BreathTrack: Detecting Regular Breathing Phases from Unannotated Acoustic Data Captured by a Smartphone

BASHIMA ISLAM*†, University of North Carolina at Chapel Hill
MD MAHBUBUR RAHMAN*, Digital Health Lab, Samsung Research America
TOUSIF AHMED, Digital Health Lab, Samsung Research America
MOHSIN YUSUF AHMED, Digital Health Lab, Samsung Research America
MD MEHEDI HASAN, Digital Health Lab, Samsung Research America
VISWAM NATHAN, Digital Health Lab, Samsung Research America
KOROSH VATANPARVAR, Digital Health Lab, Samsung Research America
EBRAHIM NEMATI, Digital Health Lab, Samsung Research America
JILONG KUANG, Digital Health Lab, Samsung Research America
JUN ALEX GAO, Digital Health Lab, Samsung Research America

Breathing biomarkers, such as breathing rate, fractional inspiratory time, and inhalation-exhalation ratio, are vital for monitoring the user's health and well-being. Accurate estimation of such biomarkers requires breathing phase detection, i.e., inhalation and exhalation. However, traditional breathing phase monitoring relies on uncomfortable equipment, e.g., chestbands. Smartphone acoustic sensors have shown promising results for passive breathing monitoring during sleep or guided breathing. However, detecting breathing phases using acoustic data can be challenging for various reasons. One of the major obstacles is the complexity of annotating breathing sounds due to inaudible parts in regular breathing and background noises. This paper assesses the potential of using smartphone acoustic sensors for passive unguided breathing phase monitoring in a natural environment. We address the annotation challenges by developing a novel variant of the teacher-student training method for transferring knowledge from an inertial sensor to an acoustic sensor, eliminating the need for manual breathing sound annotation by fusing signal processing with deep learning techniques. We train and evaluate our model on the breathing data collected from 131 subjects, including healthy individuals and respiratory patients. Experimental results show that our model can detect breathing phases with 77.33% accuracy using acoustic sensors. We further present an example use-case of breathing phase-detection by first estimating the biomarkers from the estimated breathing phases and then using these biomarkers for pulmonary patient detection. Using the detected breathing phases, we can estimate fractional inspiratory time with 92.08% accuracy, the inhalation-exhalation ratio with 86.76% accuracy, and the breathing rate with

*Both authors contributed equally to this research.
†Pursued this work while interning with Samsung Research America.

Authors' addresses: Bashima Islam, University of North Carolina at Chapel Hill, bashima@cs.unc.edu; Md Mahbubur Rahman, m.rahman2@samsung.com, Digital Health Lab, Samsung Research America; Tousif Ahmed, Digital Health Lab, Samsung Research America, tousif.a@samsung.com; Mohsin Yusuf Ahmed, Digital Health Lab, Samsung Research America, mohsin.ahmed@samsung.com; Md Mehedi Hasan, Digital Health Lab, Samsung Research America, mehedi2003@gmail.com; Viswam Nathan, Digital Health Lab, Samsung Research America, viswam.nathan@samsung.com; Korosh Vatanparvar, Digital Health Lab, Samsung Research America, korosh.v@samsung.com; Ebrahim Nemati, Digital Health Lab, Samsung Research America, e.nemati@samsung.com; Jilong Kuang, Digital Health Lab, Samsung Research America, jilong.kuang@samsung.com; Jun Alex Gao, Digital Health Lab, Samsung Research America, alex.gao@samsung.com.

2474-9567/2021/9-ART124 $15.00
https://doi.org/10.1145/3478123

91.74% accuracy. Moreover, we can distinguish respiratory patients from healthy individuals with up to 76% accuracy. This paper is the first to show the feasibility of detecting regular breathing phases towards passively monitoring respiratory health and well-being using acoustic data captured by a smartphone.

CCS Concepts: • **Human-centered Computing → Ubiquitous and Mobile Computing**; • **Information Systems →** *Data Mining*.

Additional Key Words and Phrases: Breathing, Smartphone, Audio, Teacher-Student Model, Respiratory Diseases

ACM Reference Format:
Bashima Islam, Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Md Mehedi Hasan, Viswam Nathan, Korosh Vatanparvar, Ebrahim Nemati, Jilong Kuang, and Jun Alex Gao. 2021. BreathTrack: Detecting Regular Breathing Phases from Unannotated Acoustic Data Captured by a Smartphone. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 124 (September 2021), 22 pages. https://doi.org/10.1145/3478123

1 INTRODUCTION

Breathing rate is one of the most critical vital signs. Abnormalities in breathing rate per minute can often indicate severe underlying diseases [16]. For example, breathing rate elevation can be a strong predictor of cardiac arrest [9, 18]. Due to its importance for our health and well-being monitoring, breathing rate measurement in uncontrolled settings has been a critical research topic [4, 5, 33, 70, 73]. While minute-level average breathing rate is unquestionably a significant biomarker, more sophisticated health and well-being monitoring applications require other fine-grained breathing biomarkers such as fractional inspiratory time and inhalation-exhalation ratio. For example, reduced fractional inspiratory time can predict respiratory disease deterioration [10], and the inhalation-exhalation ratio can detect psycho-social stress [7, 49] and social interaction [6]. Breathing phase (inhalation and exhalation) detection in seconds enables the extraction of these fine-grained breathing biomarkers along with breathing rate. Moreover, detecting and tracking the breathing phases is itself beneficial for monitoring health and well-being. Real-time inhalation and exhalation tracking can better guide the user's breathing exercises to enhance mindfulness and calmness [19, 70]. Recent work [11] on breathing phase detection-based wheeze sound analysis shows that it is possible to estimate the severity of a pulmonary patient using a smartphone. Among the findings were indications that the presence of wheeze in the inhalation phase indicates a more severe lung condition in comparison with the presence of wheeze in the exhalation phase.

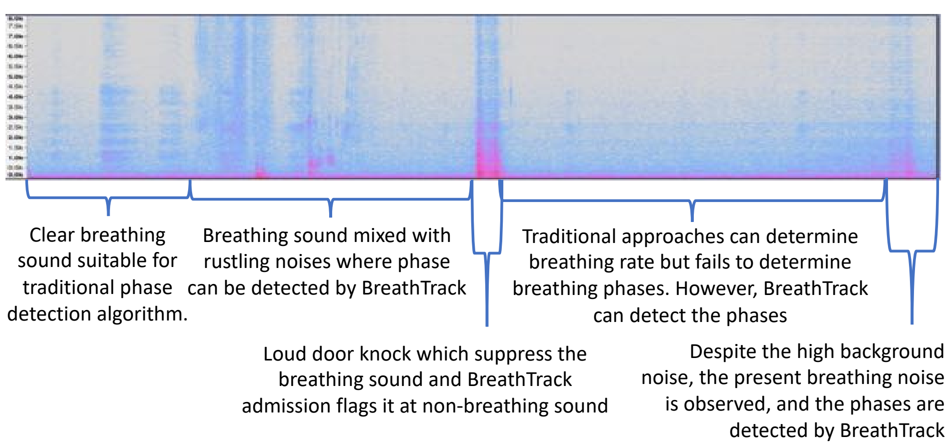

Fig. 1. Spectrogram of a regular breathing audio in the dataset. It shows breathing sounds and other environmental noises.

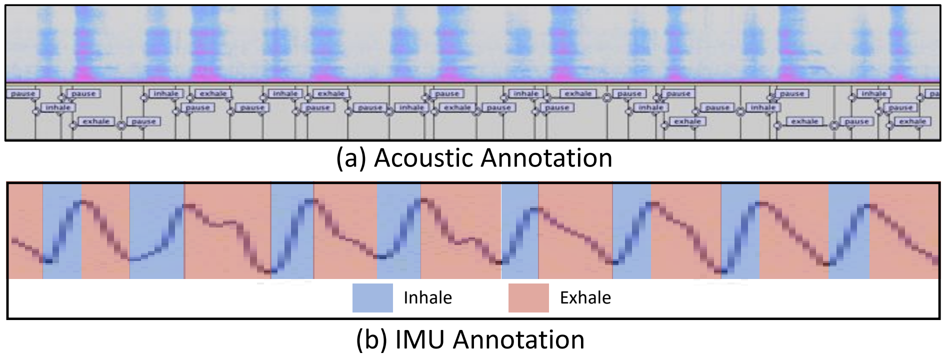

Fig. 2. (a) Regular breathing sound recording without background noise. The human annotator observes pauses in the acoustic data. (b) However, actual regular breathing does not have any pause as seen in the corresponding inertial signal.

Despite having many applications, current breathing monitoring solutions in everyday life use constraining and uncomfortable equipment, such as chestbands [7]. Although several works have utilized inertial sensors embedded in smartphones for breathing biomarker detection, they require active user participation, which cannot be performed continuously [2, 11]. Recently sensors embedded in smartwatches have shown promising results for estimating breathing rate in everyday life [33]. However, such approaches are limited to breathing rate estimation, and the adoption of smartwatches is still low (21%) compared to smartphone adoption (85% among US adults). Therefore, a smartphone-based breathing phase detector would be a more valuable and ubiquitous solution. Among all the sensors embedded in smartphones, acoustic sensors (microphones) are the most promising candidate to realize this goal as audio can be captured on-body or off-body when the phone is near the user.

Previous works show that smartphone microphones can capture breathing sounds and detect breathing phases when the phone is near the user during sleep [63] and breathing exercises [70]. This paper shows the feasibility of using a smartphone's microphone to analyze regular breathing sounds in uncontrolled scenarios to detect dynamically varying breathing phases. However, the detection model requires labeled data for training and validation, and annotating regular breathing sounds is challenging. There are three significant obstacles to annotate regular breathing sounds compared to sleep breathing and breathing exercises. First, regular breathing sounds are less audible and are often quite challenging to annotate due to background noise (Figure 1), unlike the breathing sound captured around the nose [15] or by a specialized device such as a stethoscope [55]. Next, even when the presence of breathing can be confirmed in an audio clip, a portion of the sound can be inaudible. This inaudibility occurs due to breathing mechanics where the beginning of each breathing phase has the highest airflow, which gradually reduces towards the phase switching point. Therefore, we may observe a pause between the breathing phases in the audio recording, although there is no pause between the two consecutive breathing phases during regular breathing. To illustrate, in Figure 2(a), we demonstrate that even in a clean acoustic breathing signal, there are apparent pauses which impose difficulty for the human annotators. To confirm that there are no pauses, we analyze the motion of the chest during breathing and visualize the corresponding motion signal in Figure 2(b). Finally, the duration of the breathing phases are highly dynamic in regular breathing and vary from 0.2 seconds to 5 seconds [2] compared to fixed duration in controlled breathing [70].

To overcome these challenges, we present a novel approach to automatically annotate regular breathing sounds and detect the breathing phases using acoustic signals captured by a smartphone. We develop a novel variant of the teacher-student knowledge transfer method to annotate the breathing sounds utilizing motion sensors. This approach uses a lightweight signal processing technique on motion sensor data as a teacher that teaches the dynamically varying breathing phases for the acoustic sensor-based deep learning student model. The deep learning-based student model automatically extracts the common patterns from both audible and inaudible audio data. After training the student model once, our approach only requires the acoustic data.

This paper focuses on pulmonary patients because such a breathing monitoring system will be highly beneficial for this population. Besides, studies show that breathing is heavier for respiratory patients than that for healthy individuals [39]. Therefore, we train our model using regular breathing data collected from 131 subjects (91 chronic respiratory patients and 40 healthy subjects). Furthermore, unlike previous works requiring the users to breathe in through the nose and breathe out through the mouth and follow a particular pattern (e.g., inhale for 4-second inhale followed by a 2-second pause, 4-second exhale) [70], we do not guide users' breathing during data collection. Instead, we capture both acoustic and motion (accelerometer and gyroscope) data simultaneously when the user is sitting or is in a supine posture. We request to place the smartphone on the chest for the convenience of the multi-modal data collection from real, untrained patients. This study design also addresses the problem shown in previous works where data collection using more than one device suffers from time synchronization issues [3]. Note that the orientation and placement of the smartphone were unguided.

Our model first identifies the onset of the breathing in an audio stream with 96% accuracy to tackle the presence of high background noise. Then, it pinpoints the inhalation and exhalation phases with 77% accuracy. Moreover, we show that our model enables accurate extraction of fine-grained breathing biomarkers, including fractional inspiratory time, inhalation-exhalation ratio, breathing rate with 92.08%, 86.76%, and 91.74% accuracy, respectively. Finally, we further present how these biomarkers can be utilized for differentiating respiratory patients from healthy individuals.

The key contributions of this paper are summarized below.

- We show the feasibility of detecting regular breathing phases using smartphone acoustic data to enable the passive extraction of fine-grained breathing biomarkers.
- We propose a novel variant of the teacher-student training method by fusing signal processing with deep learning techniques to train the breathing phase-detection model using unannotated acoustic data. It transfers knowledge from the inertial sensor (IMU) to acoustic data. This model detects fine-grained breathing phases in dynamically varying regular breathing cycles using smartphone acoustic signal. Our proposed model can achieve 7.22% higher accuracy than the state-of-the-art acoustic breathing phase detector [70].
- Through a study with 131 unguided subjects, we present an example use case of our proposed framework. We show that our framework can extract novel breathing biomarkers with up to 92.08% accuracy and achieves 8% higher accuracy in distinguishing respiratory patients from healthy individuals compared to the state-of-the-art speech sound-based patient detection model [45].

2 RELATED WORKS

This section describes the existing breathing detection and analysis models, and the differences between our approach and the existing methods. We also describe the background of the traditional-teacher student model and the novelty of the proposed variation presented in this paper.

2.1 Breathing Rate and Breathing Phase Detection

2.1.1 Breathing Detection and Breathing Rate Estimation. Breathing is a complex physiological process of inhaling oxygen from the air and exhaling carbon dioxide from the body. Breathing rate is one of the critical vital signs and can be extracted from other physiological signals, airflow, and respiratory-related body movements [36]. A sensor or device next to the mouth or nose is a common approach to estimating airflow/breathing [4]. Breathing-related body movements can also be easily extracted by using inertial measurement units (IMU) attached to chest bands [54, 58], head-mounted devices [26], and wrist-worn devices [27, 73]. Several of those works either focus on user-initiated breathing measurement [26] or covers constrained environments such as sleep [44, 73]. Previous works also explored contact-less breathing measurement with a motion capture system [72], a smartphone microphone [59], and analysis of high-frequency wireless signals [57]. All previous approaches mentioned above show the feasibility of actively or passively detecting breathing using mobile sensors and focusing on breathing rate estimation. In this work, we move this effort beyond breathing rate estimation and show that regular breathing phases (inhalation and exhalation) can also be monitored using acoustic data collected using a smartphone. Our approach enables the mobile sensors to extract fine-grained digital breathing biomarkers (e.g., fractional inspiratory time, inhalation exhalation ratio) to track pulmonary patients in the free-living scenario.

A few recent works highlighted the other benefits of extracting breathing phases using mobile sensors. For example, in our previous work, mWheeze [11], we presented an algorithm to extract breathing phase information using smartphone IMU data. We showed that it improves the accuracy of pulmonary patient classification. Our other work, mLung [2], used a similar approach to detect breathing phases and used it as a natural windowing technique to segment the audio stream to train machine learning models to detect lung sound events for pulmonary patients. However, those approaches work only if the user consciously holds their phone on the chest to capture the chest motion using the IMU sensor. While our model also utilizes the breathing-related chest motions, it transfers the knowledge of breathing phases from the IMU to audio data. Thus, our approach can be more convenient to the user as it offers hands-free interaction and can work with only the acoustic data captured by the smartphone around the user.

2.1.2 Acoustic Sensor Based Breathing Phase Detection during Mindful Breathing Exercises. A class of smartphone-based breath detection works focused on controlled breathing to provide bio-feedback for meditation [23, 70]. The contemporary related work is Breeze [70] that detects breathing phases (inhale, exhale and pause) using a smartphone microphone to provide bio-feedback for a structured breathing exercise. Despite the similarities, there are some significant differences between our model and Breeze. Breeze focuses on a particular regulated breathing exercise, consisting of four seconds of inhalation, followed by two seconds of exhalation and four seconds of pause. On the other hand, we focus on natural breathing where conscious breath holding is not present. Besides, the duration of regular breathing phases is more dynamic. Finally, during data collection, Breeze provides regulated instruction to inhale with the nose and exhale through the mouth where the sound source's difference contributes to the acoustic signature. In our data collection, we asked users to breathe normally without restricting the nose or mouth breathing. Therefore, our model is the first work that focuses on natural breathing phase monitoring. In addition to breathing phase monitoring, our approach can extract critical and fine-grained breathing biomarkers for assessing respiratory conditions, including inhalation-exhalation ratio, fractional inspiratory time, and breathing rate using the acoustic sensor.

2.1.3 Acoustic Sensor-based Detection and Assessment. Chauhan et al. [13] presented a breathing-based technique for user authentication using smartphone acoustic data. However, the user needs to hold the phone close to the user's nose. Previous works on acoustic-based respiratory assessment mostly involve lung obstruction or pulmonary function measurement using a smartphone microphone [20, 32, 52, 77]. Exhalesense and SpiroSmart [32, 52] used smartphone microphones to estimate lung obstruction from forced exhalation audio data. SpiroCall [20] relied on the user's exhalation sound through a call-in service on any phone using the standard telephone channel to measure pulmonary function. Although these approaches have an outstanding performance for predicting lung obstruction, they require forced effort from the user. The lack of reasonable effort from the users can result in an invalid prediction. Although we assessed the quality of the effort in our previous work [52], forced exhalation for mobile spirometry has a potential risk for flaring up the respiratory condition. On the other hand, BodyBeats [55] designed a special microphone to distinguish different non-speech body sounds. However, this approach requires the user to attach a custom device that includes a special bone conduction transducer.

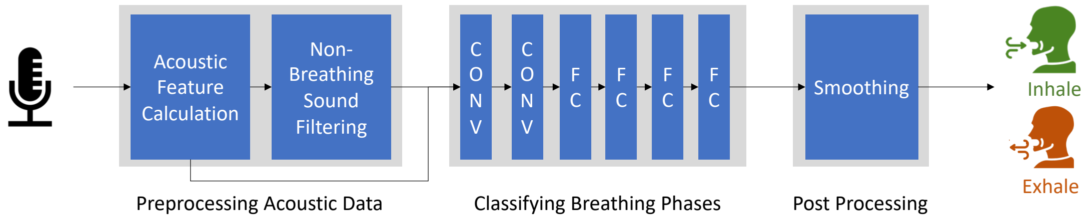

Fig. 3. Overview of BreathTrack

Contrary to the previous approaches, our approach does not need any effort from the users or any additional hardware other than a smartphone.

2.1.4 Acoustic Sensor-based Breathing Analysis for Sleep Monitoring. Ren et al. [60] presented breathing rate estimation and sleep apnea detection using smartphones and earbuds. This paper shows that the commodity mobile devices (e.g., earbuds) can capture breathing sounds without body contact. However, Their algorithms primarily focus on the overall breathing rate rather than detecting breathing phases for fine-grained breathing biomarkers. Rosenwien et al. [63] detected breathing phases (inhalation/exhalation/noise) during sleep using acoustic data captured by a directional microphone nearby the users. Nonetheless, they need manually labeled breathing phases and other acoustic noises, which is hard to achieve for regular breathing. Our work does not require labeled acoustic data and uses cross-modality knowledge transfer for automatically labeling acoustic breathing phases. While these studies show the feasibility of non-contact breathing phase detection using smartphone microphones, this paper is the first work demonstrating smartphone acoustic data for breathing phase detection during waking hours.

2.2 Teacher-Student Knowledge Transfer Models

Teacher-Student is a training method first proposed for knowledge distillation which transfers knowledge from a cumbersome DNN network (teacher) to a compressed DNN network (student) by allowing the student to learn from the teacher [28, 31]. This method allows the smaller network (student) to approximate the original function learned by the deeper network (teacher). In this technique, a compressed network (student) is trained to mimic the output of the larger network (teacher) instead of training on the raw data directly. This training technique exploits how the deeper network learns hierarchical abstraction of the features.

Later, this training method has been exploited to develop multi-task models from multiple uni-task models [37], interpret DNN models [78], and transfer knowledge from different sensing modalities [80]. This paper utilizes the teacher-student training method to transfer knowledge from the inertial sensing modality to the acoustic sensing modality. Unlike prior works that require annotated data in the teacher domain to get the pre-trained teacher model, we design a new variant of the teacher-student architecture by combining signal processing and deep learning to eliminate the need for audio annotation in the teacher domain.

3 BREATHING PHASE DETECTION MODEL

In this section, we develop a classification model to detect breathing phases – inhale or exhale – for an acoustic signal of 500ms duration. The model consists of three steps – pre-processing acoustic data, classifying breathing phases, and post-processing. We also introduce a novel variant of the teacher-student training method to train the classifier using both inertial and acoustic sensing data. Note that we only need the acoustic sensor at runtime or at testing time. This section first describes the feasibility of breathing phase detection using acoustic data, followed by the classification model's description.

First, we check the feasibility of acoustic breath phase detection. We analyze the inhale and exhale acoustic signals collected from both patient and healthy subjects to understand their acoustic signatures' variability in our collected dataset. This analysis provides us with the required observations to understand the feasibility of acoustic phase detection with the collected dataset. We further observe that the frequency distribution of inhalation and exhalation phases varies, and exhalation phases have a higher mean amplitude than inhalation phases. The average energy of inhalation and exhalation phases are 19.08 dB and 69.76 dB respectively. The signal-to-noise ratio (SNR) of inhalation and exhalation phases are -44.98 dB and -36.12 dB, respectively. The SNR and energy measurements also reflect lower amplitude in the inhalation phase. Our observations are aligned with the prior work analyzing the breathing sound [13]. From the time-domain spectrogram analysis of inhale and exhale phases in Figure 2(a), we also notice that the inhalation duration is smaller than the exhalation duration. This distinct acoustic signature allows us to classify the breathing phases. We further utilize this information along with their temporal variations during the post-processing step to improve the detection accuracy.

3.1 Pre-processing Acoustic Data

Acoustic data pre-processing includes – (1) acoustic feature calculation and (2) non-breathing sound filtering.

3.1.1 Acoustic Feature Calculation. Before extracting the feature, the acoustic signal passes through a second-order Butterworth highpass filter [69] and a Savitzky-Golay filter [68] to eliminate background noise and unwanted frequencies. Next, we segment each 1-minute filtered signal into 20 500 ms frames and extract acoustic features for each frame. Finally, we determine the optimal frame length by analytically exploring different segment durations (50 ms, 100 ms, 200 ms, and 1s), among which 500 ms performed the best. Based on prior work [70], we choose Mel Frequency Cepstral Coefficient (MFCC) [74] as our acoustic feature and calculate MFCC bands with a hop length of 256. With an audio sampling rate of 11.2KHz, this provides us a 40×22 dimensional output feature matrix. We use this feature matrix for non-breathing sound filtering and as the input vector of the breath-phase classifier. We have explored other acoustic features, e.g., log Spectrogram and a combination of MFCC and log Spectrogram, and observed similar performance.

3.1.2 Non-Breathing Sound Filtering. We develop a non-breathing sound filter that works as an admission control for the breathing sound by distinguishing breathing sound from all other sounds in the environment. These other sounds or noises include – single and multi-person speech, machine beeps, blows, knocks, paper rustles, taps, furniture movements, whispers, phone notifications, claps, throat clearing, and abdominal sounds. Unlike previous works [70] that discard any non-breathing sound, our model admits sounds where breathing is present for most of the segment. This breath detection module aims to identify the presence of noises that suppress breathing sounds. The breathing sounds that go to the breath phase detectors also contain background noises, including door closing, human speech from a distance, vacuum cleaners, car sounds from the nearby parking lot, airplane sound from nearby airports, and elevator notifications. Our model only uses the MFCC features to distinguish breathing sound from noises due to its suitability for audio event detection [1, 14, 65]. A Random Forest (RF) classifier [34] uses the principal components of the MFCC feature to differentiate breathing sound from the rest. Finally, we perform a majority voting to detect whether each 1-minute audio signal is breath or not. If the 1-minute audio is classified as breath sound, we perform the breathing phase detection; otherwise, we discard it.

3.2 Classifying Breathing Phases

3.2.1 Breath-Phase Classifier. We design a deep neural network model to classify breathing phases – inhalation and exhalation. This model consists of a Convolution Neural Network (CNN) model, which takes normalized 40×22 dimensional MFCC from the *Acoustic Feature Calculation* step as inputs. The network consists of two convolution layers with four filters of $3 \times 3 \times 1$ size and the Rectified Linear Unit (ReLU) as the activation function. To handle over-fitting, we use Batch Normalization and Dropout layers with a dropout rate of 40%. These convolution layers aim to learn the distinguishable feature representation of each acoustic data needed to detect breathing phases. We then flatten this 2-D feature representation and use three fully connected layers with 500, 200, and 50 units and ReLU activation functions. Finally, we use the Sigmoid activation function at the last fully connected layer and binary cross-entropy loss (Equation 1) to train the model. Here, p and q are the probability distribution of the true label and the predicted label, and $p \in \{y, 1-y\}$ and $q \in \{\hat{y}, 1-\hat{y}\}$. In this model, the fully connected layers aim to differentiate between inhale and exhale breathing phases.

$$H(p,q) = -\sum_i p_i \log q_i = -y \log \hat{y} - (1-y) \log(1-\hat{y}) \tag{1}$$

The biggest challenge for training the acoustic breathing phase detection is the annotation of acoustic breathing data due to the inaudible boundaries of breathing phases (Figure 2). By listening to the acoustic breathing data, we observed pauses between inhale and exhale phases. However, pause does not exist in regular breathing. This observed pause is the resultant breathing sounds in the phases' boundary, which is inaudible to human ears. For example, previous works, e.g., Breeze [70], show pause as a breath phase in controlled breathing during breathing exercises where a 2s pause (breath-holding) activity is present. However, we focus on regular breathing where breath-holding or pause is absent, but not the whole phase of the inhalation or exhalation is audible. Thus human annotation using an acoustic signal is not accurate for regular breathing. We also examine the inertial measurement unit (IMU) data and observed that IMU does not experience any pause phase, which further justifies our assumption.

Thus, we do not use the annotations using the acoustic signals for developing our model. To illustrate, we observe pause phases in both the inhale-exhale and exhale-inhale sequences in the acoustic data (Figure 2(a)). However, the breathing waveform extracted from the IMU data shows that there is no such pause in the breathing signal (Figure 2(b)). Moreover, during annotation, only 54.62% of breathing audio clips were audible. Hence, the annotator was only able to annotate 54% of the audio breathing clips.

Though acoustic signals alone might not be sufficient to train this acoustic breath phase classifier, other sensors, e.g., inertial measurement units (IMU), have shown promising results with signal-processing approaches [54]. Therefore, we aim to transfer the knowledge from the inertial modality to the auditory and deep learning models. To achieve this goal, we utilize a teacher-student training method [80], where the teacher network supervises the student network and enables the student to leverage the teacher domain's knowledge.

In this approach, the teacher model takes IMU data as input and outputs the breathing phase, which guides the student's acoustic sensing modality to classify correctly. This approach eliminates the need for acoustic data annotation and improves performance by transferring knowledge. However, IMU data are often noisy and suffer from drifting over time. Thus, unprocessed IMU data is not suitable for correct manual annotation. Besides, IMU provides six-channel data (3-axis accelerometer and 3-axis gyroscope), among which one channel represents the breathing pattern. Which channel provides the best representation depends on the orientation of the IMU. As we do not instruct the user to hold the smartphone in a specific way, any of these six channels might hold the breathing phase representation. Therefore, it is not feasible to find the best channel manually during manual annotation. Moreover, the low sampling rate of IMU does not provide enough data to train an effective deep learning model. On the other hand, the promising performance of signal processing-based approaches on IMU data for breathing applications does not require manual annotation to train the model and performs necessary filtering and smoothing to tackle noise and drift. Therefore, our solution to the short-coming of the existing teacher-student model lies in the signal processing domain.

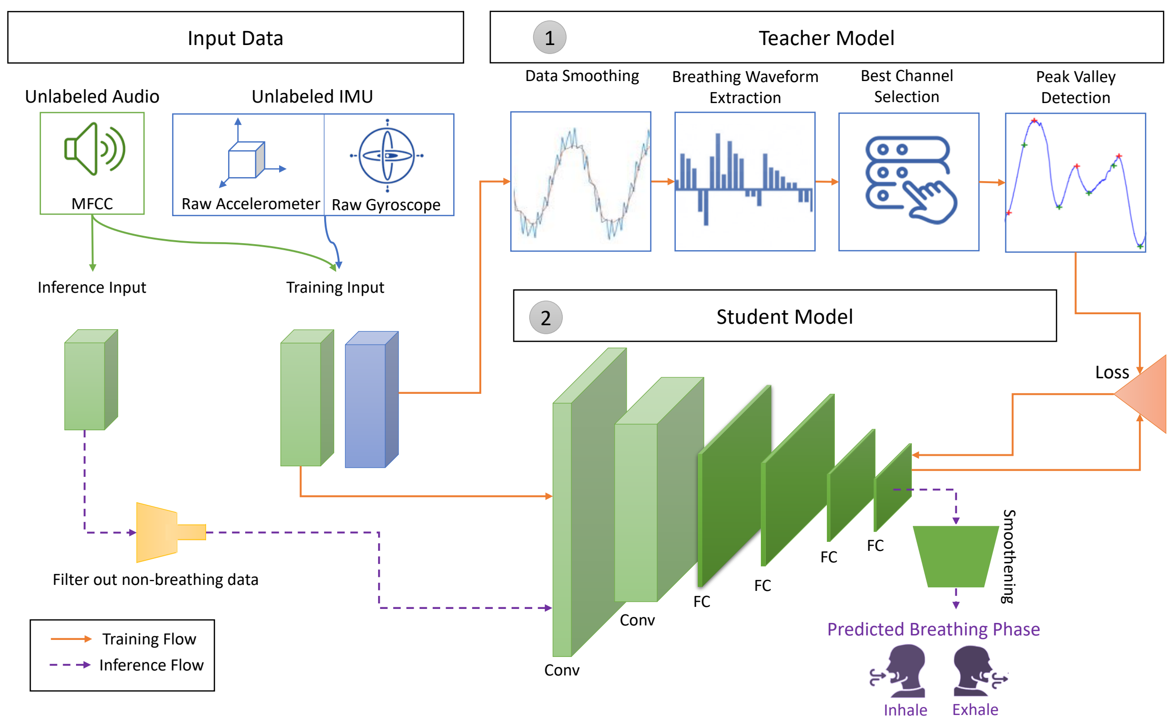

Fig. 4. Flow diagram of training acoustic breath phase-detection model (student model) with the inertial signal processing algorithm (teacher model) using the proposed variant of Teacher-Student training method. During training, the teacher model first estimates the phase labels and the student model utilizes them for the backpropagation. During inference, BreathTrack only uses the student model.

3.2.2 Proposed Variant of Teacher-Student Training Method. To address the challenges mentioned above, we propose a new variant of the teacher-student training method by replacing the teacher model's deep neural network that processes the IMU data with a signal processing-based breath phase detection algorithm (Figure 4). Unlike the existing teacher-student network [80] where the teacher modality annotation is needed, our proposed approach does not require any annotation. In this approach, the designed deep neural network model is the *Student Model*, which learns from the signal processing-based teacher model described below.

The teacher model is a signal processing model with 3-axis linear acceleration and 3-axis rotational motion signals as input from the IMU. When the phone is held against the chest, this IMU-based signal processing algorithm selects the best axis based on the maximum periodicity in the signal waveform derived by using Fourier transformation. Then it estimates the respiration cycle from the expansion and contraction of the chest while breathing by applying the Savitzky-Golay filter [68] on the selected data stream. Finally, by applying the Kalman filter as a post-processor, the algorithm accurately identifies breathing phases [54]. One comparison of the phone IMU-derived breathing phases to the chest band reference signal after time synchronization is depicted in Figure 5. More details on multi-modal breathing signal synchronization is reported in our previous works [3, 51]. While training the student model, we use these estimated breath phases to calculate the loss function, which supervises

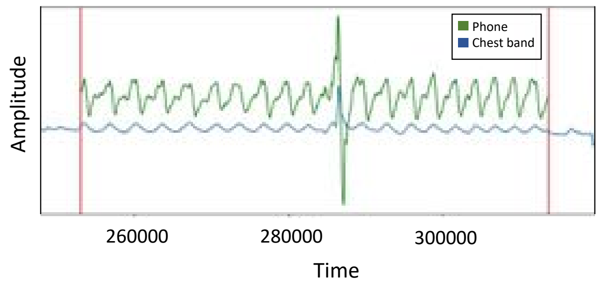

Fig. 5. Comparison between the phone sensor-based and chest band based breathing signals after time synchronization. The two red vertical lines indicated the start and end of a breathing session. The number of breaths per minute (20 BPM) is almost identical in both signals. This observation highlights the quality of the teacher model's domain knowledge to transfer the breathing phase information to the acoustic model.

the training of the student model. For example, when a 500ms segment consists of inhalation and exhalation phases, we choose the phase label that takes most of the duration. During the training of the model, this phase label from the inertial data is considered as y in Equation 1 and the predicted phase label from the acoustic signal is considered as $\hat{y}$.

3.3 Post-Processing Step.

In this step, we exploit the sequential relation and temporal characteristics of the breathing phases among the consecutive segments to the dynamic breathing phase duration. Though 500 ms segmentation benefits the breathing phases' dynamic duration, it hinders our breathing phase classifier from considering the lowest breathing phase duration. To understand how the performance can be improved, we analyze the training data to understand the minimal exhalation and inhalation phase duration and use this domain knowledge to implement a post-processing step to smooth the model's output. We observed that among the 20 estimated phase labels outputted from the breathing phase classifier, a misclassified label could often be identified by observing its neighbor labels. Therefore, we identify an anomalous inhale among exhale phases or exhale among inhale phases by observing the estimated neighbor phase labels and rectify any anomalous labeling. This smoothing reduces the anomalies in data and increases the performance by up to 4%.

4 EXPERIMENTAL SETUP

We have conducted a study to collect breathing data from pulmonary patients and healthy subjects using Samsung Galaxy Note 8 smartphones and other reference measurement devices including chest band and spirometer. In addition, we have collected audio, accelerometer, gyroscope from the phone and breathing waveform, and breathing rate using a chest band. The patients were carrying their emergency medication and inhalers for emergencies during or after the protocol, and the Institutional Review Board (IRB) approved the study.

The data collected from this study has been used in other research works, including – wheeze detection [2, 11], cough detection [46, 48], breathing rate estimation from 6-axis IMU [51], lung function estimation [46, 52], and speech analysis [45, 66, 67]. More details of the study data collection and annotation is published in mLungStudy [53]. For the completeness of the paper, we are highlighting the relevant part of the study below.

4.1 Systems and Applications

• We used two Samsung Galaxy Note 8 smartphones to collect the data - one for the participant, one for the researcher. We develop an android app for the participant's phone to continuously record sensor data, including audio with 44.1KHz 32-bit sample stereo, which is downsampled at 22.05KHz, inertial measurement unit (accelerometer, gyroscope, magnetometer) at around 100Hz. We develop another Android app for the researcher's phone to annotate the start-time and end-time of each task, collect breath counts from the participants, and take notes about each task and data collection session (the entire recording period for one participant).

• Participants wore a Bluetooth-enabled patient monitoring chest band (Zephyr Bioharness Bio-module 3) around the chest underneath their clothing to passively collect breathing reference data. We collect both breathing waveform and breathing rate as references from the chest band. We use this data to compare the breathing phases derived by the phone IMU sensor data (Figure 5) and compare the protocol breathing pattern with the out-of-protocol (passive) breathing pattern (Section 5.3).

• A portable Spirometer connected to a smartphone shows the pulmonary function test's (PFT) flow-volume curve and gives real-time data quality feedback on the connected smartphone. This device gives us an objective annotation of the subject's pulmonary condition. Based on the American Thoracic Society (ATS) and European Respiratory Society (ERS), someone has pulmonary obstructions if the lung function parameter (FEV1/FVC ratio) is less than 70% [25].

4.2 Study Description

4.2.1 Participant Recruitment. We have recruited pulmonary patients through a recruiting firm. For further confirmation, we have designed a pre-screening questionnaire that includes the following questions - '*have you ever been diagnosed with any lung condition?*', '*how severe is your asthma/COPD condition?*', '*how frequently do you use rescue inhalers?*', '*do you experience breathlessness?*', '*what medication are you taking for your pulmonary condition?*', '*how often do you cough?*'. We have considered patients who provided consistent answers on the above pre-screening questionnaires and suffered from mild, moderate, and severe pulmonary diseases. Moreover, we excluded participants with vocal chord dysfunction, severe cardiac issues, or pregnancy to reduce the non-pulmonary confounding conditions.

4.2.2 Study Protocol. This study is part of our comprehensive initiative, which aims to leverage the power of wearables and smartphones for early detection and continuous monitoring of chronic lung patients. This initiative includes quality data collection, reliable annotation, pulmonary biomarker extraction and classification, and user's privacy protection mechanisms development. We have collected regular breathing sounds, vocalization, and pulmonary symptoms such as cough, wheeze, and lung function measurements through a portable spirometer called GoSpiro. In this paper, we focus on the modeling of the regular breathing sound. We have collected regular breathing data in two positions - sitting and supine since these are the most common and comfortable postures for the subjects.

• *Sit-silent:* We asked the participants to breathe naturally at their regular pace silently in a sitting position. They have also silently counted breaths while holding the phone on the chest for one minute. Note that we do not instruct the participants to hold the phone in a specific orientation. One breath cycle consisting of an inhalation followed by an exhalation counted as one. For healthy adults, breath count varies from 12 to 20 in a minute, and for lung patients, it may go beyond 30.

• *Supine-Silent:* We instructed the participants to breathe naturally at their regular pace, silently in a supine position. They have also silently counted the breaths while holding the phone on the chest for one minute. Unlike previous works [70], we do not provide any breathing instructions, such as inhale with the nose and exhale with the mouth, to the subjects. Thus, we collect more natural data where the same person can breathe in with both nose and mouth. Besides, we do not fix the orientation of the phone for the participants. During our data collection process, we did not perform any soundproofing or use any anechoic chambers. Thus, our breathing sessions have background noises, including door closing, human speech from a distance, vacuum cleaners, car noises from the nearby parking lot, airplane sound from nearby airports, and elevator notifications. The presence of such background noises demands a robust detection model that is resilient to such sounds.

4.3 Participants

We completed several iterations through pilot studies and mock-ups to make the study rigorous in data quality and variability. Then we have collected data from 131 subjects. Among them, 91 are chronic lung patients (Male 41, Female 50), including 69 asthma patients, 9 COPD patients, and 13 with both conditions. The average age of these patients was 42.93 ± 19.49 years. Among the remaining 40 healthy subjects, 26 were male, and 14 were female. Each patient received a $150 gift card for participation. It is a minimally invasive, low-risk study, and nobody needed their emergency medicine during or after the protocol.

4.4 Patient Class Annotation

We classified our subjects into healthy or patients by exploiting the respiratory biomarkers extracted using our BreathTrack model. For this classification, we need the labels for each subject. We have two types of labels - subjective self-reported labels and objective spirometry-based labels.

4.4.1 Subjective Annotation. During the study, we provide a questionnaire to the subjects to self-report their age, pulmonary condition. We have utilized this questionnaire to decide whether a subject is a healthy or respiratory patient. Moreover, we also determine the patients' severity based on that questionnaire (e.g., 'how severe is your pulmonary condition? mild/moderate/severe?').

4.4.2 Objective Annotation. This study used an FDA[1]-approved Bluetooth Low Energy (BLE) connected, portable spirometer called GoSpiro to collect objective lung function parameters, including Forced Expiratory Volume in 1 second (FEV1%). We have collected spirometry data up to three times to choose the best effort from them as a ground truth respiratory condition. Based on the ATS/ERS guideline, we objectively label the subject as a patient if their FEV1% is below 70% [25].

4.5 Data Size

We segment each one-minute audio clip from a user into 500 ms segments, which is the input of the proposed framework. In our experiments, we have 11,618 breath audio segments and 98,841 non-breath audio segments. The breath sound detection, other sounds, and breath sound data distribution are 8.51:1, which creates a class imbalance. To address the class imbalance, we perform undersampling. Undersampling removes data examples from the training dataset that belongs to the majority class to better balance the class distribution. In our case, it reduces the skew from 8.51:1 to a 1:1 class distribution. We split the dataset into 80% training and 20% testing with 10-fold cross-validation.

We perform a subject-wise train and test split during the breath phase-detection where any data collected from the subjects in the test dataset are absent in the training dataset. We use 80% of the subjects to train the acoustic breath phase-detection model and 20% of the subjects to test the model. We have 10,494 audio segments in the training dataset and 1124 audio segments in the test dataset. The distribution of inhale and exhale in the training and test dataset are 1:1.23 and 1:1.22, respectively.

5 EXPERIMENTAL RESULTS

In this section, we evaluate the model performance using the collected data described in Section 4.

[1]Food and Drug Administration

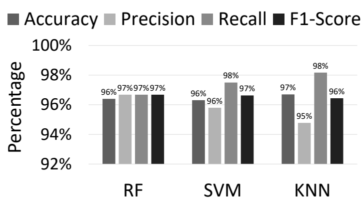

Fig. 6. Performance of the non-breathing sound filter

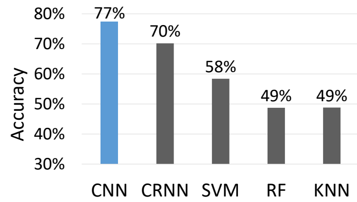

Fig. 7. Performance of different classifiers to detect breathing phases.

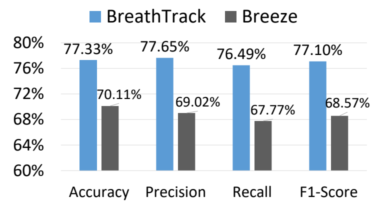

Fig. 8. Comparison of BreathTrack with Breeze

5.1 Evaluation of Non-Breathing Sound Filter

We compare the performance across three different classifiers inspired by prior works to evaluate the breath detection module's performance [60, 64, 77]. First, we choose the Support Vector Machine (SVM) as prior works achieved significant accuracy using MFCC and SVM to distinguish respiratory events from other noise sounds [60]. Next, we choose Random Forest (RF) due to its performance in obstructive sleep apnea detection from breathing sounds [64]. Finally, we compare K-Nearest Neighbour (KNN) achieved comparable performance in lung sound analysis [22]. In Figure 6, we observe that though all three classifiers: Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN), show similar accuracy (≈96%) and F1-score(≈97%); KNN and SVM have lower precision than the rest. Thus, based on our results, we choose Random Forest to distinguish breath from other sounds.

5.2 Evaluation of Breath Phase Detection

To evaluate the breath phase detection module, we first compare the proposed convolution deep network with three classic machine classifiers (SVM, RF, and KNN) previously used in breath detection from acoustic data [12, 22, 60, 64]. Next, we compared BreathTrack with a convolutional recurrent neural network (CRNN) model used in previous works [70] for controlled breathing phase detection using audio. Figure 7 shows that BreathTrack' breath phase detector achieves 17%–28% higher accuracy than the classic machine learning classifiers and 7% higher accuracy than a CRNN model. Though recurrent neural network performs better in understanding sequence, we did not observe such performance due to the lack of adequate data for each person. Later in this section, we discuss this observation in more detail.

We compare BreathTrack with a recently proposed acoustic regulated breath phase-detection model, Breeze [70], that uses a convolutional recurrent neural network model for more rigorous evaluation. However, as Breeze only focuses on regulated breathing where each breathing phase's duration is controlled, the proposed post-processing is unsuitable for uncontrolled regular breathing. Therefore, for a fair comparison, we are reporting the performance of BreathTrack before post-processing. Furthermore, unlike Breeze, we focus on regular breathing and thus do not have a breath-hold or pause phase.

Figure 8 shows that BreathTrack has 8.53%, 8.63%, and 8.72% higher F1-Score, Precision, and Recall than Breeze, respectively. Breeze is developed to monitor the breathing phase during breathing training, where each breath phase has a regulated duration (4-2-4 breath instruction). As recurrent neural network (RNN) shows superior performance in capturing this temporal relation, it performed well for the Breeze regulated dataset. However, this paper focuses on regular uncontrolled breathing where different people have different inhale-exhale duration. As we do not provide any instruction to the subjects regarding how long they should breathe in or breathe out, there are different inhale-exhale sequences in the dataset. On the other hand, we have only 20 samples for each subject, and as a result, each sequence of inhale-exhale does not have many data samples. As we know, training an RNN model requires multiple data samples of each sequence; the lack of adequate data samples for each sequence hinders the RNN from learning the sequences correctly. Thus, the CRNN model proposed by Breeze performs poorly for unregulated and uncontrolled breathing.

Moreover, during the data collection process of Breeze, subjects were given specific instructions to inhale through the nose and exhale through the mouth. Breathing through the nose and mouth has distinctive sound patterns due to structural differences in the airways. In BreathTrack, we focus on regular uncontrolled breathing where the subjects were not given any restrictions during the data collection. As a result, different people not only have different inhale-exhale duration, they also have different inhale or exhale signatures as a result of how they inhale or exhale. Furthermore, the duration of the natural breathing can widely vary from one breath to the other. Perhaps, these are the reasons for having lower accuracy from the Breeze model on our regular breathing dataset.

5.3 Comparison between Protocol Breathing and Passive Breathing Patterns

As mentioned in Section 4, we have asked our participants to breathe normally at their regular pace for one minute while holding the phone on the chest and quietly sitting down (termed here as 'sit-silent' segments) in this study. We further examine whether our protocol changed their regular breathing patterns (e.g., number of breaths per minute, namely, breathing rate). Besides passively collecting breathing rate using the Zephyr bio-harness chest band (mentioned in Section 4), we recorded the whole study session's audio. Furthermore, we have annotated quiet-sitting segments (termed as 'passive segments') outside of our protocol, e.g., the transition between the tasks. We have extracted the breathing rate from our protocol segment (sit-silent) and outside of the protocol segment (passive segment) and compare the average breathing rate. Here null hypothesis ($H0$): "there is no difference between the sit-silent breathing and passive breathing" and alternative hypothesis: "sit-silent breathing is different from passive breathing." We test our hypothesis using a two-tailed paired t-test and observe the $p - value = 0.21$. Therefore, we failed to reject the null hypothesis, indicating no differences between regular breathing in our protocol and passive breathing.

6 FINE-GRAINED RESPIRATORY BIO-MARKER EXTRACTION AND PATIENT CLASSIFICATION

Regular breathing phase detection from acoustic data has numerous use cases, including detecting respiratory patients, fitness tracking, mindfulness training, social interaction monitoring, and emotion tracking. This section demonstrates an example using passive breathing phase-detection for distinguishing respiratory patients (COPD and Asthma) from healthy subjects. We have collected subjective (self-reported) and objective (labeled using FDA-approved medical-grade devices) patient labels from the participants during the study. Using these collected data, we show the potential use case of respiratory patient detection utilizing breathing phase-based fine-grained biomarkers extracted from the acoustic data.

6.1 Patient Detection Using Breathing Phases

In this paper, along with detecting pulmonary patients, we develop detection models to detect Chronic Obstructive Pulmonary Disease (COPD) patients, asthma patients, elderly pulmonary patients, and severe patients. We consider pulmonary patients having an age of 60 or above as elderly pulmonary patients. We also distinguish patients with mild and moderate pulmonary diseases from severe pulmonary patients. There are two steps to achieve this goal – (1) respiratory biomarker estimation and (2) patient classification.

6.1.1 Respiratory Biomarker Estimation. We use the predicted breath phase labels of 1-minute audio to estimate five important biomarkers for respiratory patient assessment. These biomarkers are – fractional inspiration time, inhale-exhale ratio, inspiration duration, exhalation duration, and breathing rate.

Respiration rate is the number of breath cycles, i.e., one inhalation and one exhalation, taken in a minute. The average respiration rate for a healthy subject is 12-20 breaths per minute (BPM). Therefore, respiration rate is an early indicator for respiratory condition deterioration as it elevates during respiratory deterioration for chronic patients (e.g., asthma, COPD [10]) and infectious patients (e.g., COVID-19 [21]).

Inhalation duration is the duration of the inhale phase in a breath cycle, while exhalation duration is the exhalation phase's duration in a breath cycle. First, we calculate the mean inhalation and exhalation duration for all breath cycles in a minute. Then we calculate the mean inhale-exhale ratio overall breathing cycles for a 1-minute segment. The average inhale-exhale (I/E) ratio for a healthy subject for regular breathing is around 0.9. However, the I/E ratio can be lower when the patients have difficulty breathing (especially exhaling) or during speech or stress. Therefore, these are important markers to distinguish speech breathing [6], stressful breathing [29], and respiratory disease diagnosis [43].

Fractional inspiration time is the ratio between the inhalation duration and the breath cycle duration in a breath cycle. We estimate the mean fractional inspiration time of the 1-minute signal. Fractional inspiratory time can vary across patient conditions [71, 75]. For example, due to airway obstructions caused by inflammation or viral infection, it can be reduced compared to the patient's baseline or healthy subject.

6.1.2 Patient Classification. We utilize the estimated respiratory biomarkers to distinguish between a pulmonary patient and a healthy subject. It uses a Random Forest classifier to detect pulmonary patients. However, we face a class imbalance problem due to the severe skew in the patient class distribution (91 patients and 40 healthy subjects). If not addressed, due to this class imbalance, the trained classifier model ignores the minority class. To compensate for this imbalance, we perform undersampling with 10-fold cross-validation. It randomly removes data examples from the training dataset that belong to the majority class and balances the class distribution.

We calculate the Spearman correlation coefficient along with the p-values between the biomarkers and different patient classes. For the pulmonary patient detection, we observe that inhalation duration has the highest correlation (0.202) with the significant p-value (0.048). On the other hand, COPD patient classes have a stronger correlation with the respiration rate (-0.221, p-value: 0.032) and exhalation duration (0.203, p-value: 0.047), which matches COPD patients' facts have prolonged exhalation time [38]. On the other hand, we observe that respiration rate (0.228, p-value: 0.025) and inhalation duration (0.287, p-value: 0.005) have a stronger correlation with asthma patients. Finally, we notice that the pulmonary patient's severity is mainly related to the fractional inspiration time (-0.223, p-value: 0.028). This analysis shows the potential of pulmonary patient detection using the estimated respiratory biomarkers.

6.2 Evaluation

6.2.1 Respiratory Bio-marker Estimator. In Figure 9, we illustrate the performance of the estimated biomarkers by our model. We calculate the performance with Mean Absolute Percentage Error (MAPE) for the five biomarkers. Figure 9 shows that the inhale-exhale ratio experiences the highest error (13%) because it is affected by inhalation duration and exhalation duration. We also observe that exhalation duration has a higher error rate than the inhalation duration because inhalation and exhalation are harder to distinguish by the acoustic model, especially at the edges. Moreover, pulmonary patients (e.g., COPD, asthma) have more difficulty in exhalation due to airway obstructions and may have stacked exhalation. However, we can more accurately estimate a commonly used breathing biomarker called respiratory rate with $\approx$ 92% accuracy (error rate 8.26%). Respiration rate is one of the critical breathing biomarkers that can be very useful for healthy subjects (to measure overall fitness) and respiratory patients (e.g., COPD or COVID-19) to assess breathlessness since breathlessness often means a higher respiration rate [35]. We also more accurately estimate another novel breathing biomarker called fractional inspiratory time, which has more discriminatory power in distinguishing respiratory patients from healthy subjects and assessing the severity of the respiratory patients.

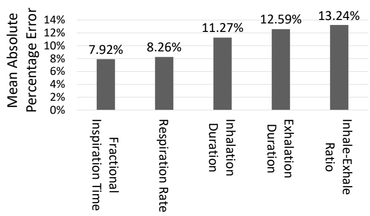

Fig. 9. Performance of the estimated Biomarkers using predicted breathing phases from BreathTrack.

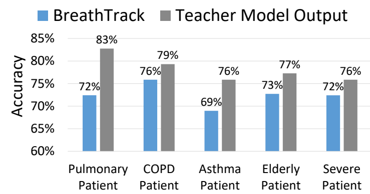

Fig. 10. Performance of patient classifier using the estimated Biomarkers from predicted breathing phase.

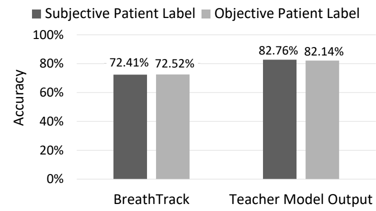

Fig. 11. Performance of pulmonary patient detection from estimated Biomarkers.

6.2.2 Patient Classification. Figure 10 shows the accuracy of the pulmonary, COPD, asthma, elderly pulmonary patient, and disease severity prediction model from biomarkers. Here, the patient class labels are self-reported by the subjects via our questionnaires, named subjective patient labels. From Figure 10, we observe that the acoustic model achieves the highest accuracy in detecting COPD patients. We also observe that model's highest accuracy drop occurs during pulmonary patient detection with 11% more error than the ground truth biomarkers. Asthma patient detection has the lowest accuracy (69%), which may occur because the asthma patients were not suffering from an attack during the data collection. If the patients are not under asthma attack or deteriorated, the difference between them and healthy subjects is often minimal. Moreover, we can observe that the COPD patient classification shows the highest accuracy and lowest gap between our model and the teacher model since COPD patients often have labored breathing symptoms. Therefore, it is likely that their breathing sounds will be more audible and more accurately be detected.

Figure 11,compares the pulmonary patient classifier's accuracy by using subjective and objective patient labels. As described in Section 4.4, subjective patient ground truths are collected using the questionnaire, and the objective patient ground truths are determined using FEV1% collected during Spirometry tests. We observe that model performance is similar for both the subjective patient label and the objective patient label proving the correctness of the self-reported patient labels.

Moreover, we analyze impurity-based importance named Gini importance of different biomarkers on detecting patients. The Gini importance of a feature is computed as the (normalized) total reduction of the criterion brought by that feature. We observe that fractional inspiration time has the highest importance in distinguishing respiratory patients, asthma patients, and severe patients. However, the inhale-exhale ratio has similar importance for severe patient detection. Furthermore, inhale-exhale ratio and exhalation duration have more significant importance than other biomarkers in identifying COPD patients and elderly respiratory patients.

Finally, we compare the patient detection accuracy with the state-of-the-art acoustic pulmonary patient detector [45]. However, unlike previous work, which focused on natural speech to detect pulmonary patients, we focus on breathing phases detected from regular breathing. As a result, our pulmonary patient classifier achieves 72% accuracy compared to 68% accuracy achieved in the state-of-the-art [45]. Moreover, using natural speech to detect patients requires the participants to read texts actively. On the other hand, our patient classifier uses passively estimated breathing phases from natural breathing.

7 DISCUSSION, LIMITATIONS, AND FUTURE WORKS

In this section, we discuss the advantages and limitations our approaches.

7.1 Training and Deployment Methodology

In Figure 12 and Figure 13, we are showing the training and expected passive deployment methodology of BreathTrack. During the training period (Figure 12), we need to collect both acoustic (microphone) and inertial

Fig. 12. Training method of BreathTrack

Fig. 13. Expected Deployment Method of BreathTrack

(IMU) data simultaneously. Therefore, the users hold the smartphone on the chest to simultaneously collect both inertial (IMU) and acoustic (microphone) data. It allows capturing the breathing sound from the mouth or nose using the microphone and the chest displacement during breathing using the accelerometer and gyroscope sensors. Using data from these two modalities, we train the breath phase-detection model as described in Section 3. Note that both IMU and microphone data are needed only for the training period. Besides, we require training only once to develop the breath phase-detection model. During our evaluation in Section 5, the participants in the train and test sets were mutually exclusive. Thus, we do not require retraining for every user.

BreathTrack only requires the acoustic (microphone) data in deployment. It eliminates the need to hold the phone on the chest, and the phone can be placed nearby the user (as shown in Figure 13). In this example use case, the user works with his computer and the smartphone resting on the table close to the user, passively collecting acoustic breathing data. In this scenario, the distance between the source and the smartphone may vary. The quality of sound captured by the smartphone microphone depends on the distance, intensity of breathing sound, and background noise [44]. From the collected dataset, we observe that breathing from a pulmonary patient is heavier than a healthy subject as patients often suffer from airway obstructions. Thus, breathing sounds will be more audible for pulmonary patients.

This paper shows the feasibility of detecting regular breathing phases using acoustic data for fine-grained breathing biomarker extraction to monitor pulmonary patients. We keep the variation in model performance due to variation in the distance as future work. Moreover, a recent study [40] presented methods to make the acoustic model device-independent. Therefore, we also envision that acoustic sensing-based BreathTrack can be modified for the audio captured by earbuds or smart glasses through microphone-agnostic transfer learning. The most significant advantage of passively sensing breathing sound using earbuds or smart glasses is that the distance between the source and the microphone is relatively fixed due to their placement on-body. Moreover, the earbud model can also utilize the inertial sensors and fuse them with microphone data for more accurate breathing phase analysis [62].

7.2 Benefits of Tracking Regular Breathing Phases

Recent studies show that continuous monitoring of health biomarkers, e.g., respiratory rate, pulse rate, blood sugar, carries significant importance for both chronic and asymptomatic patients. During the current pandemic situation due to COVID-19, the importance of continuous health monitoring is even more evident as 25%–80% of people infected with COVID-19 are unaware of being a carrier of the virus [50]. Stanford's study shows that by continuously monitoring simple biomarkers, e.g., heart rate, respiration rate, and step count, early detection of COVID-19 is possible with 85% accuracy [41].

Tracking and assessment of regular breathing can be performed either actively or passively. Active respiratory assessment requires the user to perform a series of tasks specially designed for the assessment process. These tasks include forced exhalation [52], voluntary coughing [47], script reading [8], or holding the smartphone to the chest [11]. Although active assessments are feasible for occasional spot-checks, they are not suitable for continuous monitoring, which is significant for symptomatic and asymptomatic patients.

On the other hand, passive assessment can automatically capture the breathing and other interesting sounds (e.g., speech) to derive pulmonary condition continuously. Thus, it enables more fine-grained, longitudinal analysis to detect pulmonary deterioration earlier for better treatment and management of the chronic condition. While this method does not need active participation, controlling the quality of the recordings can be challenging and poses significant privacy concerns [24].

7.3 Privacy Risk of Passive Acoustic Sensing

Privacy is one of the major concerns for any acoustic sensing system. Prior works on preserving privacy in acoustic sensing primarily focus on speech and remove background noises [30]. Though breath sound is often considered the background noise, it is the target sound in this paper. Our prior works of speech obfuscation [76] and residual removal [30] can be exploited to preserve privacy during passive acoustic sensing. For example, to preserve privacy during breath monitoring using residual removal, a two-step method can be deployed. First, we can utilize the extensive literature on voice activity detection (VAD) to extract the speech signal [42, 56, 61, 79]. Next, we can use this speech signal as the noise and remove it from the original signal with residual removal algorithms [17]. The absence of speech in the resultant signal will preserve privacy and enhance the performance of the classifiers. Moreover, in our breath detection module, we differentiate breath sounds from others, including speech, which minimizes privacy risk.

7.4 Self-awareness and Automated Assessment

When the pulmonary patients are having difficulty breathing, they will be self-aware about their pulmonary condition. However, through continuous assessment, smartphone models can quantify the intensity of the deterioration that can facilitate early or just-in-time intervention. For example, in the recent COVID-19 pandemic, millions of respiratory patients are asymptomatic, unaware of their viral infection, and actively spreading the virus. Therefore, enabling smartphones to detect viral infections in asymptomatic patients may inform the patient about the infection early and contribute significantly to society.

7.5 Assessment Accuracy and Clinical Applications

The temporal precision of our phase-detection model is in seconds with an accuracy of around 77%. When we use this model to extract breathing biomarkers in minute-level temporal precision towards clinical applications, the accuracy increases to 92%. Usually, the minimum requirement of accuracy in clinical applications varies from application to application. Our patient classification model is around 76% accurate in identifying chronic respiratory patients from healthy individuals. In contrast, the average accuracy of the rapid, consumer-grade test for infectious respiratory disease (COVID-19) is 72% for the patients with symptoms and 58% for the patients without symptoms. It is still proven to be helpful during the high demand time of the tests. Similarly, we believe that our biomarker extraction accuracy and patient classification accuracy for rapid, consumer-grade respiratory tests will be helpful to millions of chronic respiratory patients with smartphones.

8 CONCLUSION

This paper presents a novel approach for automatically annotating breathing sounds and detecting dynamically varying breathing phases using smartphone acoustic sensors. We develop a novel variant of the teacher-student model that transfers knowledge from inertial sensors into an acoustic sensing model and fuses signal processing with deep learning methods to reduce the burden of data annotation. We use detected breathing phases to estimate novel and known breathing biomarkers, including respiratory rate, inhalation-exhalation ratio, and fractional inspiratory time with up to 92% accuracy. Finally, we utilize these estimated biomarkers to determine respiratory patients with up to 76% accuracy. This paper is the first work to show the feasibility of using natural breathing sounds captured by user's smartphones to track their breathing phases, fine-grained respiratory markers, and respiratory health.

REFERENCES

[1] Ahmad Rami Abu-El-Quran, Rafik A Goubran, and Adrian DC Chan. 2006. Security Monitoring Using Microphone Arrays and Audio Classification. *IEEE Transactions on Instrumentation and Measurement* (2006).

[2] Mohsin Y Ahmed, Md Mahbubur Rahman, Viswam Nathan, Ebrahim Nemati, Korosh Vatanparvar, and Jilong Kuang. 2019. mLung: Privacy-preserving Naturally Windowed Lung Activity Detection for Pulmonary Patients. In *IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN).*

[3] Tousif Ahmed, Mohsin Y Ahmed, Md Mahbubur Rahman, Ebrahim Nemati, Bashima Islam, Korosh Vatanparvar, Viswam Nathan, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao. 2020. Automated Time Synchronization of Cough Events from Multimodal Sensors in Mobile Devices. In *ACM International Conference on Multimodal Interaction (ICMI).*

[4] Farah Q AL-Khalidi, Reza Saatchi, Derek Burke, H Elphick, and Stephen Tan. 2011. Respiration Rate Monitoring Methods: A Review. *Wiley Online Library on Pediatric Pulmonology* (2011).

[5] H. Aly and M. Youssef. 2016. Zephyr: Ubiquitous Accurate Multi-sensor Fusion-based Respiratory Rate Estimation Using Smartphones. In *IEEE International Conference on Computer Communications (INFOCOM).*

[6] Rummana Bari, Roy J Adams, Md Mahbubur Rahman, Megan Battles Parsons, Eugene H Buder, and Santosh Kumar. 2018. rConverse: Moment by Moment Conversation Detection Using a Mobile Respiration Sensor. *ACM Interactive, Mobile, Wearable and u=Ubiquitous Technologies (IMWUT)* (2018).

[7] Rummana Bari, Md Mahbubur Rahman, Nazir Saleheen, Megan Battles Parsons, Eugene H Buder, and Santosh Kumar. 2020. Automated Detection of Stressful Conversations Using Wearable Physiological and Inertial Sensors. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2020).

[8] E Blackstock, V Nathan, E Nemati, MM Rahman, K Vatanparvar, D McCaffrey, N Mendez-Lozano, BD Levy, BA Raby, and CH Fanta. 2020. Pause Duration During Scripted Speech and Cough Count as Acoustic Biomarkers for Dynamic Airflow Obstruction. In *American Thoracic Society Asthma Clinical Studies.*

[9] Biykem Bozkurt and Douglas L. Mann. 2003. Shortness of Breath. *Circulation* (2003).

[10] S Burge and JA Wedzicha. 2003. COPD Exacerbations: Definitions and Classifications. *European Respiratory Journal* (2003).

[11] 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 *ACM Conference on Human Factors in Computing Systems (CHI).*

[12] Soujanya Chatterjee, Md Mahbubur Rahman, Ebrahim Nemati, Viswam Nathan, Korosh Vatanparvar, and Jilong Kuang. 2019. mLung++: Automated Characterization of Abnormal Lung Sounds in Pulmonary Patients Using Multimodal Mobile Sensors. In *Adjunct Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing.*

[13] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing Acoustics-based User Authentication. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys).*

[14] Chloé Clavel, Thibaut Ehrette, and Gaël Richard. 2005. Events Detection for an Audio-based Surveillance System. In *IEEE International Conference on Multimedia and Expo.*

[15] Phil Corbishley and Esther Rodriguez-Villegas. 2007. Breathing Detection: Towards a Miniaturized, Wearable, Battery-operated Monitoring System. *IEEE Transactions on Biomedical Engineering* (2007).

[16] Michelle A Cretikos, Rinaldo Bellomo, Ken Hillman, Jack Chen, Simon Finfer, and Arthas Flabouris. 2008. Respiratory Rate: the Neglected Vital Sign. *The Medical journal of Australia* (2008).

[17] F Douarche, L Buisson, S Ciliberto, and A Petrosyan. 2004. A Simple Noise Subtraction Technique. *Review of scientific instruments* (2004).

[18] John F Fieselmann, Michael S Hendryx, Charles M Helms, and Douglas S Wakefield. 1993. Respiratory Rate Predicts Cardiopulmonary Arrest for Internal Medicine Inpatients. *Journal of general internal medicine* (1993).

[19] Jérémy Frey, May Grabli, Ronit Slyper, and Jessica R Cauchard. 2018. Breeze: Sharing Biofeedback Through Wearable Technologies. In *ACM Conference on Human Factors in Computing Systems (CHI).*

[20] 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 *ACM Conference on Human Factors in Computing Systems (CHI).*

[21] Trisha Greenhalgh, Gerald Choon Huat Koh, and Josip Car. 2020. Covid-19: a Remote Assessment in Primary Care. *British Medical Journal* (2020).

[22] Morten Grønnesby, Juan Carlos Aviles Solis, Einar Holsbø, Hasse Melbye, and Lars Ailo Bongo. 2017. Feature Extraction for Machine Learning Based Crackle Detection in Lung Sounds from a Health Survey. *arXiv preprint arXiv:1706.00005* (2017).

[23] Tian Hao, Chongguang Bi, Guoliang Xing, Roxane Chan, and Linlin Tu. 2017. MindfulWatch: A Smartwatch-based System for Real-time Respiration Monitoring during Meditation. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2017).

[24] Gillian R Hayes, Shwetak N Patel, Khai N Truong, Giovanni Iachello, Julie A Kientz, Rob Farmer, and Gregory D Abowd. 2004. The Personal Audio Loop: Designing a Ubiquitous Audio-based Memory Aid. In *Springer International Conference on Mobile Human-Computer Interaction*.

[25] Matthew J Hegewald, Dave S Collingridge, Thomas W DeCato, Robert L Jensen, and Alan H Morris. 2018. Airflow obstruction categorization methods and mortality. *Annals of the American Thoracic Society* (2018).

[26] Javier Hernandez, Yin Li, James M Rehg, and Rosalind W Picard. 2015. Cardiac and respiratory parameter estimation using head-mounted motion-sensitive sensors. *EAI Endorsed Transactions on Pervasive Health and Technology* (2015).

[27] Javier Hernandez, Daniel McDuff, and Rosalind W Picard. 2015. Biowatch: estimation of heart and breathing rates from wrist motions. In *IEEE International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*.

[28] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531* (2015).

[29] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*.

[30] Md Tamzeed Islam, Bashima Islam, and Shahriar Nirjon. 2017. SoundSifter: Mitigating Overhearing of Continuous Listening Devices. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*.

[31] Seung Wook Kim and Hyo-Eun Kim. 2017. Transferring Knowledge to Smaller Network with Class-distance Loss. (2017).

[32] 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 *ACM Conference on Ubiquitous Computing (UbiComp)*.

[33] Daniyal Liaqat, Mohamed Abdalla, Pegah Abed-Esfahani, Moshe Gabel, Tatiana Son, Robert Wu, Andrea Gershon, Frank Rudzicz, and Eyal De Lara. 2019. WearBreathing: Real World Respiratory Rate Monitoring Using Smartwatches. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2019).

[34] Andy Liaw, Matthew Wiener, et al. 2002. Classification and Regression by RandomForest. *R news* (2002).

[35] Libretexts. 2020. Breathing Patterns. https://med.libretexts.org/Bookshelves/Anatomy_and_Physiology/Book:_Anatomy_and_Physiology_(Boundless)/21:_Respiratory_System/21.5:_Mechanics_of_Breathing/21.5D:_Breathing_Patterns

[36] Haipeng Liu, John Allen, Dingchang Zheng, and Fei Chen. 2019. Recent development of respiratory rate measurement technologies. *Physiological measurement* 40, 7 (aug 2019), 07TR01. https://doi.org/10.1088/1361-6579/ab299e

[37] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Improving multi-task deep neural networks via knowledge distillation for natural language understanding. *arXiv preprint arXiv:1904.09482* (2019).

[38] Stephen H Loring, Mauricio Garcia-Jacques, and Atul Malhotra. 2009. Pulmonary characteristics in COPD and mechanisms of increased work of breathing. *Journal of applied physiology* (2009).

[39] L Pekka Malmberg, Leena Pesu, and AR Sovijärvi. 1995. Significant Differences in Flow Standardised Breath Sound Spectra in Patients with Chronic Obstructive Pulmonary Disease, Stable Asthma, and Healthy Lungs. *BMJ Thorax* (1995).

[40] Akhil Mathur, Anton Isopoussu, Fahim Kawsar, Nadia Berthouze, and Nicholas D Lane. 2019. Mic2mic: Using Cycle-consistent Generative Adversarial Networks to Overcome Microphone Variability in Speech Systems. In *IEEE International Conference on Information Processing in Sensor Networks (IPSN)*.

[41] Tejaswini Mishra, Meng Wang, Ahmed A Metwally, Gireesh K Bogu, Andrew W Brooks, Amir Bahmani, Arash Alavi, Alessandra Celli, Emily Higgs, Orit Dagan-Rosenfeld, et al. 2020. Early Detection Of COVID-19 Using A Smartwatch. *medRxiv* (2020).

[42] Mohammad Hossein Moattar and Mohammad M Homayounpour. 2009. A Simple but Efficient Real-time Voice Activity Detection Algorithm. In *IEEE European Signal Processing Conference*.

[43] Hadi Moghadas-Dastjerdi, Mohammadreza Ahmadzadeh, and Abbas Samani. 2017. Towards Computer Based Lung Disease Diagnosis Using Accurate Lung Air Segmentation of CT Images in Exhalation and Inhalation Phases. *Elsevier Expert Systems with Applications* (2017).

[44] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless Sleep Apnea Detection on Smartphones. In *ACM international conference on Mobile Systems, Applications, and Services (MobiSys)*.

[45] Viswam Nathan, Korosh Vatanparvar, Md Mahbubur Rahman, Ebrahim Nemati, and Jilong Kuang. 2019. Assessment of Chronic Pulmonary Disease Patients Using Biomarkers from Natural Speech Recorded by Mobile Devices. In *IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN)*.

[46] 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 *IEEE Engineering in Medicine & Biology Society (EMBC)*.

[47] Ebrahim Nemati, Md J Rahman, Erin Blackstock, Viswam Nathan, Md. M Rahman, Korosh Vatanparvar, and Jilong Kuang. 2020. Estimation of the Lung Function Using Acoustic Features of the Voluntary Cough. In *IEEE Engineering in Medicine and Biology Society*

(EMBC).

[48] Ebrahim Nemati, Md Mahbubur Rahman, Viswam Nathan, Korosh Vatanparvar, and Jilong Kuang. 2020. A Comprehensive Approach for Classification of the Cough Type. In *IEEE Engineering in Medicine & Biology Society (EMBC)*.

[49] Kurt Plarre, Andrew Raij, Syed Monowar Hossain, Amin Ahsan Ali, Motohiro Nakajima, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Santosh Kumar, Marcia Scott, et al. 2011. Continuous Inference of Psychological Stress from Sensory Measurements Collected in the Natural Environment. In *IEEE international conference on Information Processing in Sensor Networks*.

[50] Roz Plater. 2021. 50 Percent of People with COVID-19 Aren't Aware They Have Virus. https://www.healthline.com/health-news/50-percent-of-people-with-covid19-not-aware-have-virus

[51] Md Mahbubur Rahman, Mohsin Yusuf Ahmed, Tousif Ahmed, Bashima Islam, Viswam Nathan, Korosh Vatanparvar, Ebrahim Nemati, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao. 2020. BreathEasy: Assessing Respiratory Diseases Using Mobile Multimodal Sensors. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 41–49.

[52] Md Mahbubur Rahman, Tousif Ahmed, Ebrahim Nemati, Viswam Nathan, Korosh Vatanparvar, Erin Blackstock, and Jilong Kuang. 2020. ExhaleSense: Detecting High Fidelity Forced Exhalations to Estimate Lung Obstruction on Smartphones. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*.

[53] Md Mahbubur Rahman, Viswam Nathan, Ebrahim Nemati, Korosh Vatanparvar, Mohsin Ahmed, and Jilong Kuang. 2019. Towards Reliable Data Collection and Annotation to Extract Pulmonary Digital Biomarkers Using Mobile Sensors. In *EAI International Conference on Pervasive Computing Technologies for Healthcare*.

[54] Md Mahbubur Rahman, Ebrahim Nemati, Viswam Nathan, and Jilong Kuang. 2018. InstantRR: Instantaneous Respiratory Rate Estimation on Context-aware Mobile Devices. In *EAI International Conference on Body Area Networks*.

[55] Tauhidur Rahman, Alexander Travis Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. 2014. BodyBeat: a Mobile System for Sensing Non-speech Body Sounds. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*.

[56] Javier Ramırez, José C Segura, Carmen Benıtez, Angel De La Torre, and Antonio Rubio. 2004. Efficient Voice Activity Detection Algorithms Using Long-term Speech Information. *Elsevier Speech communication* (2004).

[57] Ruth Ravichandran, Elliot Saba, Ke-Yu Chen, Mayank Goel, Sidhant Gupta, and Shwetak N Patel. 2015. WiBreathe: Estimating Respiration Rate Using Wireless Signals in Natural Settings in the Home. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*.

[58] Tuomas Reinvuo, Manne Hannula, Hannu Sorvoja, Esko Alasaarela, and R Myllyla. 2006. Measurement of Respiratory Rate with High-resolution Accelerometer and EMFit Pressure Sensor. In *IEEE Sensors Applications Symposium*.

[59] Yanzhi Ren, Chen Wang, Yingying Chen, and Jie Yang. 2014. Poster: Hearing Your Breathing: Fine-grained Sleep Monitoring Using Smartphones. In *ACM International conference on Mobile Computing and Networking*.

[60] Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen. 2015. Fine-grained Sleep Monitoring: Hearing Your Breathing with Smartphones. In *IEEE Conference on Computer Communications (INFOCOM)*.

[61] Philippe Renevey and Andrzej Drygajlo. 2001. Entropy Based Voice activity Detection in Very Noisy Conditions. In *European Conference on Speech Communication and Technology*.

[62] Tobias Röddiger, Daniel Wolffram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards Respiration Rate Monitoring Using an In-ear Headphone Inertial Measurement Unit. In *ACM International Workshop on Earable Computing (EarComp)*.

[63] T Rosenwein, Eliran Dafna, Ariel Tarasiuk, and Yaniv Zigel. 2014. Detection of Breathing Sounds during Sleep Using Non-contact Audio Recordings. In *IEEE International Conference of Engineering in Medicine and Biology Society*.

[64] T Rosenwein, Eliran Dafna, Ariel Tarasiuk, and Yaniv Zigel. 2015. Breath-by-breath Detection of Apneic Events for OSA Severity Estimation Using Non-contact Audio Recordings. In *IEEE Engineering in Medicine and Biology Society (EMBC)*.

[65] J-L Rouas, Jérôme Louradour, and Sébastien Ambellouis. 2006. Audio Events Eetection in Public Transport Vehicle. In *IEEE Intelligent Transportation Systems Conference*.

[66] 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 *ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*.

[67] 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 *IEEE International Conference on Pervasive Computing and Communications (PerCom)*.

[68] Ronald W Schafer. 2011. What is a Savitzky-Golay Filter?[lecture notes]. *IEEE Signal Processing Magazine* (2011).

[69] Ivan W Selesnick and C Sidney Burrus. 1998. Generalized Digital Butterworth Filter Design. *IEEE Transactions on signal processing* (1998).

[70] Chen-Hsuan Shih, Naofumi Tomita, Yanick X Lukic, Álvaro Hernández Reguera, Elgar Fleisch, and Tobias Kowatsch. 2019. Breeze: Smartphone-based Acoustic Real-time Detection of Breathing Phases for a Gamified Biofeedback Breathing Training. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2019).

[71] Urmila Shivaram, Joseph Donath, Faroque A Khan, and Judith Juliano. 1987. Effects of Continuous Positive Airway Pressure in Acute Asthma. *Karger Respiration* (1987).

[72] Ranganatha Sitaram, Tomas Ros, Luke Stoeckel, Sven Haller, Frank Scharnowski, Jarrod Lewis-Peacock, Nikolaus Weiskopf, Maria Laura Blefari, Mohit Rana, Ethan Oblak, et al. 2017. Closed-loop Brain Training: the Science of Neurofeedback. *Nature Reviews Neuroscience* (2017).

[73] Xiao Sun, Li Qiu, Yibo Wu, Yeming Tang, and Guohong Cao. 2017. Sleepmonitor: Monitoring Respiratory Rate and Body Position During Sleep Using Smartwatch. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2017).

[74] Vibha Tiwari. 2010. MFCC and Its Applications in Speaker Recognition. *Citeseer International Journal on Emerging Technologies* (2010).

[75] MARTIN J Tobin, M JEFFREY Mador, SUSAN M Guenther, ROBERT F Lodato, and MARVIN A Sackner. 1988. Variability of Resting Respiratory Drive and Timing in Healthy Subjects. *Journal of Applied Physiology* (1988).

[76] Korosh Vatanparvar, Viswam Nathan, Ebrahim Nemati, Md Mahbubur Rahman, and Jilong Kuang. 2019. A Generative Model for Speech Segmentation and Obfuscation for Remote Health Monitoring. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*.

[77] Varun Viswanath, Jake Garrison, and Shwetak Patel. 2018. SpiroConfidence: Determining the Validity of Smartphone Based Spirometry Using Machine Learning. In *IEEE Engineering in Medicine and Biology Society (EMBC)*.

[78] Junpeng Wang, Liang Gou, Wei Zhang, Hao Yang, and Han-Wei Shen. 2019. Deepvid: Deep Visual Interpretation and Diagnosis for Image Classifiers via Knowledge Distillation. *IEEE transactions on visualization and computer graphics* (2019).

[79] Kyoung-Ho Woo, Tae-Young Yang, Kun-Jung Park, and Chungyong Lee. 2000. Robust Voice Activity Detection Algorithm for Estimating Noise Spectrum. *Electronics Letters* (2000).

[80] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7356–7365.