ACM DIGITAL LIBRARY • Association for Computing Machinery • acm open

SHORT-PAPER

Utilizing Speech as a Biosignal for Monitoring Respiratory Health and Beyond

SEJAL BHALLA, University of Toronto, Toronto, ON, Canada

EYAL DE LARA, University of Toronto, Toronto, ON, Canada

ALEX MARIAKAKIS, University of Toronto, Toronto, ON, Canada

Utilizing Speech as a Biosignal for Monitoring Respiratory Health and Beyond

Sejal Bhalla
University of Toronto
Toronto, Ontario, Canada
sejal@cs.toronto.edu

Eyal de Lara
University of Toronto
Toronto, Ontario, Canada
delara@cs.toronto.edu

Alex Mariakakis
University of Toronto
Toronto, Ontario, Canada
mariakakis@cs.toronto.edu

Abstract

Speech arises from the complex coordination of respiratory, neurological, cardiovascular, and muscular systems, making it a rich yet underutilized biosignal for health monitoring. This research explores the use of speech captured via mobile and wearable devices to assess and monitor respiratory health, with a focus on individuals with chronic obstructive pulmonary disease (COPD). It introduces methods for passive, continuous symptom monitoring using natural speech and explores structure-preserving speech representations to estimate lung function from acoustic features. Building on the insights from speech-based respiratory assessments, this thesis proposes a disease-agnostic speech diagnostic framework powered by self-supervised learning. By enabling low-burden, scalable assessments from natural speech, it supports equitable access to early diagnostics and personalized care across a broad spectrum of health conditions.

CCS Concepts

• **Human-centered computing** → *Ubiquitous and mobile devices*; • **Computing methodologies** → *Artificial intelligence*; • **Applied computing** → **Health informatics**; **Consumer health**.

Keywords

Mobile health, speech analysis, machine learning, foundation model, remote monitoring

1 Introduction

Chronic respiratory diseases, including chronic obstructive pulmonary disease (COPD) and asthma, are among the most widespread and debilitating health conditions globally. Affecting over 450 million people and causing millions of deaths annually [25], these diseases impose a substantial burden on individuals and healthcare systems alike. Effective disease management relies on regular assessments through lung function testing and symptom tracking, which are essential to guide treatment, monitor progression, and anticipate exacerbations. However, gold-standard spirometry requires active effort and clinical oversight, making it impractical for frequent use at home [3, 16]. Barriers such as patient frailty, the cost of lung function testing, and limited health literacy further restrict access, especially in underserved populations [12, 22]. As a result, there is growing interest in leveraging ubiquitous technologies to capture digital biomarkers of lung health remotely, passively, and with minimal user burden.

1.1 Problem Statement

Existing out-of-clinic monitoring techniques using mobile and wearable devices often depend on high-effort tasks, such as the six-minute walk test [10] or forced breathing maneuvers [17, 23, 36, 37] to elicit measurable indicators of lung function. While effective, these methods impose physical demands on users, limiting adherence and making them unsuitable for frequent monitoring, especially among individuals with severe illness. This highlights the need for sensing modalities that are minimally burdensome, unobtrusive, and seamlessly integrated into daily life.

Speech presents a physiologically grounded and low-burden alternative. Its production is tightly coupled with respiratory mechanics, as airflow through the vocal folds encodes information about pulmonary function. Unlike sporadic symptoms such as coughing or wheezing, speech offers a naturally occurring signal that can reflect both chronic impairment and acute changes in respiratory condition. Importantly, it can be passively captured using microphones in smartphones and smartwatches, enabling unobtrusive monitoring in everyday environments.

Despite its potential, speech remains underutilized in respiratory health monitoring. Lung function estimation studies have shown moderate accuracy [11], while symptom monitoring efforts often rely on controlled recordings [2, 28, 34, 39], limiting real-world applicability. Both directions require distinct yet complementary improvements; the former demands robust analysis of passively collected speech, and the latter calls for representations that capture the physiological instead of linguistic characteristics of speech. Addressing these challenges is key to building scalable, low-burden respiratory monitoring systems using ubiquitous devices.

1.2 Thesis Overview

The overall goal of my research is to leverage speech in novel ways for monitoring different health conditions, with a focus on respiratory health. This entails two objectives: (1) developing

systems for tracking daily variations in symptoms and (2) estimating critical lung function metrics.

To achieve the first objective, I leveraged a longitudinal speech dataset collected from COPD patients in real-world settings, paired with their daily symptom reports. I designed a robust speech processing pipeline, called PulmoListener [5], capable of handling the unique challenges posed by real-world audio, such as high noise levels and multiple speakers. To gain a deeper understanding of the underlying physiological processes during symptom exacerbations, I investigated the relationships between speech-derived features, physiological variables (e.g., heart rate variability and physical activity), and respiratory health outcomes. This analysis provided insights into the complex interplay between respiratory function, autonomic regulation, and vocal production, uncovering composite biomarkers of respiratory health.

To achieve the second objective, I explored speech representations and features tied to respiratory performance to reduce the error in computing key metrics of pulmonary function [4]. I will expand the scope of this research by creating a generalized speech diagnostic framework. This framework will leverage the complexity of vocal production to identify and monitor conditions beyond respiratory illnesses, establishing speech as a versatile biosignal for diverse health applications.

2 Related Work

This section reviews prior work on respiratory assessments, with an emphasis on the broader use of speech as a biosignal. It outlines the strengths and limitations of existing methods and describes how my research advances these efforts toward more robust and generalizable speech-based health diagnostics.

2.1 Ubiquitous Respiratory Assessments

Traditional assessments of chronic respiratory diseases such as COPD rely heavily on spirometry, which measures airflow limitation through parameters like forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC). While essential, these measures do not capture the full complexity of COPD, a multifaceted disease involving both pulmonary and systemic manifestations [7, 20, 21]. Therefore, modern respiratory assessments incorporate a combination of parameters—lung mechanics, airflow measurements, patient-reported symptoms, and vital signs—to facilitate holistic clinical decision making.

2.1.1 Symptoms. Self-reported symptoms articulate the challenges patients face in everyday life. Prior studies have demonstrated success in recognizing individual respiratory symptoms such as cold [34], coughing [26, 28, 38, 39], wheezing [8], and fever [6] using smartphones and smartwatches. However, most of these systems are designed to detect isolated symptoms and do not model respiratory health holistically, largely due to the lack of diverse, symptom-rich clinical datasets.

Sedaghat et al. [33] addressed this limitation by using passively collected speech and daily symptom scores to infer overall respiratory condition, marking a shift from symptom classification to severity estimation. However, their system's performance was limited because of manual feature extraction. My work has built

on this by using a more rigorous speech processing pipeline that accounts for temporal speech dynamics in free-living contexts.

2.1.2 Lung Function. Lung function estimation using mobile devices falls into two categories: (1) systems that record forced exhalation via embedded microphones [17, 23, 37], and (2) systems that infer lung function from acoustic features during prompted speech tasks [11, 29]. While the former closely resembles clinical spirometry, it places considerable burden on users and often requires supervision. The latter is more convenient but has demonstrated limited performance. Moreover, many studies focus on disease classification rather than estimating clinical metrics, limiting their clinical utility. My work has addressed these gaps by treating speech as a biosignal instead of a linguistic artifact, exploring new feature representations and modeling approaches that improve the precision of lung function estimation.

2.2 Speech as a Biosignal

Speech is not only affected by respiratory processes but also by neurological, psychological, cardiovascular, and muscular systems. Therefore, speech is a multidimensional biosignal with applications that extend beyond respiratory illnesses. Over the years, there has been a substantial body of work that has explored the use of speech processing for diagnostics [15, 18, 31, 33, 34], but most systems are usually tailored to a single disease. For example, models trained to detect dysarthria may perform poorly on tasks involving COVID-19 due to their reliance on disease-specific biomarkers.

Recent advances in self-supervised learning (SSL) [9, 19] offer a path toward universal speech representations, yet generalization remains a challenge as many models fail to perform well on unseen datasets due to confounding factors like noise, gender, and device variability [32, 41]. WavRx [40] takes a step toward disease-agnostic modeling using WavLM embeddings, but its scope is limited to only a few health conditions.

To address these limitations, my future research will expand the scope of speech diagnostics to a broader range of health conditions. Inspired by foundation modeling approaches in ECG and PPG [1], I propose physiologically-grounded training objectives that minimize the influence of confounding factors, thereby improving generalizability across health states and recording environments. Ultimately, this work will culminate in a scalable, disease-agnostic diagnostic framework that integrates robust speech processing with physiologically informed machine learning, enabling comprehensive health assessment through natural speech.

3 Methodology

This section presents three interconnected efforts toward robust speech-based health monitoring systems. The first focuses on predicting daily symptom severity in COPD patients using passively collected speech and wearable sensor data. The second targets lung function estimation using novel phoneme-aware speech representations, and the third outlines a future direction for a disease-agnostic framework built on self-supervised learning.
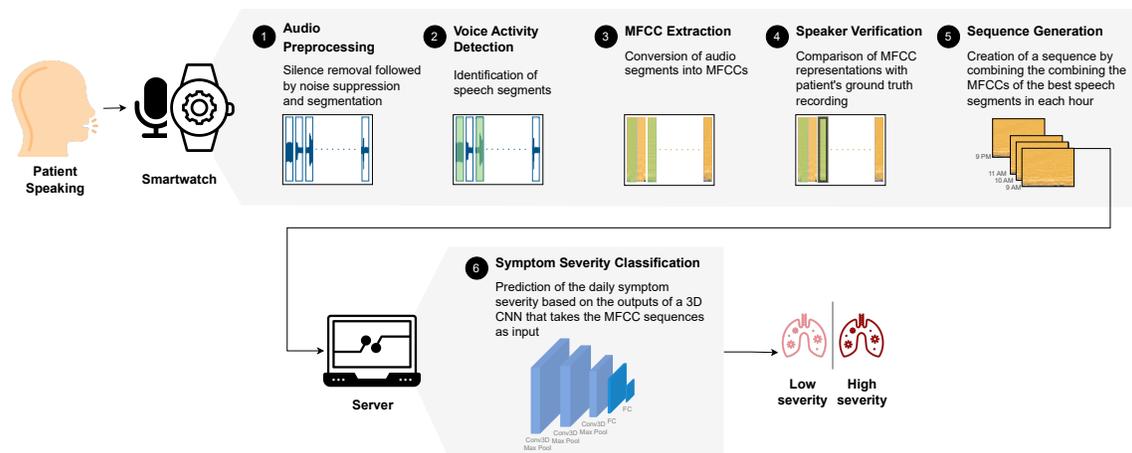
Figure 1: The system overview of PulmoListener. Audio recorded within a given hour is split, processed, and sorted such that the segment most likely containing the patient's own voice is identified. Features are extracted from those segments across all hours of the day to produce a daily feature vector that can be used for symptom severity classification.

3.1 Symptom Monitoring using Passively Collected Speech

PulmoListener [5] is an end-to-end speech processing pipeline designed to tackle the challenges of using real-world speech to track daily variations in COPD symptom severity. As shown in Figure 1, it combines voice activity detection and speaker verification to isolate speech segments relevant to speech analysis. These segments are then processed into sequences of features over the course of a day to account for diurnal modifiers of speech independent of respiratory health. PulmoListener was evaluated on the wearCOPD v2 dataset [35]—a longitudinal dataset comprising passively recorded smartwatch audio and daily symptom self-reports from eight individuals with COPD over an average duration of 164 ± 92 days. I investigated how this system could be used both for predicting same-day symptom severity and forecasting symptom severity several days in advance.

In a follow-up study[1], I extended this investigation to examine the relationship between interpretable vocal features and COPD symptom progression using an expanded version of the dataset. I analyzed how acoustic speech parameters correlate with daily self-reported symptoms and occurrences of exacerbations through linear mixed-effects models. Additionally, I incorporated other physiological variables, such as step count and heart rate variability, to explore their combined predictive value. By examining these factors concurrently, this work contributes to a more holistic understanding of COPD and supports the development of comprehensive, multimodal monitoring strategies.

3.2 Phoneme-Aware Speech Analysis for Lung Function Estimation

Prior work has used temporal and spectral features from natural speech to estimate lung function metrics, often relying on fixed-length windows to segment audio [11, 33]. This approach results in a heterogeneous feature space that requires training on large, diverse datasets for robust performance. Alternatively, some studies prompt users to produce sustained phonemes (e.g., /s/ or /aa/), yielding more consistent features but relying on unnatural tasks that may not generalize to real-world speech [13, 14, 24, 29]. To bridge these approaches, I leveraged phoneme-level segments extracted from natural speech for lung function estimation.

For this analysis, I used the in-lab subset of the wearCOPD v2 dataset [35], which includes 11 COPD patients who read scripted passages and completed spirometry testing. After using forced alignment to match the audio recordings with their scripts, various phonation, prosodic, and spectral features were then calculated for each identified phoneme. I then conducted a correlation analysis to evaluate associations between these features and lung function metrics. This was done for both traditional sliding window-based features and phoneme segments to assess which segmentation strategy yields stronger associations with lung function.

3.3 Disease-Agnostic Speech Diagnostic Framework (Future Work)

Building on insights from my previous studies, I will conclude my thesis by developing a self-supervised foundation model for speech-based health monitoring. To achieve this, I will curate large-scale speech datasets spanning respiratory (COPD and asthma), neurodegenerative (Parkinson's and Alzheimer's), psychological (depression and emotion), cardiovascular (heart failure and stroke), and cardiometabolic conditions (hypertension and diabetes). This will allow me to benchmark how well various models, including the proposed self-supervised approach, generalize across diseases.

The model will be trained on both healthy and pathological speech data to improve robustness and reduce reliance on generic speech corpora. Inspired by structure-preserving objectives used in foundation models for other physiological signals (e.g., ECG and PPG) [1, 27], I will incorporate physiological constraints into the pretraining phase. Specifically, I will design a novel pretext task

---

[1]This work was recently accepted in Nature Scientific Reports.

that preserves the phonetic structure of speech during representation learning. Since individual phonemes engage the respiratory, neurological, and muscular systems in distinct ways, the variation in their representations can reveal different physiological indicators of disease.

4 Evaluation

This section summarizes the empirical findings from completed studies and outlines the future directions that build on these results to advance speech-based health diagnostics.

4.1 Key Findings

PulmoListener achieved an average sensitivity of 0.79 ± 0.03 and specificity of 0.83 ± 0.05 when classifying daily symptom severity in COPD patients. It also demonstrated the ability to forecast symptom severity up to four days in advance, with a sensitivity of 0.75 ± 0.02 and specificity of 0.74 ± 0.07. These results highlight the feasibility of using passively collected natural speech for real-world symptom monitoring and early intervention.

To better characterize which aspects of speech contribute to these predictions, I examined the association between vocal features and COPD outcomes. Categories of features reflecting phonation, prosody, and articulation each exhibited unique relationships with symptom severity and exacerbation risk. For instance, phonation features like jitter and shimmer decreased on days with elevated symptom scores, while prosodic markers such as the mean fundamental frequency and its variability were positively associated with worsening conditions. These associations were further modulated by physiological variables. For example, during periods of low HRV, higher jitter was linked to worsening symptoms. Reduced physical activity similarly intensified these associations, aligning with prior evidence that inactivity is a marker of exacerbation risk.

For lung function estimation, Figure 2 shows the distribution of correlation magnitudes between phoneme-level audio features and FEV1. While the range reflects variability in predictive power, six of the fourteen phonemes resulted in correlations that exceeded the baselines, demonstrating the utility of analyzing speech at the phoneme level. Similar results were observed when the dataset was analyzed with respect to FVC as the target lung function metric. The disparate combinations of phonemes and audio features surfaced by these experiments illustrate the utility of restricting analysis to individual phonemes. Further, the variability in correlation direction across different phonemes highlights that distinct sounds may capture unique aspects of lung function. Since each phoneme engages specific articulatory and respiratory mechanisms, aggregating them into unstructured segments risks obscuring these informative patterns.

4.2 Future Work

The next phase of my research will focus on developing a disease-agnostic speech diagnostic model. As outlined in Section 3.3, this model will leverage phonetic training objectives and self-supervised learning. To ensure robust and reproducible evaluation, I will construct a benchmark spanning general-purpose and disease-specific speech tasks. Recognizing that speech
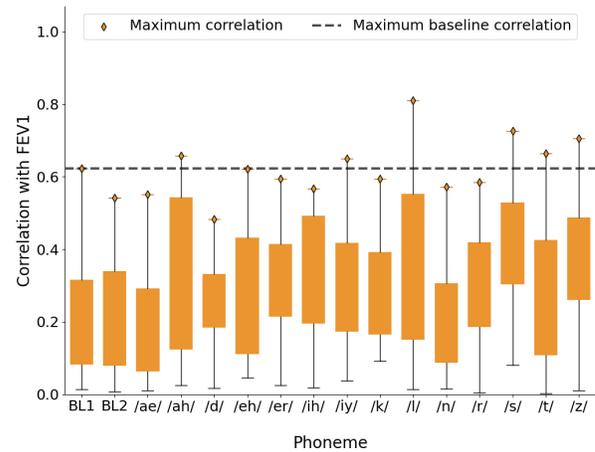

Figure 2: Box-and-whisker plots showing how the distribution of correlation magnitudes between audio features and FEV1 varies across different phonemes. The two leftmost distributions correspond to phoneme-agnostic baselines BL1 and BL2 with window lengths of 2 seconds and 120 milliseconds, respectively.

patterns vary across age, gender, and ethnicity, I will assess the fairness of the model using frameworks such as HEAL [30], refining the training objectives and architecture to improve both accuracy and equity across populations.

5 Expected Contribution

This research makes four key contributions toward advancing speech as a biosignal for health monitoring:

(1) It introduces a robust pipeline for passive, continuous speech monitoring in real-world settings, enabling daily assessment of symptom severity and early detection of deterioration in individuals with COPD.
(2) It develops methods for integrating interpretable vocal features with other physiological signals to identify multimodal markers of COPD progression.
(3) It improves upon prior approaches to speech-based lung function estimation by developing novel phoneme-aware speech representations that align more closely with underlying respiratory physiology.
(4) It lays the foundation for a disease-agnostic model by leveraging self-supervised learning and large-scale, multi-condition speech datasets to support generalization across health conditions.

By addressing key limitations in digital health systems, this work supports scalable, low-cost health monitoring using everyday devices. Additionally, the introduction of standardized tools and benchmarks will support reproducibility and accelerate future work in speech-based health diagnostics.

Acknowledgments

References

[1] Salar Abbaspourazad, Oussama Elachqar, Andrew C. Miller, Saba Emrani, Udhyakumar Nallasamy, and Ian Shapiro. 2024. Large-scale Training of Foundation Models for Wearable Biosignals. arXiv:2312.05409 [cs.LG] https://arxiv.org/abs/2312.05409

[2] Forsad Al Hossain, Andrew A. Lover, George A. Corey, Nicholas G. Reich, and Tauhidur Rahman. 2020. FluSense: A Contactless Syndromic Surveillance Platform for Influenza-Like Illness in Hospital Waiting Areas. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 1 (mar 2020), 28 pages. doi:10.1145/3381014

[3] Mirthe Bakker, Joanne Sloots, Clara Van Ommeren, Alexandra Kleberger, Monique Tabak, Tanja Effing, Gerard Linssen, Martijn Grinovero, Job Van Der Palen, and Anke Lenferink. 2019. Adherence in patients with COPD and heart failure using an eHealth self-management intervention. *European Respiratory Journal* 54, suppl 63 (2019). doi:10.1183/13993003.congress-2019.PA744

[4] Sejal Bhalla, Tien Han, Andrea Gershon, Robert Wu, Eyal de Lara, and Alex Mariakakis. 2025. Phoneme-Aware Acoustic Analysis of Natural Speech for Lung Function Assessment. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. doi:10.1109/ICASSP49660.2025.10888421

[5] Sejal Bhalla, Salaar Liaqat, Robert Wu, Andrea S. Gershon, Eyal de Lara, and Alex Mariakakis. 2023. PulmoListener: Continuous Acoustic Monitoring of Chronic Obstructive Pulmonary Disease in the Wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 86 (sep 2023), 24 pages. doi:10.1145/3610889

[6] Joseph Breda, Mastafa Springston, Alex Mariakakis, and Shwetak Patel. 2023. FeverPhone: Accessible Core-Body Temperature Sensing for Fever Monitoring Using Commodity Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 3 (mar 2023), 23 pages. doi:10.1145/3580850

[7] M. Cazzola, W. MacNee, F. J. Martinez, K. F. Rabe, L. G. Franciosi, P. J. Barnes, V. Brusasco, P. S. Burge, P. M. A. Calverley, B. R. Celli, P. W. Jones, D. A. Mahler, B. Make, M. Miravitlles, C. P. Page, P. Palange, D. Parr, M. Pistolesi, S. I. Rennard, M. P. Rutten-van Mölken, R. Stockley, S. D. Sullivan, J. A. Wedzicha, and E. F. Wouters. 2008. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *European Respiratory Journal* 31, 2 (2008), 416–469. arXiv:http://erj.ersjournals.com/content/erj/31/2/416.full.pdf doi:10.1183/09031936.00099306

[8] Soujanya Chatterjee, Md Mahbubur Rahman, Tousif Ahmed, Nazir Saleheen, Ebrahim Nemati, Viswam Nathan, Korosh Vatanparvar, and Jilong Kuang. 2020. Assessing Severity of Pulmonary Obstruction from Respiration Phase-Based Wheeze-Sensing Using Mobile Sensors. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376444

[9] Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Jian Wu, Michael Zeng, Xiangzhan Yu, and Furu Wei. 2022. WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing. *IEEE Journal of Selected Topics in Signal Processing* 16, 6 (Oct. 2022), 1505–1518. doi:10.1109/jstsp.2022.3188113

[10] Qian Cheng, Joshua Juen, Shashi Bellam, Nicholas Fulara, Deanna Close, Jonathan Silverstein, and B. Schatz. 2017. Predicting Pulmonary Function from Phone Sensors. *Telemedicine journal and e-health : the official journal of the American Telemedicine Association* 23 (03 2017). doi:10.1089/tmj.2017.0008

[11] Keum San Chun, Viswam Nathan, Korosh Vatanparvar, Ebrahim Nemati, Md Mahbubur Rahman, Erin Blackstock, and Jilong Kuang. 2020. 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)*. 1–10. doi:10.1109/PerCom45495.2020.9127380

[12] RJ Cooper, J Colligan, S Hamilton, E Finlayson, M Duffy, J Gilliatt, M Swanson, A Giangreco, and EK Sage. 2021. Evaluation of myCOPD, a digital self-management technology for people with COPD, in a remote and rural population. *Thorax* 76, A169.2–A170. doi:10.1136/thorax-2020-BTSabstracts.295

[13] Biswajit Das, Khalid Daoudi, Jiri Klempir, and Jan Rusz. 2019. Towards disease-specific speech markers for differential diagnosis in parkinsonism. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 5846–5850.

[14] Muzeyyen Dogan, Emel Eryuksel, Ismail Kocak, Turgay Celikel, and Mehmet Ali Sehitoglu. 2007. Subjective and Objective Evaluation of Voice Quality in Patients With Asthma. *Journal of Voice* 21, 2 (2007), 224–230. doi:10.1016/j.jvoice.2005.11.003

[15] G. M. Dyukova, Zh. M. Glozman, E. Yu. Titova, E. S. Kriushev, and A. A. Gamaleya. 2010. Speech Disorders in Right-Hemisphere Stroke. *Neuroscience and Behavioral Physiology* 40, 6 (July 2010), 593–602. doi:10.1007/s11055-010-9301-9

[16] Lydia J. Finney, Stefan Avey, Dexter Wiseman, Anthony Rowe, Matthew J. Loza, Patrick Branigan, Christopher S. Stevenson, Frédéric Baribaud, Jadwiga A. Wedzicha, Ioannis Pandis, and Gavin C. Donaldson. 2023. Using an electronic diary and wristband accelerometer to detect exacerbations

and activity levels in COPD: a feasibility study. *ERJ Open Research* 9, 6 (2023). arXiv:https://publications.ersnet.org//content/erjor/9/6/00366-2023.full.pdf doi:10.1183/23120541.00366-2023

[17] Mayank Goel, Elliot Saba, Maia Stiber, Eric Whitmire, Josh Fromm, Eric C. Larson, Gaetano Borriello, and Shwetak N. Patel. 2016. SpiroCall: Measuring Lung Function over a Phone Call. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5675–5685. doi:10.1145/2858036.2858401

[18] Melisa Gumus, Danielle D. DeSouza, Mengdan Xu, Celia Fidalgo, William Simpson, and Jessica Robin. 2023. Evaluating the utility of daily speech assessments for monitoring depression symptoms. *Digital Health* 9 (2023), 20552076231180523. doi:10.1177/20552076231180523

[19] Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. 2021. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 29 (Oct. 2021), 3451–3460. doi:10.1109/TASLP.2021.3122291

[20] P W Jones. 2001. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 56, 11 (2001), 880–887. arXiv:https://thorax.bmj.com/content/56/11/880.full.pdf doi:10.1136/thorax.56.11.880

[21] P. W. Jones, G. Harding, P. Berry, I. Wiklund, W-H. Chen, and N. Kline Leidy. 2009. Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 34, 3 (2009), 648–654. arXiv:http://erj.ersjournals.com/content/erj/34/3/648.full.pdf doi:10.1183/09031936.00102509

[22] Christina Emme Kira Marie Skibdal and Henrik Hansen. 2022. Listen to Me! – A Mixed-Methods Study of Thoughts and Attitudes Towards Participation in Pulmonary Telerehabilitation Among People with Severe and Very Severe COPD Who Declined Participation in Pulmonary Rehabilitation. *Patient Preference and Adherence* 16 (2022), 2781–2798. arXiv:https://www.tandfonline.com/doi/pdf/10.2147/PPA.S380832 doi:10.2147/PPA.S380832 PMID: 36281352.

[23] Eric C. Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N. Patel. 2012. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 280–289. doi:10.1145/2370216.2370261

[24] Youri Maryn, Paul Corthals, Paul Van Cauwenberge, Nelson Roy, and Marc De Bodt. 2010. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice* 24, 5 (2010), 540–555.

[25] Sara Momtazmanesh, Sahar Saeedi Moghaddam, Seyyed-Hadi Ghamari, Elaheh Malakan Rad, Negar Rezaei, Parnian Shobeiri, and all. 2023. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. *eClinicalMedicine* 59 (2023), 101936. doi:10.1016/j.eclinm.2023.101936

[26] Ebrahim Nemati, Md. Juber Rahman, Erin Blackstock, Viswam Nathan, Md. Mahbubur Rahman, Korosh Vatanparvar, and Jilong Kuang. 2020. Estimation of the Lung Function Using Acoustic Features of the Voluntary Cough. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, Montreal, QC, Canada, 4491–4497. doi:10.1109/EMBC44109.2020.9175986

[27] Arvind Pillai, Dimitris Spathis, Fahim Kawsar, and Mohammad Malekzadeh. 2024. PaPaGei: Open Foundation Models for Optical Physiological Signals. *arXiv preprint arXiv:2410.20542* (2024).

[28] Vishwajith Ramesh, Korosh Vatanparvar, Ebrahim Nemati, Viswam Nathan, Md Mahbubur Rahman, and Jilong Kuang. 2020. CoughGAN: Generating Synthetic Coughs that Improve Respiratory Disease Classification. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, Montreal, QC, Canada, 5682–5688. doi:10.1109/EMBC44109.2020.9175597

[29] Nazir Saleheen, Tousif Ahmed, Md Mahbubur Rahman, Ebrahim Nemati, Viswam Nathan, Korosh Vatanparvar, Erin Blackstock, and Jilong Kuang. 2020. Lung Function Estimation from a Monosyllabic Voice Segment Captured Using Smartphones. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 10, 11 pages. doi:10.1145/3379503.3403543

[30] Mike Schaekermann, Terry Spitz, Malcolm Pyles, Heather Cole-Lewis, Ellery Wulczyn, Stephen Pfohl, Donald Martin, Ronnachai Jaroensri, Geoff Keeling, Yuan Liu, Stephanie Farquhar, Qinghan Xue, Jenna Lester, Cían Hughes, Patricia Strachan, Fraser Tan, Peggy Bui, Craig Mermel, Lily Peng, and Po-Hsuan Chen. 2024. Health equity assessment of machine learning performance (HEAL): a framework and dermatology AI model case study. *eClinicalMedicine* 70 (03 2024), 102479. doi:10.1016/j.eclinm.2024.102479

[31] Klaus R. Scherer, Felix Burkhardt, Uwe D. Reichel, Florian Eyben, and Björn W. Schuller. 2024. Using voice analysis as an early indicator of risk for depression in young adults. doi:10.48550/ARXIV.2411.11541 Version Number: 1.

[32] Guilherme Schu, Parvaneh Janbakhshi, and Ina Kodrasi. 2022. On Using the UA-Speech and Torgo Databases to Validate Automatic Dysarthric Speech Classification Approaches. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2022), 1–5. https://api.semanticscholar.org/CorpusID:253553250

[33] Tina Sedaghat, Salaar Liaqat, Daniyal Liaqat, Robert Wu, Andrea Gershon, Tatiana Son, Tiago H. Falk, Moshe Gabel, Alex Mariakakis, and Eyal de Lara. 2022. Unobtrusive Monitoring of COPD Patients using Speech Collected from Smartwatches in the Wild. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 818–823. doi:10.1109/PerComWorkshops53856.2022.9767283

[34] Xiao Sun, Zongqing Lu, Wenjie Hu, and Guohong Cao. 2015. SymDetector: Detecting Sound-Related Respiratory Symptoms Using Smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 97–108. doi:10.1145/2750858.2805826

[35] Robert Wu, Daniyal Liaqat, Eyal de Lara, Tatiana Son, Frank Rudzicz, Hisham Alshaer, Pegah Abed-Esfahani, and Andrea S Gershon. 2018. Feasibility of Using a Smartwatch to Intensively Monitor Patients With Chronic Obstructive Pulmonary Disease: Prospective Cohort Study. *JMIR Mhealth Uhealth* 6, 6 (14 Jun 2018), e10046. doi:10.2196/10046

[36] Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2023. EarSpiro: Earphone-Based Spirometry for Lung Function Assessment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 188 (jan 2023), 27 pages. doi:10.1145/3569480

[37] Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2023. EarSpiro: Earphone-Based Spirometry for Lung Function Assessment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 188 (jan 2023), 27 pages. doi:10.1145/3569480

[38] Wenlong Xu, Guoqiang He, Chen Pan, Dan Shen, Ning Zhang, Peirong Jiang, Feng Liu, and Jingjing Chen. 2022. A forced cough sound based pulmonary function assessment method by using machine learning. *Front. Public Health* 10 (Oct. 2022), 1015876. doi:10.3389/fpubh.2022.1015876

[39] Xuhai Xu, Ebrahim Nemati, Korosh Vatanparvar, Viswam Nathan, Tousif Ahmed, Md Mahbubur Rahman, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao. 2021. Listen2Cough: Leveraging End-to-End Deep Learning Cough Detection Model to Enhance Lung Health Assessment Using Passively Sensed Audio. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 42 (mar 2021), 22 pages. doi:10.1145/3448124

[40] Yi Zhu and Tiago Falk. 2024. WavRx: a Disease-Agnostic, Generalizable, and Privacy-Preserving Speech Health Diagnostic Model. *IEEE Journal of Biomedical and Health Informatics* (2024), 1–14. doi:10.1109/jbhi.2024.3454550

[41] Yi Zhu, Mohamed Imoussaïne-Aïkous, Carolyn Côté-Lussier, and Tiago H. Falk. 2023. Investigating Biases in COVID-19 Diagnostic Systems Processed with Automated Speech Anonymization Algorithms. In *3rd Symposium on Security and Privacy in Speech Communication*. 46–54. doi:10.21437/SPSC.2023-8