanparvar, E. Nemati, M. M. Rahman, E. Blackstock, and J. Kuang, "Towards Passive Assessment of Pulmonary Function from Natural Speech Recorded Using a Mobile Phone," in *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2020, pp. 1–10.

[16] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231216003325

[17] C. Infante, D. Chamberlain, R. Fletcher, Y. Thorat, and R. Kodgule, "Use of cough sounds for diagnosis and screening of pulmonary disease," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, 2017, pp. 1–10.

[18] A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology*, vol. 1, 2002, pp. 182–183.

[19] J. Wells, "IPA transcription systems for English," *PG Bulletin, the bulletin of teachers of English phonetics in Chile and abroad*, vol. 9, 8 2001. [Online]. Available: https://www.phon.ucl.ac.uk/home/wells/ipa-english-uni.htm

[20] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. M. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, and J. Wanger, "Standardisation of spirometry," *European Respiratory Journal*, vol. 26, no. 2, pp. 319–338, 2005. [Online]. Available: https://erj.ersjournals.com/content/26/2/319

[21] B. L. Graham, I. Steenbruggen, M. R. Miller, I. Z. Barjaktarevic, B. G. Cooper, G. L. Hall, T. S. Hallstrand, D. A. Kaminsky, K. McCarthy, M. C. McCormack, C. E. Oropez, M. Rosenfeld, S. Stanojevic, M. P. Swanney, and B. R. Thompson, "Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement," *American Journal of Respiratory and Critical Care Medicine*, vol. 200, no. 8, pp. e70–e88, 10 2019.

[22] F. Pedregosa FABIANPEDREGOSA, V. Michel, O. Grisel OLIVIERGRISEL, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot andÉdouardand, a. Duchesnay, and F. Duchesnay EDOUARDDUCHESNAY, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," Tech. Rep., 2011. [Online]. Available: http://scikit-learn.sourceforge.net.

[23] M. Kos, Z. Kačič, and D. Vlaj, "Acoustic classification and segmentation using modified spectral roll-off and variance-based features," *Digital Signal Processing: A Review Journal*, vol. 23, no. 2, pp. 659–674, 2013.

[24] Mikhail Korobov and Konstantin Lopuhin, "ELI5," 2022. [Online]. Available: https://github.com/TeamHG-Memex/eli5

[25] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320315000989

[26] Medical technologies guidance, "KardiaMobile for detecting atrial fibrillation," The National Institute for Health and Care Excellence, Tech. Rep., 2022. [Online]. Available: www.nice.org.uk/guidance/mtg64