DeepBreath: Breathing Exercise Assessment with a Depth Camera

WENTAO XIE†, The Hong Kong University of Science and Technology, Hong Kong SAR
CHI XU†, The Hong Kong University of Science and Technology, Hong Kong SAR
YANBIN GONG, The Hong Kong University of Science and Technology, Hong Kong SAR
YU WANG and YUXIN LIU, The First Affiliated Hospital of Guangzhou Medical University, China
JIN ZHANG, Southern University of Science and Technology, China
QIAN ZHANG, The Hong Kong University of Science and Technology, Hong Kong SAR
ZEGUANG ZHENG, The First Affiliated Hospital of Guangzhou Medical University, China
SHIFANG YANG, Guangdong Provincial People's Hospital and Southern Medical University, China

Practicing breathing exercises is crucial for patients with chronic obstructive pulmonary disease (COPD) to enhance lung function. Breathing mode (chest or belly breathing) and lung volume are two important metrics for supervising breathing exercises. Previous works propose that these metrics can be sensed separately in a contactless way, but they are impractical with unrealistic assumptions such as distinguishable chest and belly breathing patterns, the requirement of calibration, and the absence of body motions. In response, this research proposes DeepBreath, a novel depth camera-based breathing exercise assessment system, to overcome the limitations of the existing methods. DeepBreath, for the first time, considers breathing mode and lung volume as two correlated measurements and estimates them cooperatively with a multitask learning framework. This design boosts the performance of breathing mode classification. To achieve calibration-free lung volume measurement, DeepBreath uses a data-driven approach with a novel UNet-based deep-learning model to achieve one-model-fit-all lung volume estimation, and it is designed with a lightweight silhouette segmentation model with knowledge transferred from a state-of-the-art large segmentation model that enhances the estimation performance. In addition, DeepBreath is designed to be resilient to involuntary motion artifacts with a temporal-aware body motion compensation algorithm. We collaborate with a clinical center and conduct experiments with 22 healthy subjects and 14 COPD patients to evaluate DeepBreath. The experimental result shows that DeepBreath can achieve high breathing metrics estimation accuracy but with a much more realistic setup compared with previous works.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Applied computing → Health informatics**.

Additional Key Words and Phrases: breathing exercises, breathing mode, lung volume, depth camera, multitask learning

† Wentao Xie and Chi Xu are the co-first authors.
This work was done when Wentao Xie was also affiliated with Southern University of Science and Technology.

Authors' Contact Information: Wentao Xie, wxieaj@cse.ust.hk, CSE, The Hong Kong University of Science and Technology, Hong Kong SAR; Chi Xu, cxubs@cse.ust.hk, CSE, The Hong Kong University of Science and Technology, Hong Kong SAR; Yanbin Gong, ygongae@cse.ust.hk, CSE, The Hong Kong University of Science and Technology, Hong Kong SAR; Yu Wang; Yuxin Liu, State Key Lab of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, China; Jin Zhang, zhang.j4@sustech.edu.cn, RITAS, CSE, Southern University of Science and Technology, China; Qian Zhang, qianzh@cse.ust.hk, CSE, The Hong Kong University of Science and Technology, Hong Kong SAR; Zeguang Zheng, State Key Lab of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, China; Shifang Yang, yangshifang@gdph.org.cn, Department of Pulmonary and Critical Care Medicine, Guangdong Provincial People's Hospital and Southern Medical University, China.

ACM 2474-9567/2024/9-ART137
https://doi.org/10.1145/3678519

ACM Reference Format:
Wentao Xie, Chi Xu, Yanbin Gong, Yu Wang, Yuxin Liu, Jin Zhang, Qian Zhang, Zeguang Zheng, and Shifang Yang. 2024. DeepBreath: Breathing Exercise Assessment with a Depth Camera. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 137 (September 2024), 26 pages. https://doi.org/10.1145/3678519

1 Introduction

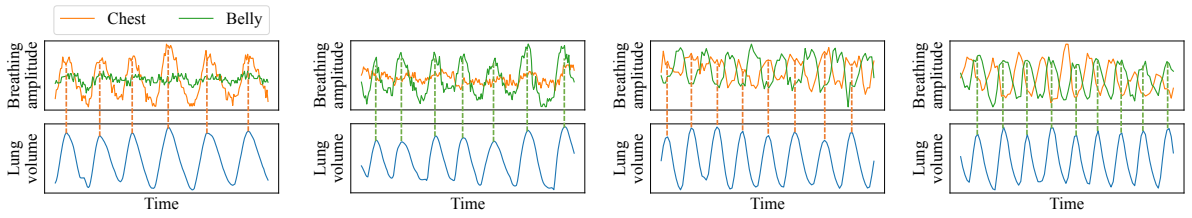
Practicing breathing exercises is essential for chronic obstructive pulmonary disease (COPD) patients [1, 22, 46]. A typical breathing exercise involves the patient breathing followed by a designated pattern, which is characterized by breathing metrics including breathing mode (belly or chest breathing) and lung volume. During a breathing exercise session, these two breathing metrics must be carefully measured as they reflect the quality of the exercise. For example, a patient is suggested to use belly breathing [17, 46], which is a measurement of breathing mode, and the patient should take deep breaths during training [74, 75], which is a measurement of lung volume.

Medical guidelines suggest pulmonary disease patients should practice breathing exercises daily [1, 2], and the ideal venues for breathing exercises are hospitals and rehabilitation centers where therapists can supervise the training exercises. However, due to time and budget constraints, the authorities recommend that patients can also take home-based training programs where the patients can practice breathing training at home [26]. Nevertheless, the effectiveness of home-based breathing exercises is not as good as the ones conducted under the supervision of therapists due to low adherence [37], which is because of the lack of supervision and feedback, where the patients have no idea about their performance.

To address this issue, researchers have developed various types of systems to assess these two breathing metrics to help patients conduct breathing exercises at home. To estimate the breathing mode, wearable sensors [5] and ultrasound sensors [23] are leveraged to measure the different heaving patterns at the chest and belly to infer the breathing mode. These designs are based on the assumption that, during chest breathing, the heaving amplitude of the chest is larger than that of the belly and vice versa. To estimate the lung volume, depth camera-based solutions are widely explored [29, 32, 47, 49, 55, 57–59, 61]. The rationale behind these systems is that a depth camera can probe the fine-grained torso surface dynamics caused by breathing. The measured torso surface dynamics can be transformed into lung volume by using a regression function for each subject.

Although promising, these designs cannot be used for daily breathing exercise assessment because of the following limitations. **Firstly**, previous breathing mode classification methods are only tested on healthy subjects [23] and may not work on pulmonary disease patients. This is because previous works are based on the assumption that breathing mode can be easily inferred by comparing the amplitudes of the chest and belly's heaving patterns (Figures 1(a)-1(b)). But this is not always true for patients whose breathing pattern can be complex and inconsistent, where the chest and belly can have indistinguishable heaving amplitudes regardless of the breathing mode (Figures 1(c)-1(d)) due to lack of practice and lung function deficiencies. **Secondly**, most previous lung volume estimation methods require calibration with a clinical spirometer before use [32, 49, 55, 57–59]. Otherwise, the subjects need to be bare-chested [29, 47]. However, this is an impractical requirement for home usage because spirometers are usually available in clinics and hospitals, and users would prefer normal clothing when conducting breathing exercises. **Thirdly**, all the previous noncontact sensor-based works are vulnerable to involuntary body motions. This is because these works are based on distance measurements, and involuntary body motions - motion artifacts (MAs) - can easily overwhelm the breathing dynamics. Therefore, these works require the users to be seated in a chair with their back firmly against the back of the chair to restrict natural body motions that happen with breathing. However, this is not the best setup for breathing exercises because, as discussed in Section 2.1. It is more appropriate to let the patient sit in a backless chair where involuntary MAs are inevitable.

In this research, our objective is to design a practical breathing exercise assessment system that can overcome the above limitations. Specifically, the designed system should be applicable to pulmonary patients' complex breathing patterns, be calibration-free, and be robust to involuntary MAs. Inspired by previous research, we opt to



(a) Chest breathing (b) Belly breathing (c) Chest breathing (d) Belly breathing

Fig. 1. Breathing patterns under different breathing modes. (a)-(b) Healthy subjects. A(chest) > A(belly) under chest breathing, and A(chest) < A(belly) under belly breathing. (c)-(d) Patients. A(chest) and A(belly) can have similar values but different phases.

design our system based on depth cameras because of their ability to capture fine-grained torso surface dynamics, which are highly related to breathing metrics. In this way, to use our design, a pulmonary disease patient just needs to sit casually in front of a depth camera with their daily clothing and conduct breathing exercises while our ready-to-use system monitors the breathing metrics and gives feedback to the patient. However, even with a high-resolution depth camera, achieving such a design is challenging. **Firstly**, it is hard to predict the patient's actual breathing mode if the breathing patterns at the chest and belly are similar. This is because by observing these similar breathing patterns with a depth camera, it is almost impossible to identify which part, the chest or belly, is actually contributing to breathing and which part is just compensation. **Secondly**, measuring lung volume without calibration is challenging. This is because the relationship between lung volume and torso surface dynamics differs for every subject due to various human factors such as age, gender, and health status. **Thirdly**, it is hard to estimate the breathing mode and lung volume with the interference of involuntary MAs, where MA-induced distance changes can easily overwhelm the one caused by breathing-induced torso surface dynamics.

To overcome the above challenges, we present DeepBreath, the first depth camera-based breathing exercise assessment system that can practically measure breathing mode and lung volume for pulmonary patients. Specifically, we have the following designs to solve the above challenges. **Firstly**, we observe that all the previous works regard breathing mode and lung volume as two independent measurements and estimate them separately. However, these two measurements are actually correlated with each other. For example, even though the breathing patterns at the chest and belly can have indistinguishable amplitudes at a certain breathing mode, the true breathing region (belly or chest) is always more correlated to the actual lung volume, as shown in Figures 1(c)-1(d). Therefore, to overcome the challenge of predicting one's breathing mode under similar chest and belly heaving amplitude, we can take the current lung volume as prior knowledge to assist in deciding the current breathing mode. In light of this, we design a multitask learning (MTL) framework to let breathing mode classification and lung volume estimation work cooperatively, where the breathing mode classifier can leverage lung volume information from its counterpart. **Secondly**, to enable calibration-free lung volume estimation, we use a data-driven approach and design a powerful UNet-based [51] lung volume regressor to achieve one-model-fit-all lung volume estimation. Also, to let the model consider human factors when predicting lung volume, we design a lightweight segmentation model to segment the human silhouette from the raw depth image and feed the segmented image to the lung volume regressor. The segmented silhouette implicitly encodes human factors such as weight, height, etc. Notably, when designing the segmentation model, we distill the segmentation power from a large, RGB image-based segmentation model - Segment Anything [34] - to our small and depth image-based model to achieve accurate and fast segmentation. **Thirdly**, to solve the challenge of involuntary MAs, we design

a temporal-aware depth image processing algorithm to compensate for involuntary body motions based on an observation that involuntary MAs are mostly slower than breathing.

We collaborate with a medical center to evaluate the performance of DeepBreath. We recruit 36 subjects (26 males, 10 females), including 14 mild to moderate COPD patients (11 males, 3 females) in the experiments. The evaluation with leave-one-subject-out (LOSO) cross-validation shows that DeepBreath can achieve an accurate breathing exercise assessment with a mean breathing mode classification F1-score of 0.92, a mean absolute lung volume error of 0.09L, and a mean breathing rate error of 0.09 BPM.

We summarize the main contributions of this work as follows. First, we propose DeepBreath, the first depth camera-based solution to practically measure breathing mode and lung volume for pulmonary disease patients, with no impractical assumptions such as consistent chest/belly breathing pattern, per-user calibration, no clothing, and no involuntary body motion. Secondly, we propose a series of techniques that overcome the limitations faced by previous works, including an MTL framework to bypass the complex breathing patterns when predicting breathing mode, a novel UNet-based model to achieve one-model-fit-all lung volume regression, a lightweight, depth image-based human silhouette segmentation model with knowledge distilled from a state-of-the-art large model, a temporal-aware depth image processing algorithm to compensate for the involuntary motion artifacts. Lastly, we collaborate with a medical center to evaluate DeepBreath. The result shows that DeepBreath can accurately estimate the two breathing metrics - breathing mode and lung volume - with low delay.

2 Background

Before we dive into the details of our system design, we first discuss the general background of this research. In this section, we first briefly introduce the background of breathing exercises from a medical perspective. Then, we discuss the reason why depth cameras can be leveraged to sense breathing and other human activities.

2.1 Breathing Exercise Assessment

Breathing exercises improve patients' respiratory capacity, strengthen their breathing muscles, and enhance their overall quality of life [31]. Typically, pulmonary therapists guide patients to engage in slow and deep belly breathing. During the exercise, patients are instructed to place their hands on their chest and belly for "tactile stimulation" [52]. Also, a common clinical practice requires the patients to sit upright with no back support. In this way, the patients must support the weight of their upper body on their own. This is to help the patient better train the respiratory muscles as well as other muscles associated with breathing. The ideal breathing technique is belly breathing, involving slow inhalation, allowing the belly to expand while keeping the chest still. Exhalation is performed by gradually contracting the belly [18]. During belly breathing, it can be observed that the hand placed on the belly exhibits regular upward and downward movements synchronized with the breathing cycle, while the hand placed on the chest remains relatively still [71]. Also, therapists would suggest the patient inhale maximally to reach their inspiratory capacity (IC) to better train the inspiratory muscles.

However, individuals with COPD may have rather complex and abnormal breathing patterns due to limited lung function and lack of practice, where engaging in chest or belly breathing may unintendedly cause the motions of its counterpart. In this case, observing the chest and belly moving amplitude is not enough for therapists to decide the current breathing mode. Instead, they also need to consider the lung volume, more specifically, inhalation or exhalation, to see which of these moving patterns truly contribute to breathing to identify the true breathing mode, as shown in Figures 1(c)-1(d).

2.2 Potentials of Depth Camera for Human Sensing

Depth cameras have gained significant popularity in healthcare research [10]. With advancements in hardware technology, depth cameras are capable of providing increasingly accurate depth measurements. Microsoft Azure

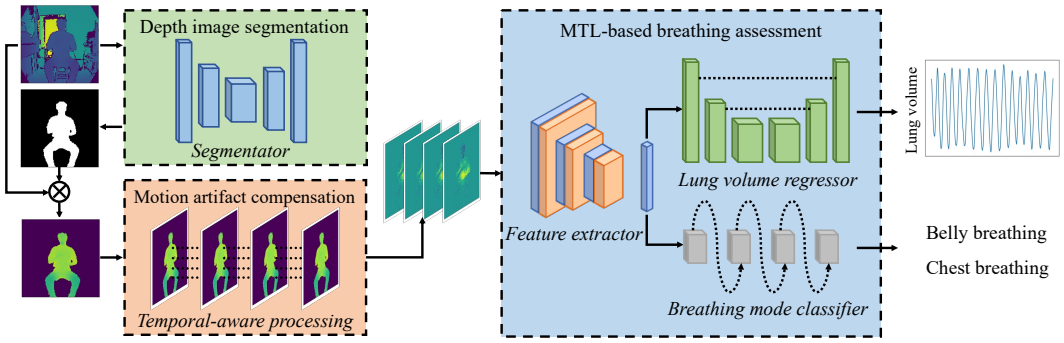


Fig. 2. DeepBreath overview.

Kinect [3], released in 2020, exhibits a system error of less than 2 *mm* within a 2-meter sensing range [36, 62]. The Azure Kinect is based on the Continuous Wave Time-of-Flight (CW ToF) technology. The CW ToF imaging system stands out for its remarkable mechanical robustness, high resolution, and low computational costs. In this technology, the light emitted from a modulated light source scatters upon encountering objects within the camera's field of view. By measuring the phase delay of the amplitude envelope between the emitted and reflected light, distance values are obtained for each pixel in the imaging array [41]. As a result of this mechanism, CW ToF technology achieves exceptional precision, making it well-suited for capturing subtle physiological changes and body dynamics.

3 System Design

This section elaborates on the software design of DeepBreath. The design contains three modules that operate sequentially - a depth image segmentation module, a motion artifact compensation module, and an MTL-based breathing assessment module. The system diagram of DeepBreath is shown in Figure 2. In the depth image segmentation module, we use a knowledge distillation method to design an accurate and lightweight segmentation model that discards the background of a depth image and only preserves the human subject's silhouette. In the motion artifact compensation module, we design a temporal-aware depth image processing algorithm to cancel out the influence of involuntary body motions. In the MTL-based breathing assessment module, we design an MTL framework to fully leverage the underlying correlation between the measurements of breathing mode and lung volume and measure these two concurrently. This design boosts the accuracy of breathing mode classification. The MTL framework consists of a shared feature extractor, a breathing mode classifier, and a novel UNet-based lung volume regressor to achieve calibration-free lung volume measurement. Note that the original depth image stream from the camera has a dimension of 576×640 and a frame rate of 15 fps. We resize it to 128×128 to reduce the model size, and we downsample it to 3 fps because it is enough to capture breathing, which usually has a frequency of no more than 0.5 Hz.

3.1 Depth Image Segmentation

The human subject in the captured depth image is mixed with the background, and it can be at any location in the depth image due to different sitting positions. Therefore, we first design a segmentation method to segment the human subject from the depth image. This would also benefit the subsequent breathing measurement tasks because the segmented human silhouette implicitly encodes human factors such as height and weight, which could be considered in measuring breathing metrics. However, the state-of-the-art segmentation models, *e.g.*,

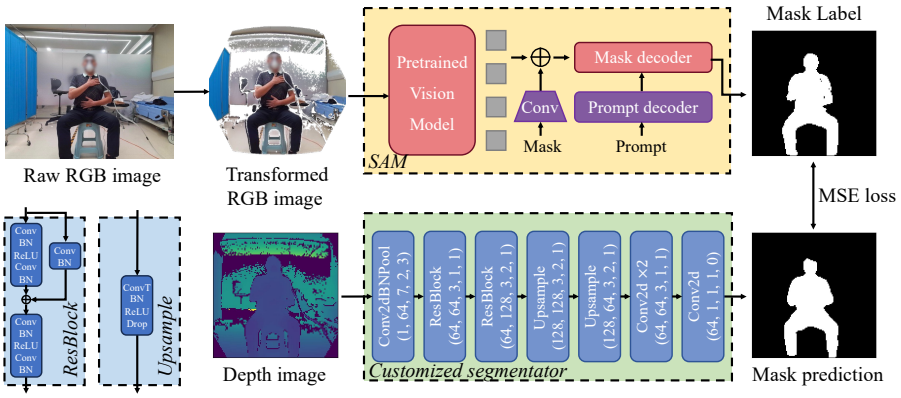


Fig. 3. SAM-supervised segmentator training.

Segment Anything Model (SAM) [34], are extremely large and run slowly on consumer-level computers. This is unsuitable for our task since the patients expect timely feedback. Also, most existing image segmentation methods are designed for RGB images, while in DeepBreath, we are dealing with depth images. In this module, we apply a SAM-supervised training scheme that leverages the power of SAM to train our lightweight, customized segmentation model that can segment a human subject from depth images. This design achieves a high-accuracy segmentation with a much faster processing speed, as discussed in Section 5.2.4. The scheme is shown in Figure 3. For more details on SAM, please refer to [34].

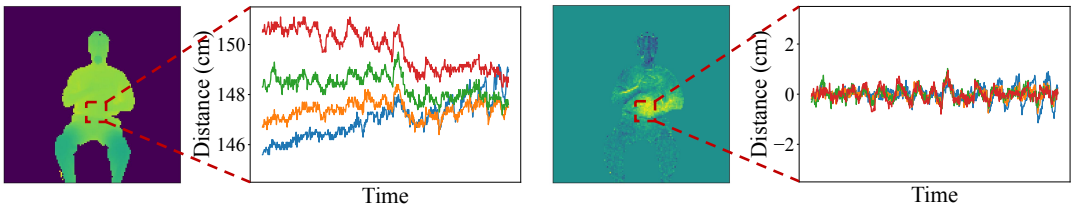
We configure our depth camera to take RGB and depth images simultaneously. Note that this is just for training the segmentator in the training phase. During the inference phase, DeepBreath only requires depth images. Due to the different perspective angles of the depth and RGB cameras, we first apply a perspective transformation to align the RGB images with the depth images. Then, the transformed RGB image is fed to SAM to generate the RGB-based mask. The produced mask at this stage will be leveraged as the ground truth for training the customized model.

The customized segmentation model adopts an autoencoder structure with only two downsampling blocks and two upsampling blocks. To further ease the training burden, we directly leveraged the pre-trained ResNet-18 [30] blocks as the encoder to reduce the training burden. The decoder contains two upsampling blocks in which a transposed convolution layer is used for upsampling.

In the inference stage, the system takes the depth image as the input and outputs the corresponding image mask. The mask is then multiplied with the depth image to extract only measurements on the human subject. After segmentation, we align the image so that the segmented human subject is in the middle of the image and crop the image on the x-axis so that the image has a dimension of 128×80 to discard irrelevant information. Notably, this dimension is an empirical pick that can cover the size of all the subjects in our dataset.

3.2 Motion Artifact Compensation

As discussed in Section 2.1, a patient is suggested to sit in a backless chair when conducting breathing exercises. This way, the patient may move the upper body unconsciously. This could easily overwhelm the distance readings of the depth images since body motions are much larger than breathing-induced distance changes. As we will show in Section 5.4, leaving these body motions in the data could largely harm the system's performance. Therefore, unconscious body motions - involuntary motion artifacts (MAs) - must be removed from the captured depth images before passing the depth images to the following breathing metrics estimation modules. Fortunately, we



(a) With MAs. (b) Without MAs.

Fig. 4. MA compensation. The four curves represent four different distance readings in the red box.

Algorithm 1: Motion Artifacts Removal

Input: M: Stacked depth images with size $N \times W \times H$; ω: window size for averaging; b: number of samples in a batch. α: factor to compute the threshold.

Output: M: The stacked images after motion artifacts removal.

$M_A \leftarrow \text{ZEROS}(N, W, H)$ // Store the averaged depth images.

$F \leftarrow \text{ONES}(N, W, H)$ // Store the flag bits for which locations need to be discarded.

for $n = 1$ *to* N **do**

  $v \leftarrow \text{MEDIAN}(\|M[n] - M[b \cdot \lfloor n/b \rfloor]\|)$ // Estimate the median variation (zeros are ignored).

  $w, h \leftarrow \text{ARGWHERE}(\|M[n] - M[b \cdot \lfloor n/b \rfloor]\| > \alpha \cdot v)$ // Locate the outliers.

  $F[n - \omega/2 : n + \omega/2, w, h] \leftarrow 0$ // Update the flag table.

end

for $n = 1$ *to* N **do**

  $M_A[n] \leftarrow \text{AVERAGE}(M[n], \omega)$ // Average a frame by with its neighbors.

end

$M \leftarrow (M - M_A) \cdot F$ // Subtract the averaged values to cancel motion artifacts.

observe that these MAs are generally slower than breathing. This allows us to design a temporal-aware depth image processing algorithm to remove MAs effectively. The core idea is to track the distance measurement on each pixel and compensate the MA for that pixel by subtracting the running mean of the distance readings of this pixel.

The algorithm takes a series of depth images as the input. An example of one depth image frame with some of its pixel readings is shown in Figure 4(a). Evidently, the breathing signal is mixed with body motions. After the compensation algorithm, as shown in Figure 4(b), the MA is removed while the breathing pattern is preserved. Notably, the algorithm constantly faces a situation where when a subject's body moves, some pixels' readings will be interrupted since the subject moves out of its region. This could be resumed after the subject returns to where they were. In this case, the pixel's reading should be discarded when the subject is absent. Otherwise, after subtracting the running mean, the result will contain large outliers that may jeopardize the following deep learning models. Therefore, we introduce a validity check mechanism for the readings of each pixel. The algorithm divides the input image sequence into batches, and it takes the first frame of each batch as the template. Then, the algorithm compares a frame to its closest template prior to it. If the pixel reading of the current frame is too different from that of the template, meaning the subject moves, the pixel at this frame will be considered an outlier and will be set to zero on the output. The threshold for determining outliers is computed dynamically based on the median of the difference of a frame and the template, with a scaling factor $\alpha = 20$. Each batch

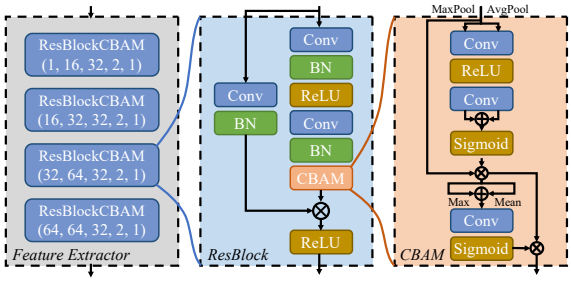
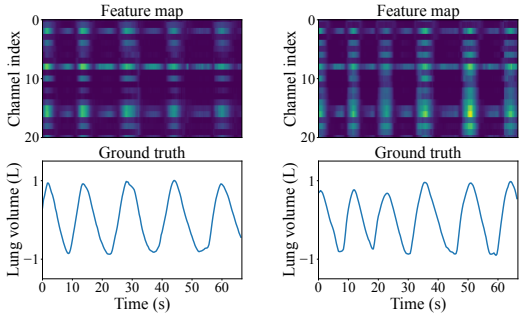


Fig. 5. Feature extractor



(a) Chest breathing (b) Belly breathing

Fig. 6. Examples of the extracted features.

contains 5 seconds of depth images. The window size used to compute the mean values is 20 seconds because our empirical study shows that this time is sufficient to compute MAs. The algorithm is summarized in Algorithm 1.

3.3 MTL-based Breathing Assessment

As we discussed in Section 1, patients can have complex and indistinguishable heaving patterns on the chest and belly regardless of the breathing modes, which makes it hard for us to classify the correct breathing mode. To overcome this challenge, we observe that knowing the current lung volume as prior knowledge can help to decide the breathing mode (Figure 1). Therefore, DeepBreath considers breathing mode and lung volume to be two correlated measurements, while in previous works, these two measurements were estimated separately. Specifically, we use an MTL framework to boost the performance of breathing mode prediction, where the knowledge of lung volume is considered when predicting the breathing mode. In addition, to achieve calibration-free lung volume estimation, we designed a novel UNet-based [51] regression model to achieve one-model-fit-all lung volume regression.

The architecture of the MTL model consists of three models. First, a shared feature extractor powered by ResNet [30] and Convolutional Block Attention Module (CBAM) [66] is used for extracting features from depth images. A separate UNet-based model is used for lung volume regression and a separate Gate Recurrent Unit (GRU) model is used for breathing mode classification. The input of this module is 20-second depth images, which have a dimension of $60 \times 128 \times 80$. The following sections discuss the design of these models in detail.

3.3.1 Depth Feature Extraction. A depth image after MA compensation is first fed to the feature extractor for representation learning. The architecture of the feature extractor is represented in Figure 5. The backbone of this model is consecutive residual convolution layers, which have long proven to be effective in extracting image-like features. However, one characteristic of our depth image is sparsity. As shown in Figure 4(b), more than half of the pixels are zeros resulting from the segmentator. In order to let our model focus more on important features rather than meaningless noise, we introduce a CBAM model to each residual block. The CBAM model first applies maximum and average pooling on the spatial domain to compute channel features and derives the channel-wise attention using convolution layers. Similarly, spatial-wise features can be derived by taking the maximum and mean values on the channel domain, and the spatial-wise attention is computed using convolution layers. Using the CBAM structure, our feature extractor concentrates more on the essential features in the depth image.

The feature extracted by this model has a dimension of $60 \times 64 \times 8 \times 5$, where 60 is the sequence length (20 seconds under 3 fps) and 64 is the channel dimension. We apply an additional adaptive average pooling layer so

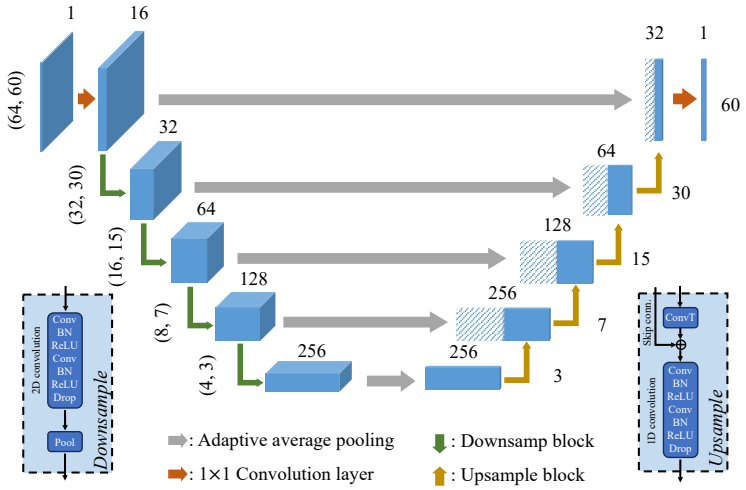


Fig. 7. Lung volume regressor.

that the final output of this model is 60×64. The feature output of this model is used for lung volume regression and breathing mode classification. Examples of the extracted features, together with the actual lung volume and breathing mode, are shown in Figure 6. Note that the original feature has 64 channels. Here, we only show the first 20 channels for clarity.

3.3.2 Lung Volume Regression. We use a UNet-based [51] model for lung volume regression. The architecture of the regressor is shown in Figure 7, which consists of an encoder part and a decoder part. The encoder and decoder are connected with skip connections as it does with classic UNet. The encoder takes the extracted feature as the input and further encodes it with convolution operations. Since the input feature is two-dimensional, the encoder has a classic 2D convolution-based architecture where each block contains two consecutive convolution layers and ends with a maximal pooling layer with a stride of 2 for dimension reduction.

The architecture of the decoder part is mostly symmetric with the encoder. However, since the target output is the lung volume signal - a 1D time series, we adjust the decoder part to align with this 1D structure. Therefore, the convolutions used in the decoder are all 1D convolutions. Specifically, in each decoder block, a transposed convolution with a stride of 2 is first leveraged to upsample the encoded feature, followed by two convolutions for signal reconstruction. After the decoder, we apply an additional convolution layer with a kernel size of 1×1 to summarize the reconstructed feature and predict lung volume. Between the encoder and the decoder, we use an adaptive average pooling layer to compress the 2D encoded feature into 1D. This pooling layer is applied both for the encoder output and the skip connections. The feature encoded by the encoder has a dimension of $256 \times 4 \times 3$, and it becomes 256×3 after the adaptive pooling layer. The final output of the decoder is 1×60.

3.3.3 Breathing Mode Classification. Similar to the lung volume regressor, the breathing mode classifier also takes the feature extracted by the feature extractor as the input. We employ a GRU for breathing mode classification for its ability to process temporal data. We feed the features from the feature extractor to the GRU and use a fully connected layer with one output unit to process the output of the last GRU unit. Finally, we use a Sigmoid layer for classification. The hidden layer of the GRU has 32 units.

3.3.4 Multitask Learning. As we discussed previously, we use the MTL scheme to predict lung volume and breathing mode simultaneously for an observation that these two metrics are correlated with each other, and predicting them cooperatively can boost the accuracy. We train the entire network in an end-to-end fashion, where the inputs are the segmented depth images, and the labels are the ground truth lung volume and the actual breathing mode.

However, these two tasks have different complexities. When training the MTL network, we find that the breathing mode predictor converges fast while the lung volume is hard to converge. This is because learning to regress lung volume is harder since its label is a sequence with values of large variances, while breathing mode prediction is relatively easy with binary labels. Therefore, the learning progress of these two should be balanced in case the model overfits one of these two tasks. In light of this, we use different learning rates for parameters of different models, and we apply different weights to the loss function to balance the convergence speed of the two tasks. Specifically, we use the L1 loss ($\mathcal{L}_1$) to train the volume regressor and the BCE loss ($\mathcal{L}_2$) to train the breathing mode predictor. The final loss function is weighted as $\mathcal{L} = \mathcal{L}_1 + \alpha \cdot \mathcal{L}_2$, where α is set to 0.1 as an empirical pick. In terms of the learning rate, we use a learning rate of $1e{-}3$ for the feature extractor and lung volume regressor and $1e{-}4$ for the breathing mode predictor.

Because of the MTL scheme, the features extracted from the feature extractor contain information on both lung volume and breathing mode. As shown in Figure 6, the features are highly correlated to the volume value. In the meantime, the features exhibit different characteristics for different breathing modes, even if the lung volume values are similar.

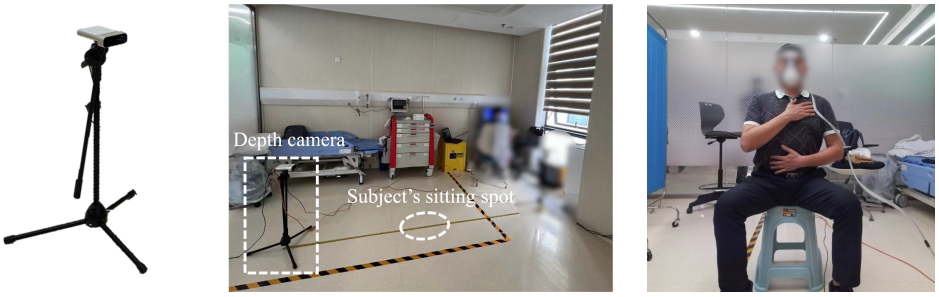
It is worth noting that the system will experience a cold start in the first 20 seconds after the system starts to run, as the models need to take 20-second depth images to give the predictions of the current state. After that, the models can give timely predictions. This waiting time can be reduced by using models with 5-second and 10-second windows at first and then gradually switching to the standard, 20-second window model after the system runs for 20 seconds. Notably, as we will discuss in Section 5.4.4, using a shorter window will degrade DeepBreath's performance. Even so, considering a normal breathing exercise session can take more than 10 minutes, errors in the first 20 seconds can be neglected. In our implementation, this 20-second window size is selected because this can cover at least one breathing cycle after examining our dataset, with the longest breathing cycle being $17s$.

4 Implementation

We develop DeepBreath on Microsoft Azure Kinect [3]. We connect the Kinect camera to a laptop running Windows 11. The depth and RGB image collection system is built with LiveScan3d [35]. The collection system concurrently captures depth and RGB images, together with their timestamps. The timestamps are later used to synchronize with the spirometer data. The Kinect is configured with an "NFOV unbinned" hardware mode where the field of view of the depth camera is $75^\circ \times 65^\circ$, the raw depth image has a resolution of 640×576, and the frame rate is 15 fps. The software system is developed with Python 3.8. The deep learning models are implemented with Pytorch 1.12 and trained with an AdamW [42] optimizer. The batch size used for training all four models is 32, and all models are trained with 100 epochs on the Google Colab platform with an NVIDIA Tesla V100 GPU.

5 Evaluation

In this section, we evaluate the performance of DeepBreath. First, we introduce our data collection protocol and experimental setting, followed by the definition of measurement metrics. Then, we conduct several experiments to demonstrate the sensing accuracy, efficiency and robustness of DeepBreath. Additionally, to show the effectiveness of our core designs, we introduce a series of ablation studies. Unless otherwise specified, we use the leave-one-subject-out (LOSO) cross-validation to present the results. For evaluations related to processing time, the



(a) Depth camera setup. (b) Experiment room. (c) Subject in a session.

Fig. 8. Data collection setup.

experiments are conducted on a PC with a 12th Gen Intel(R) Core(TM) i5-12400F CPU and an NVIDIA GeForce RTX 4060 Ti GPU.

5.1 Evaluation Setup

In this section, we introduce the experimental setup of our evaluations. We first discuss the data collection settings and protocols of the experiments. Then, we introduce a series of metrics that we will use to evaluate DeepBreath's performance.

5.1.1 Data Collection. We collaborate with a medical center and recruit 22 healthy subjects and 14 COPD patients to participate in our experiments. The patients are all diagnosed with mild to moderate COPD by the doctors. All subjects have received written informed consent. Each participant receives 100 CNY [1] compensation for the experiment. The experiment is approved by our institution's IRB[2], and the data collection process is supervised by doctors and therapists. The demography of the participants is shown in Table 1 and Figure 9, where **BMI** (Body Mass Index) is commonly used as an indicator of body fat, **FVC** (Forced Vital Capacity) indicates the overall lung capacity, and **FEV1** (Forced Expiratory Volume during the first second) is an important metric about the presence and severity of obstructive lung diseases [7]. A reduced FEV1/FVC ratio indicates airway obstruction [26]. In the experiments, all subjects wear a shirt or T-shirt by default.

The data collection takes place in a rehabilitation center with an area of around 12 m^2. After informed consent, an introduction to the experiment, and a tutorial on belly breathing, the subject is asked to sit in front of a depth camera with a distance of around 1.5 m and practice breathing exercises with hands placed on chest and belly as discussed in Section 2.1. Meanwhile, the subject also wears a plastic mask, which is connected to a spirometer (ADInstruments, New Zealand [6]) for ground truth lung volume measurement. Note that we let the subject sit on a backless chair, which is more beneficial for the patient to practice breathing but would introduce serious involuntary MAs as discussed in Section 2.1. The depth camera system and the spirometer are synchronized with the Internet timestamp. The data collection setup is shown in Figure 8.

There are two 10-minute data collection sessions for each subject. The first one is for chest breathing, and the second one is for belly breathing. For those subjects that cannot complete the entire 10-minute session, we split the 10-minute session into two 5-minute sub-sessions. By default, the breathing mode of the collected data is labeled as the designated one of that session. However, we found that some patients may fail to maintain the designated breathing mode due to fatigue or lack of practice. Therefore, we manually modified the label of the

[1]Around 13.4 USD (1 CNY ≈ 0.14 USD).
[2]The Hong Kong University of Science and Technology HREP-2023-0350

Table 1. Demographics of the subjects

Stats	COPD	Healthy
Population	14	22
Age (years)	63.23 (6.26)	22.36 (1.29)
# of female	3	7
Height (cm)	162.91 (8.73)	173.01 (7.16)
Weight (kg)	63.40 (10.81)	65.27 (9.54)
BMI (kg/m^2)	23.91 (3.69)	21.72 (2.38)
FEV1 (L)	1.30 (0.81)	/
FEV1/FVC	0.49 (0.14)	/

Format: mean (standard deviation)

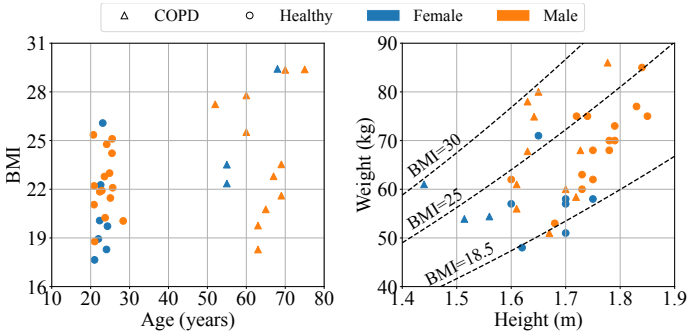


Fig. 9. Breakdown of demographics.

mistaken periods. This case only happens in two of the subjects. During the data collection, the subjects are free to withdraw at any time. In total, we have collected around 10.5 hours of data, among which 3.5 hours are from the patients. The raw depth camera and spirometer data are resampled to 3 Hz for further model training and evaluation.

5.1.2 Performance Metrics. We evaluate DeepBreath's performance with the following metrics. These metrics are comprehensive and represent different aspects of DeepBreath's performance.

- **Breathing rate (BR)**. We compute breathing rate as the number of breathing cycles in one minute (beat-per-minute, BPM). One breathing cycle is identified by locating the peak of the lung volume curve. We use mean absolute BPM error to evaluate DeepBreath's breathing rate estimation performance.
- **Breathing mode classification (BMC)**. We use precision, recall, and F1-score to evaluate DeepBreath's performance on breathing mode classification. These metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TN}+\text{FN}}, \quad \text{F1-score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

  where a positive means chest breathing, and a negative means belly breathing.
- **Lung Volume (LV)**. We use four metrics to measure the error between the predicted and actual lung volume. (i) Correlation. We compute the Pearson correlation coefficient to measure the similarity between the predicted lung volume curve and the actual one. (ii) MAE. we compute the mean absolute error (MAE) to quantify the volume measurement error. (iii) IC MAE. As discussed in Section 2.1, breathing exercise trainees are encouraged to take deep breaths so the inhaled air volume reaches their inspiratory capacity (IC) value, where IC is defined as the maximal air one can breathe in after a normal exhalation [4]. Therefore, it is important to evaluate DeepBreath's ability to predict accurate IC values. (iv) IC MAE%. The percentage error of IC is also usually used for evaluating the prediction error [32, 47]. The percentage error is computed as the IC MAE divided by the actual IC value.
- **System delay**. We also compute the system delay as the time it takes for DeepBreath to update its prediction. We separately evaluate the delay caused by each module in DeepBreath.

5.1.3 Baselines. To show the novelty of DeepBreath, we compare DeepBreath's performance against other baseline methods. We have implemented the following baselines for comparison.

- **Soleimani *et al.* [58]**. In this work, the authors design several signal-processing algorithms with calibration to evaluate lung function from depth images. For each subject, they fit a linear model to transform depth readings into lung volume measurements. They tested their method on 85 patients and achieved a high

Table 2. Overall performance

Metric	All	COPD	Healthy
Breathing rate error (BPM)	0.09 (0.14)	0.10 (0.07)	0.09 (0.16)
Breathing mode F1-score	0.92 (0.23)	0.82 (0.25)	0.96 (0.21)
Lung volume correlation	0.95 (0.05)	0.96 (0.04)	0.95 (0.05)
Lung volume MAE (L)	0.09 (0.05)	0.10 (0.04)	0.08 (0.05)
Lung volume IC MAE (L)	0.18 (0.15)	0.19 (0.17)	0.18 (0.15)
Lung volume IC MAE%	0.16 (0.12)	0.17 (0.13)	0.15 (0.11)

Format: mean (standard deviation)

correlation between the measurements from the proposed method and a spirometer. We implement their algorithms and include a calibration-free version where we test the performance on one subject using the linear model fitted from other subjects.

- **Ostadabbas *et al.* [49]**. This work proposes an automatic chest bounding algorithm and an airflow signal calculation algorithm, which integrates the depth readings and computes the tidal volume. Similarly, they assume the relationship between the depth readings and the airflow is constant for an individual and construct a prediction model for each subject. This method was tested on 14 patients. Again, we implement two versions of this baseline: one is a personalized model and the other is a calibration-free model where data from other subjects are leveraged to construct the model for the target subject.
- **DeepBreath without MTL**. In this work, we use an MTL framework to boost the performance of breathing measurements. We are interested in its performance if the lung volume regression and breath mode classification tasks run separately. Therefore, we also include a baseline where the MTL framework is deactivated to compare our MTL-based approach.

Note that for breathing mode estimation, we find there is limited research work to compare with except for BreathMentor [23]. However, BreathMentor is an ultrasound-based method that is not suitable to serve as our baseline. Therefore, we only include the without-MTL version of DeepBreath as our baseline to compare the performance of breath mode classification.

5.2 Performance Study

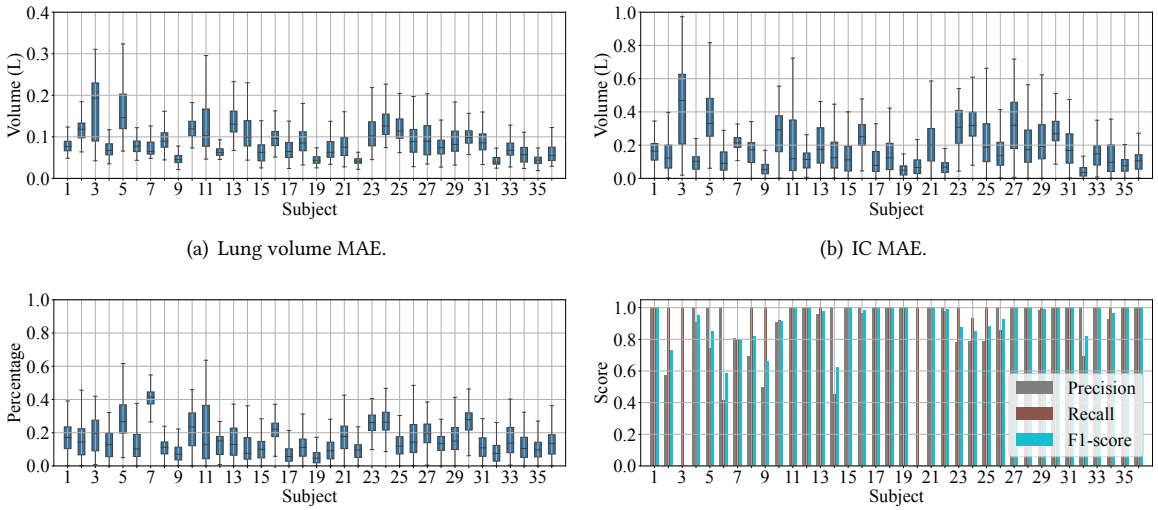
In this section, we consider DeepBreath's three modules as a whole and evaluate DeepBreath's end-to-end performance against the performance metrics discussed above.

5.2.1 Overall Performance. We run LOSO cross-validation on each subject. The overall performance is summarized in Table 2. The mean breathing rate error is 0.09 BPM. This error is extremely low compared with other noncontact solutions [9, 23, 64, 65, 68, 73]. This is reasonable since a depth camera can capture finer-grained chest measurements than other modalities, and breathing exercises, the scenario we consider, involve larger breathing amplitude than at rest. The F1-score of mode classification for healthy subjects is 0.96, which is a comparable performance with BreathMentor [23] whose F1-score is 0.97. However, BreathMentor is only evaluated on healthy subjects, while the performance on patients is unknown. On the contrary, DeepBreath can deal with patients' complex breathing patterns and shows a 0.82 F1-score in breathing mode classification. For lung volume regression, the mean correlation coefficient between the estimated one and the ground truth is 0.95, the lung volume MAE is 0.09L, the mean IC prediction error is 0.18L, and the mean IC percentage error is 0.16. In addition, considering most of the previous work adopts a per-subject calibration approach, we also implement a calibration-based version for DeepBreath as a comparison to the baseline methods. We use a subject's first 60%

Table 3. Baseline comparison.

Baselines	Soleimani *et al.* [58]		Ostadabbas *et al.* [49]		DeepBreath w/o MTL		DeepBreath	
Calibration	w/	w/o	w/	w/o	w/	w/o	w/	w/o
Lung volume MAE (L) ↓	0.12 (0.11)	0.17 (0.18)	0.13 (0.13)	0.42 (0.52)	0.05 (0.04)	0.09 (0.07)	0.04 (0.03)	0.09 (0.05)
Breath mode F1-score ↑	-	-	-	-	0.96 (0.16)	0.83 (0.19)	0.98 (0.03)	0.92 (0.24)

Format: mean (standard deviation)



(a) Lung volume MAE.

(b) IC MAE.

(c) IC MAE%.

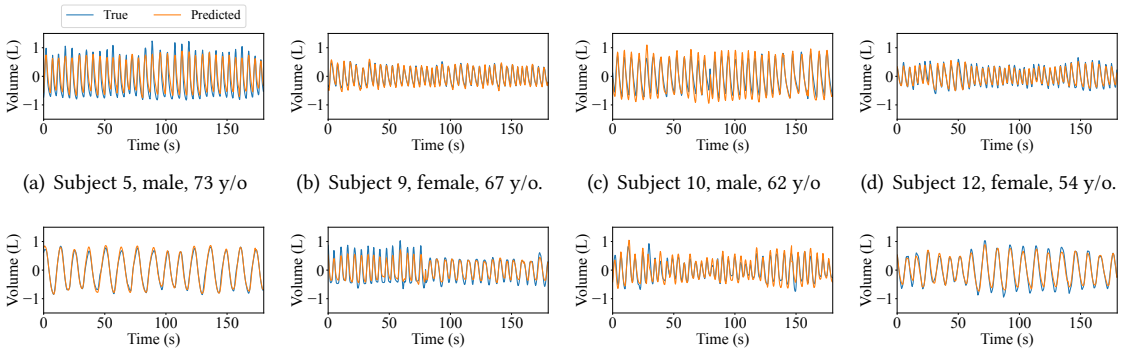
(d) Breath mode classification.

Fig. 10. Overall performance. (Subjects 1-14 are patients.)

data to train a personalized model and use the rest of the data for testing. In this case, DeepBreath can achieve extremely high performance with 0.04 L error of lung volume estimation and 0.98 F1-score for breathing mode classification.

Compared with the baseline methods, our method achieves a higher performance as shown in Table 3. For the lung volume estimation, even if our method is calibration-free by default, we can achieve a lower error compared with two of the calibration-based baselines, while these baselines work with larger errors without calibration. For breathing mode classification, our method significantly improves the accuracy compared with the baseline method, thanks to the MTL design.

5.2.2 Breathing Mode. As discussed in Section 1, breathing mode classification for patients can be very challenging because of their abnormal breathing patterns. However, by utilizing an MTL model, we can boost the accuracy by letting the breathing mode predictor consider the information of the current lung volume. The overall result is shown in Figure 10. In particular, the F1-score for patients and healthy subjects are 0.82 and 0.96. We also present the precision, recall, and F1-score for each subject in Figure 10(d). Note that two of the subjects



(a) Subject 5, male, 73 y/o (b) Subject 9, female, 67 y/o. (c) Subject 10, male, 62 y/o (d) Subject 12, female, 54 y/o.

(e) Subject 19, male, 27 y/o (f) Subject 25, female, 22 y/o (g) Subject 26, female, 21 y/o (h) Subject 27, male, 22 y/o

Fig. 11. Examples of lung volume estimation. (a)-(d) Patients. (e)-h) Healthy subjects.

(Subjects 3 and 20) do not have precision and F1-score because they withdraw from the experiments, leaving only the chest breathing data. In Section 5.4, we will evaluate the performance gain of the MTL design in terms of breathing mode classification.

5.2.3 Lung Volume. Lung volume is an important indicator of the quality of the breathing exercise. Here, we evaluate DeepBreath's lung volume regression accuracy with four metrics as discussed in Section 5.1.2. In general, DeepBreath achieves very accurate lung volume tracking without calibration. Examples of the predicted lung volume curve together with the ground truth one are shown in Figure 11. (i) **Correlation**. We compute the correlation between the predicted lung volume curve and the actual one to measure DeepBreath's ability to capture the general shape of the lung volume pattern. As presented in Table 2, the overall correlation coefficient is around 0.95, which indicates a fairly high similarity compared with the ground truth. This also can be observed from the examples in Figure 11. (ii) **Lung volume MAE**. The mean absolute error (MAE) is the most direct measurement of the prediction error. DeepBreath achieves a mean MAE of 0.09 L for all subjects, 0.1 L for patients, and 0.08 L for healthy subjects. This result indicates that DeepBreath can track lung volume quantitatively quite well. The box plots of lung volume MAE for every subject are shown in Figure 10(a). (iii) **Inspiratory capacity MAE**. Doctors and therapists care much about whether the patient reaches their IC during breathing exercises, which is a strong indicator of evaluating the patient's lung function and breathing efforts. Therefore, we also calculate the IC MAE as discussed in Section 5.1.2. DeepBreath achieves a mean IC MAE of 0.18 L for all subjects and almost no difference between the healthy subjects and patients. Figure 10(b) shows the box plots of IC MAE for every subject. Note that IC MAE is larger than the lung volume MAE. This is reasonable because IC MAE measures the errors at the peaks and valleys of the lung volume curve prediction, and estimations on these key points are much harder than the others. (iv) **Inspiratory capacity MAE%**. Lastly, we also present the IC percentage MAE. The overall error is 0.16, where the patient population has a slightly higher error of 0.17 compared with 0.15 from the healthy population. The box plots of the percentage IC MAE for every subject are shown in Figure 10(c).

5.2.4 Processing Delay. We also conduct experiments to see whether DeepBreath can support real-time operation. We simulate the real-time operation by feeding our system with streams of random input and measure the delay caused by each component accordingly. We assume a 1Hz refreshing rate in the simulation. Therefore, we feed the segmentator with one second (3 frames) of data for each run. For the motion artifact compensation module

Table 4. System delay.

Device	SEG	MAC	MTL model			Total
CPU	50.8 *ms*	20.6 *ms*	68.2 *ms*			139.6 *ms*
			EXT	**CLS**	**REG**	
			55.8 *ms*	7.0 *ms*	5.4 *ms*	
GPU	2.8 *ms*	19.3 *ms*	7.2 *ms*			29.5*ms*
			EXT	**CLS**	**REG**	
			3.6 *ms*	0.3 *ms*	3.3 *ms*	

SEG: image segmentation. MAC: MA compensation. EXT: feature extractor.
CLS: breath mode classifier. REG: lung volume regressor.

and the breathing assessment module, we feed them 20 seconds (60 frames) of data for each run because these modules require a 20-second input regardless of the refreshing rate. The simulation continues for 100 runs. Table 4 summarizes the result. This result indicates that with a CPU, the total time it takes for DeepBreath to give an output is 139.6 ms, which shows the feasibility of a 1Hz refreshing rate. In addition, the delay can be reduced to 29.5 ms if DeepBreath is run on a GPU device. However, as discussed in Section 6, future work should consider making the system real-time and evaluate the system delay thoroughly.

5.3 Robustness Study

This section provides an extra set of experiments to evaluate the robustness of DeepBreath under different real-world conditions. Specifically, we conduct experiments with altered experimental parameters: (i) the depth camera's position, (ii) the subject's orientation, (iii) the camera's view angle, (iv) the subject's clothing, (v) the subject's voluntary motions, and (vi) subject's posture. We recruit three healthy subjects to participate in the experiments. We ask these subjects to conduct breathing exercises under different conditions where each subject performs three-minute chest and belly breathing. We use the model trained in Section 5.2 to give breathing metrics estimations under these conditions in the LOSO manner.

5.3.1 Impact of Distance. In the above evaluations, the default distance between the subject and the camera is about 1.5 *m*. In this section, we change this distance to 0.5 *m*, 1 *m*, 2 *m*, and 2.5 *m*, as shown in Figure 13(a), to test the system performance. We compute DeepBreath's performance in lung volume regression and breathing mode classification under these conditions, and the results are shown in Figure 12. From this result, we can see that DeepBreath can operate with high accuracy within the range of 1 *m* to 1.5 *m*. However, if the subject sits too close to or far from the camera, the performance drops a little. Even so, we believe a proper mechanism can be designed to prevent the user from sitting in such extreme proximity. This is because, after segmentation, we can easily obtain the user's sitting distance by directly analyzing the depth images. In this case, the system can suggest the user adjust the position if a too close or far away distance is detected.

5.3.2 Impact of Subject's Orientation. During the data collection process, we found that although we explicitly instructed the participants to sit right in front of the depth camera, some participants may unconsciously adjust their orientation before or during the experiments. Therefore, DeepBreath needs to tolerate this orientation shift. We define the orientation of the subject as the angle between their facing direction and the middle of the depth camera's field of view. In this section, we vary the orientation of the subject so that he/she is 15°, 30°, and 45° with respect to the camera, as shown in Figure 13(b). Without the loss of generality, we assume the orientation leans towards the right-hand side of the subject. The result in Figure 12 indicates that DeepBreath can operate with high accuracy with orientation within 30° (60° if both sides are considered) with respect to the camera.

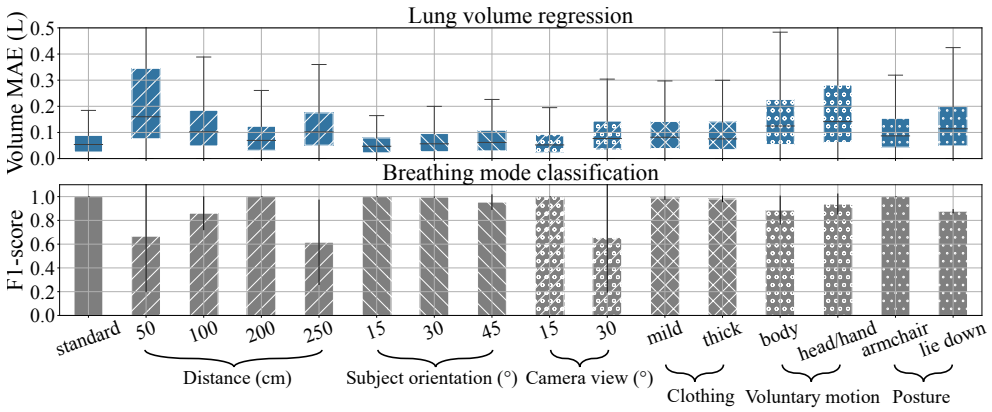
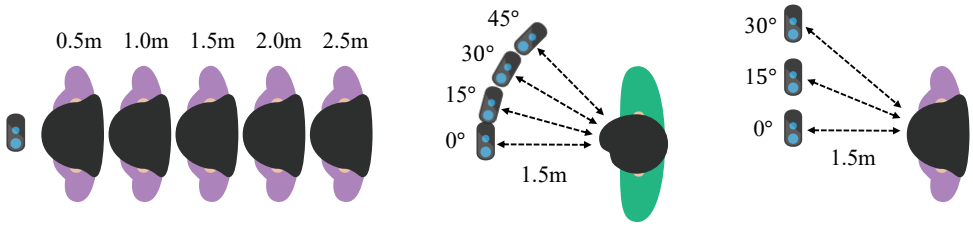


Fig. 12. Robustness study.



(a) Distance. (b) Subject's orientation. (c) Camera's view angle.

Fig. 13. Robustness study - camera position.

5.3.3 Impact of Camera's View Angle. It is important to consider the case when the subject does not sit right in the middle of the camera's field of view. Therefore, we adjust the camera's view angle so that the subject sits at 15° and 30° with respect to the middle of the camera's field of view while keeping the subject's orientation unchanged. The setup for this experiment is shown in Figure 13(c). Without the loss of generality, we adjust the camera's location to the right-hand side of the subject. Note that because the single-side field-of-view under our hardware configuration is 37.5° as discussed in Section 4, when the camera has a 30° view angle with respect to the subject, the subject will be located at the edge of the captured depth image. Therefore, we do not introduce more view angle variations. The result shown in Figure 12 indicates DeepBreath can work with high accuracy with a view angle within 15° (30° if both sides are considered). When the view angle increases to 30°, we see a little performance drop. This is reasonable since, in these cases, some parts of the human body may be out of the image and this can jeopardize the performance of both the segmentator and the breathing measurement models.

5.3.4 Impact of Clothing. By default, all the subjects wear a single-layer shirt or T-shirt during the experiments. In this evaluation, we test DeepBreath's performance when the user wears clothes of different thicknesses. We let the subjects wear clothes with three different levels of thickness, including (i) thin clothing, where the subject wears a T-shirt; (ii) mild clothing, where the subject wears a T-shirt and a hoody; and (iii) thick clothing, where the subject wears a T-shirt, a hoody, and a jacket. The clothes used in this evaluation are shown in Figure 14. The result in Figure 12 shows that compared with only a T-shirt, wearing a thicker cloth slightly increases the errors in

(a) A T-shirt. (b) A hoody. (c) A jacket.

Fig. 14. Robustness study - clothing.

Fig. 15. Robustness study - lying down.

lung volume estimation, but the estimations are still accurate with MAE less than 0.1 L. This result demonstrates DeepBreath's robustness when facing different clothing. We believe this is because a typical breathing exercise requires the subject to place their hands on the chest and belly, and this action ensures firm contact between the body and the cloth so that the chest vibration patterns can be captured by the camera regardless of the clothing.

5.3.5 Impact of Voluntary Motions. As discussed in previous sections, we use an MA-compensation algorithm to cancel out the influences caused by involuntary body motions. In addition, we are also interested in DeepBreath's robustness in facing voluntary motions, which are much larger and more random than involuntary ones. We include this evaluation to test DeepBreath's limit in tolerating body motions. Specifically, we have two experiments with different types of voluntary motions, including (i) body motions, where the subject deliberately swings their body back-and-forth or left-to-right, and (ii) head and hand motions, where the subject randomly swings their head and hands voluntarily. The results are shown in Figure 12. Compared with normal involuntary motions, the lung volume estimation under these two cases has a larger error of around 0.12 L. This result indicates that large voluntary motions will introduce large errors in lung volume estimation, and users of DeepBreath should try not to move their bodies deliberately. Even though lung volume estimation is affected by large voluntary body motions, the breathing mode classification performance still maintains a high accuracy with an around 0.88 F1-score under these two cases.

5.3.6 Impact of Body Posture. The default posture for breathing exercises in this research is sitting in a chair without back support. This posture is, as discussed in Section 2.1, optimal for breathing exercises. However, in practical use, the users may not strictly follow this recommendation and may have other preferred postures. In this evaluation, we test DeepBreath's robustness in facing different body postures. We introduce the following common body postures: (i) the subject sits comfortably in an armchair, and (ii) the subject lies down on a yoga mat. Notably, in the second case, we configured the depth camera to face downward and be around 1.5 meter above the ground, as shown in Figure 15. The results are presented in Figure 12. When the subject sits in an armchair, the performance of lung volume estimation degrades a little bit. We hypothesize that this is because when sitting in an armchair, the subject's body inclines due to relaxation, and this will increase errors in volume estimation. Even such, the error is still very low with MAE less than 0.1L. Regarding the subject lying down on a mat, the performance degrades further with lung volume MAE around 0.1 L and breathing mode classification F1-score of 0.87. This is because the estimation models have only seen sitting postures in the training set and may not generalize well to other postures. We believe increasing the training dataset diversity could effectively boost the performance of other postures like lying down.

5.4 Technical Effectiveness Study

In this section, we study the effectiveness of each technical design in DeepBreath. We systematically change or remove some parts of the architecture, train our deep-learning models again and test the performance on the new models. In this evaluation, both the full model and its variations are validated using five-fold cross-validation. The results are shown in Figure 16.

5.4.1 Benefits of Multitasking. We split the MTL model into two separate models, a lung volume regressor and a mode classifier, and train and test them separately. Specifically, for the lung volume regression, we connect the UNet-based regressor directly to the feature extractor, and a similar setup also applies to the breathing mode classifier. Notably, we observe that the performance drops sharply once we split the mode classification model from the MTL framework without modification. This is probably because the original network parameters are tuned to best fitting the MTL scheme. For fairness, we further optimize the mode classification model. Specifically, we reduce the size of the feature extractor from four layers to three layers, and we add an additional max pooling layer after the GRU. As Figure 16 shows, although the MTL framework seems to have little enhancement for lung volume regression, the task of breathing mode classification benefits a lot from it. By leveraging the MTL framework, we see an around 10% F1-score gain for all subjects. The gains for the patients and healthy subjects are 11% and 9%, respectively.

5.4.2 Benefits of Segmentation. Next, we test the system performance with and without body shape segmentation. The result is shown in the third and fifth columns in Figure 16. To summarize, introducing the segmentation module helps reduce the lung function regression error by 10%, and the overall breathing mode classification F1-score is increased from 0.85 to 0.89. In addition, we implement a vanilla segmentation algorithm and compare its performance with our segmentation model. This segmentation algorithm segments a depth image through thresholding. It first sets the distance reading at the middle of an image as the baseline distance, assuming the subject is located in the middle of the image. Then, it discards the pixels whose distance reading is 50 cm away from the baseline. The result shows this segmentation method achieves an F1-score of 0.86 which is similar with no segmentation at all. We believe this is because a pure thresholding-based algorithm cannot cope with subjects with various shapes and the influence of environmental objects, while our segmentation model can segment the subject accurately.

5.4.3 Benefits of Motion Artifact Compensation. As we mention in Section 3.2, involuntary MA is one of the factors affecting the quality of both mode classification and lung volume regression. Hence, in this section, we compare the accuracy of lung volume prediction with and without MA compensation. The results are presented in columns four and five in Figure 16. Based on these comparisons, we can conclude that our MA compensation algorithm significantly improves the system's performance. For breathing mode classification, the model almost cannot work once we drop the MA compensation module. Similarly, lung volume prediction can have large errors without MA compensation.

5.4.4 Impact of Prediction Window Size. As discussed in Section 3.3.4, in addition to the standard 20-second window model, DeepBreath leverages models with 5-second and 10-second windows at the very beginning of a session to reduce the cold start time. However, due to the fact that one cycle of human breathing can take more than 10 seconds, a shorter window may not be sufficient to cover an entire breathing cycle. In particular, the longest breathing cycle logged in our dataset is around 17s, from subject 20. Therefore, the performance will be affected if a shorter window is utilized. Here, we compare the performance when a 5-second, 10-second, and 30-second window is used, respectively. The results are shown in the last three columns in Figure 16. Compared with the standard 20-second window, the performance drops when shorter windows are used, especially for the performance of breathing mode classification. We believe this is because it is hard for the model to identify the

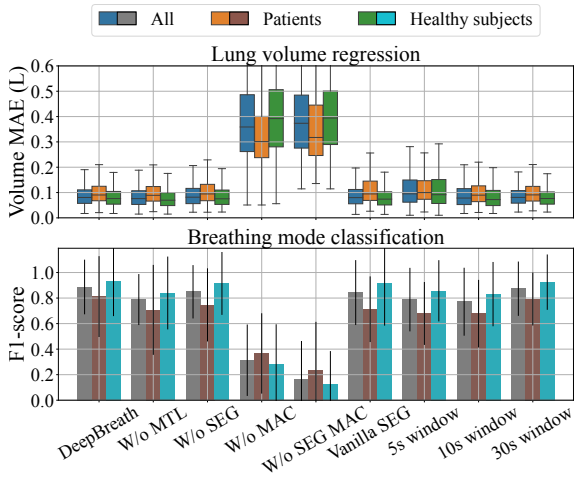


Fig. 16. Technical effectiveness study. (SEG: image segmentation. MAC: motion artifact compensation.)

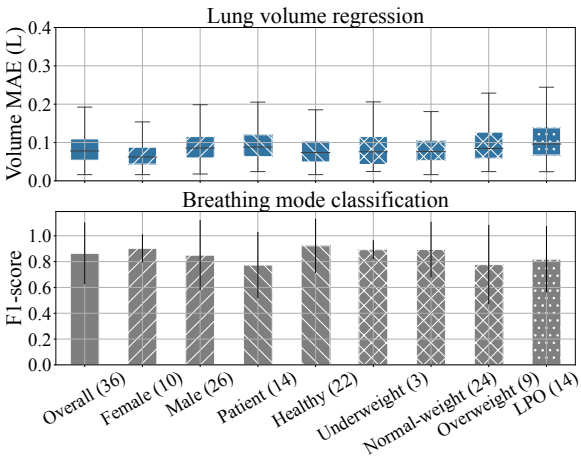


Fig. 17. Demographic study.

correct breathing mode if it does not see a full breathing cycle. Notably, the performance improves a little when a 30-second window is used. While this improvement is not significant, we believe a 20-second window can well balance the trade-off between the cold start time and the performance.

5.5 Demographic Study

The demographics of our subjects are shown in Table 1 and Figure 9. This section demonstrates our system performance on different demographic groups. We evaluate our system in terms of gender, health condition, and BMI value. Here, subjects are grouped into underweight (BMI<18.5), normal-weight (18.5<BMI<24.9), and overweight (BMI>25) according to the international standard. The result in Figure 17 shows no significant variance among these groups. Notably, there is a slight performance degradation in the overweight group. We hypothesize that this might be because the observable breathing pattern of this group is less than the others. Also, the overweight group contains mostly patients, and DeepBreath's performance is slightly worse than that of healthy subjects, as discussed in previous sections. However, more experiments with a larger population are needed for verification. Potentially, this problem can be solved if a personalized model is used as discussed in Section 5.2.1. In addition, we observe that there is a gap in ages between the healthy subjects and the patients. In the previous sections, we conducted leave-one-subject-out and random five-fold cross-validations where the training set contained data from both the healthy subjects and the patients. It is unclear how the heterogeneity between patients and healthy subjects can affect performance. Therefore, here we conduct an additional leave-patient-out (LPO) validation, where we train the model on data from the healthy subjects and test the model on patients, to hint at the model's generalizability to other user groups. The result presented in the last column of Figure 17 shows no significant performance degradation, showing our model can be generalized from the healthy population to the patients on our dataset.

5.6 Performance of the Segmentator

In this section, we evaluate the performance of our segmentator. As a comparison, we also run experiments on the teacher model, the Segment Anything Model (SAM) [34], and compare its performance with our segmentator.

(a) Top 25% (b) Bottom 25%

Fig. 18. Segmentation examples. (Left: ours. Right: SAM.)

Table 5. Comparison of segmentation models.

	SAM ViT-h	SAM ViT-b	Ours
Total params:	641M	94M	0.8M
Power* (J):	140	31.2	0.4
Loading (s)	3.97 (3.66)	0.85 (0.80)	0.13 (0.12)
Processing (s)	22.90 (1.02)	6.20 (0.26)	0.02 (0.002)
mIoU	-	0.87	0.92

Format: CPU (GPU)
* : Measured by energy consumption per image.

There are multiple versions of SAM, among which the one with pretrained ViT-h is the largest while the one with pretrained ViT-b is the smallest. We compare these two models with our customized segmentator regarding processing time and performance, where images segmented by ViT-h-based SAM are regarded as the gold standard. The results are shown in Table 5. For processing time, it takes up to about 23 seconds to segment an image if tested on a CPU. Even with GPU and the ViT-b-based SAM, it still takes 0.26s to process a frame, which is not enough to achieve real-time processing considering our 3fps frame rate, while our segmentation only takes 0.02s. For performance, our segmentator outperforms the ViT-b-based SAM model in terms of mean intersect over union (mIoU). We also measure the power consumption of these models by computing the average energy it takes for the model to segment one image. The result shows that our segmentator is much more power-efficient than the other two. Some segmentation examples are shown in Figure 18.

6 Discussions

In this section, we will discuss the limitations of DeepBreath and point out the future directions of this study.

Population diversity. As shown in Figure 9, the demographics of the participants are biased. Specifically, the healthy population in this research is mostly 20-30 years old, while the COPD patients are all above 50 years old. Also, it is unclear how well DeepBreath would work on 30-50 years old subjects since data from this age group is missing. Therefore, future research should target increasing the diversity of the recruited subjects and test the performance of DeepBreath on a wider range of age groups.

Robustness in complex scenarios. In the current setup, we ask the subjects to sit in front of the camera at a distance of around 1.5m while facing right at the camera. Although our evaluations in Section 5.3 show that our system is robust to other sitting positions and orientations, the performance may drop when facing more extreme and complex situations. To further improve DeepBreath's robustness, we could enhance our training dataset with a greater variety of sitting positions and orientations. Also, we could leverage more advanced domain adversarial training frameworks [33] to increase the generalizability of our deep-learning models.

User experience study. Another thing worth noting is that even though we show through simulation that DeepBreath supports real-time operation, we have not yet implemented the system in real-time. Future work should consider making the system work in real-time so that actual breathing exercise guiding systems can be developed for the benefit of COPD patients. After that, future work should also study the patients' experience when using the system to study whether the design can truly enhance adherence to patients' breathing exercises and, if necessary, introduce gamification to DeepBreath to make the system more engaging.

Form factor. The depth camera we use in this work is Microsoft Azure Kinect [3], a commercial product in the market. However, this device is not so widely seen in common households. Even so, with depth cameras becoming available in smartphones and tablets [8], we see the potential of integrating DeepBreath on these mobile devices and leveraging the existing depth cameras on these devices to supervise breathing exercises. Also,

there are various types of dedicated and more cost-efficient depth sensors in the market, future studies should also investigate the feasibility of leveraging these depth sensors to measure breathing.

Privacy issues. Users of DeepBreath may raise privacy concerns when using it. This is because a depth camera will take depth images of the user and may also capture depth images of the user's background which, in our scenario, is the user's household. To minimize the risk of privacy leakage, the deep-learning models and algorithms that power DeepBreath should run locally. Also, privacy-preserving deep-learning frameworks such as federated learning [72] can be applied when training or updating deep-learning models.

7 Related Works

In this section, we summarize the related works of this paper. The reviewed works are grouped into two categories: breathing measurement methods and the applications of depth cameras in rehabilitation.

7.1 Breathing Assessment

Breath monitoring has attracted considerable interest for its indication of fundamental body function, including physiological and mental aspects. We start with summarizing the clinical practice for breathing assessment, followed by discussing other methods in the research field.

7.1.1 Clinical Solutions. Currently, most respiratory rate (RR) monitoring methods rely on measuring airflow changes, typically through the use of a nasal cannula. However, this approach often causes discomfort to patients [13, 40]. Furthermore, for a comprehensive assessment of lung volume, more extensive respiratory information is required. Spirometry and body plethysmography are commonly employed techniques that estimate airflow using differential pressure sensors [20, 24]. When it comes to evaluating breathing patterns, ultrasound imaging has emerged as an effective method for assessing diaphragm motion [63]. Nevertheless, these techniques are time-consuming and cumbersome. In practice, doctors are accustomed to monitoring a patient's actual breathing mode by placing their hands on the patient's chest and abdomen [27, 63].

7.1.2 Mobile Solutions. To detect breathing rate, integrated IMU sensors are utilized to capture subtle motions induced by respiration [28, 38, 60]. Some other works have successfully extracted RR from PPG signals [39, 44]. Additionally, contactless sensing methods are acting as promising alternatives. Some work leverages microphones to detect RR [11, 56, 67], and radio-frequency (RF) based techniques, like WiFi, millimeter wave, ultra-wideband, and active acoustic-based techniques, can do better, especially in diverse environments [9, 23, 64, 65, 68, 73]. Moreover, RGB and depth cameras have also been explored to measure RR [12, 43, 50, 53]. However, only RR is insufficient for accurate breathing exercise assessment. Regarding the breathing mode classification, BreathMentor [23] utilizes a microphone array to detect distinct modes. Nevertheless, this work only focuses on a healthy population. Recently, in Orlova *et al.* [48], a motion capture system is used to determine the proportions of different breathing modes. For lung volume measurement, Chu *et al.* [19] employ three strain sensors attached to the human body to scale the air exhalation, and Sharma *et al.* [54] use two sensors applied to the chest and abdomen in order to detect the rough breathing waveform. Meanwhile, contactless signals, such as radar and acoustics, have also revealed great power to track breathing waveform [45, 69, 76], yet none of them can predict the real lung volume. The utilization of depth cameras for estimating lung volume has been widely explored [29, 32, 47, 49, 55, 57–59]. However, most of the existing approaches are not suitable for deployment in everyday training scenarios. They are either influenced by body swings during deep breathing [57], require per-subject calibration with a clinical spirometer [32, 49, 55, 57–59], require participants to be shirtless and in close proximity to the camera [47], or require the use of two or more cameras [29, 32].

7.2 Depth Camera Applications for Rehabilitation

With VR techniques, depth cameras can be exploited to improve adherence in rehabilitation programs [15]. Several studies [14, 16, 70] have shown their ability to assess the patients' completion of specific movements. For example, Cho *et al.* [21] employs depth cameras to implement immersive virtual prism adaptation therapy for stroke patients. In Grooten *et al.* [25], one depth camera is utilized to capture a pre-defined set of movements, enabling the evaluation of posture, balance, and scoliosis while differentiating between healthy individuals and patients. Besides, Dubois *et al.* [21] investigate the gait parameters with depth cameras to predict the risk of falls.

8 Conclusion

This research proposes DeepBreath, a novel depth camera-based breathing exercise assessment system that overcomes the limitations of existing methods. DeepBreath applies a multitask learning framework to achieve high-accuracy breathing mode estimation and uses a data-driven approach with a novel UNet-based deep-learning model to achieve calibration-free lung volume estimation. The system is designed to be resilient to motion artifacts with a temporal-aware motion artifact removal algorithm and a lightweight silhouette segmentation model to enhance performance. We collaborated with a medical center and recruited 22 healthy subjects and 14 COPD patients to validate our design. The experimental results show that DeepBreath achieves high accuracy, strong robustness, and a much simpler setup compared with the previous works.

Acknowledgments

The authors would like to thank Christopher Remde for his help in deploying LiveScan3D [35], from which we built our data collection system. The authors would also like to thank Dr Rongchang Chen from the First Affiliated Hospital of Guangzhou Medical University and his team for their support in using their facilities. This research is supported by Hong Kong RGC under Contract CERG 16206122, 16204820, AoE/E-601/22-R, Contract R8015, and 3030_006. Dr Qian Zhang, Dr Zeguang Zheng and Dr Shifang Yang are the corresponding authors.

References

[1] 2022. Breathing Exercises | America Lung Association. https://www.lung.org/lung-health-diseases/wellness/breathing-exercises Accessed Nov 22, 2023.
[2] 2022. Diaphragmatic Breathing Exercises and Benefits. https://my.clevelandclinic.org/health/articles/9445-diaphragmatic-breathing Accessed Jan 17, 2024.
[3] 2023. Azure Kinect DK - Develop AI Models | Microsoft Azure. https://azure.microsoft.com/en-us/products/kinect-dk Accessed Nov 24, 2023.
[4] 2023. Inspiratory Capacity. https://www.ncbi.nlm.nih.gov/mesh/?term=Inspiratory+Capacity Accessed Nov 25, 2023.
[5] 2023. SenseEcho. https://www.sensecho.com/ Accessed Nov 22, 2023.
[6] 2023. Spirometry | ADInstruments. https://www.adinstruments.com/research/human/respiratory/spirometry/ Accessed Nov 25, 2023.
[7] 2024. Pulmonary Function Tests. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/pulmonary-function-tests Accessed Jan 29, 2024.
[8] 2024. Use Face ID on your iPhone or iPad Pro. https://support.apple.com/en-hk/HT208109 Accessed Jan 26, 2024.
[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart Homes That Monitor Breathing and Heart Rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 837–846.
[10] Md Atiqur Rahman Ahad, Anindya Das Antar, and Omar Shahid. 2019. Vision-based Action Understanding for Assistive Healthcare: A Short Review.. In *CVPR Workshops*. 1–11.
[11] Tousif Ahmed, Md Mahbubur Rahman, Ebrahim Nemati, Mohsin Yusuf Ahmed, Jilong Kuang, and Alex Jun Gao. 2023. Remote Breathing Rate Tracking in Stationary Position Using the Motion and Acoustic Sensors of Earables. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 325, 22 pages.
[12] Flavia Benetazzo, Alessandro Freddi, Andrea Monteriù, and Sauro Longhi. 2014. Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results. *Healthc Technol Lett* 1, 3 (Sep 2014), 81–86.

[13] K. Bhavani-Shankar, H. Moseley, A. Y. Kumar, and Y. Delph. 1992. Capnometry and anaesthesia. *Can J Anaesth* 39, 6 (Jul 1992), 617–632.
[14] Bruno Bonnechère, Victor Sholukha, Lubos Omelina, Bart Jansen, and S Van Sint Jan. 2016. Validation of Trunk Kinematics Analysis through Serious Games Rehabilitation Exercises Using the Kinect™ Sensor. In *Proceedings of the 4th Workshop on ICTs for Improving Patients Rehabilitation Research Techniques* (Lisbon, Portugal) *(REHAB '16)*. Association for Computing Machinery, New York, NY, USA, 45–48.
[15] Elizabeth B Brokaw, Emily Eckel, and Bambi R Brewer. 2015. Usability evaluation of a kinematics focused Kinect therapy program for individuals with stroke. *Technol Health Care* 23, 2 (2015), 143–151.
[16] Y. J. Chang, S. F. Chen, and J. D. Huang. 2011. A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. *Res Dev Disabil* 32, 6 (2011), 2566–2570.
[17] Yu-Fen Chen, Xuan-Yi Huang, Ching-Hui Chien, and Jui-Fen Cheng. 2017. The Effectiveness of Diaphragmatic Breathing Relaxation Training for Reducing Anxiety. *Perspectives in Psychiatric Care* 53, 4 (2017), 329–336.
[18] Denesh K Chitkara, Miranda Van Tilburg, William E Whitehead, and Nicholas J Talley. 2006. Teaching diaphragmatic breathing for rumination syndrome. *Am J Gastroenterol* 101, 11 (Nov 2006), 2449–2452.
[19] Michael Chu, Thao Nguyen, Vaibhav Pandey, Yongxiao Zhou, Hoang N Pham, Ronen Bar-Yoseph, Shlomit Radom-Aizik, Ramesh Jain, Dan M Cooper, and Michelle Khine. 2019. Respiration rate and volume measurements using wearable strain sensors. *NPJ Digit Med* 2 (2019), 8.
[20] CP Criée, S Sorichter, HJ Smith, P Kardos, R Merget, D Heise, D Berdel, D Köhler, H Magnussen, W Marek, et al. 2011. Body plethysmography–its principles and clinical use. *Respir Med* 105, 7 (Jul 2011), 959–971.
[21] Amandine Dubois and François Charpillet. 2014. A gait analysis method based on a depth camera for fall prevention. *Annu Int Conf IEEE Eng Med Biol Soc* 2014 (2014), 4515–4518.
[22] Marcelo Fernandes, Alberto Cukier, and Maria Ignêz Zanetti Feltrim. 2011. Efficacy of diaphragmatic breathing in patients with chronic obstructive pulmonary disease. *Chronic respiratory disease* 8, 4 (2011), 237–244.
[23] Yanbin Gong, Qian Zhang, Bobby H.P. NG, and Wei Li. 2022. BreathMentor: Acoustic-Based Diaphragmatic Breathing Monitor System. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 53 (jul 2022), 28 pages.
[24] Brian L Graham, Irene Steenbruggen, Martin R Miller, Igor Z Barjaktarevic, Brendan G Cooper, Graham L Hall, Teal S Hallstrand, David A Kaminsky, Kevin McCarthy, Meredith C McCormack, et al. 2019. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med* 200, 8 (Oct 2019), e70–e88.
[25] Wilhelmus Johannes Andreas Grooten, Lisa Sandberg, John Ressman, Nicolas Diamantoglou, Elin Johansson, and Eva Rasmussen-Barr. 2018. Reliability and validity of a novel Kinect-based software program for measuring posture, balance and side-bending. *BMC Musculoskelet Disord* 19, 1 (Jan 2018), 6.
[26] David MG Halpin, Gerard J Criner, Alberto Papi, Dave Singh, Antonio Anzueto, Fernando J Martinez, Alvar A Agusti, and Claus F Vogelmeier. 2021. Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. The 2020 GOLD science committee report on COVID-19 and chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 203, 1 (2021), 24–36.
[27] Hidetaka Hamasaki. 2020. Effects of Diaphragmatic Breathing on Health: A Narrative Review. *Medicines (Basel)* 7, 10 (Oct 2020).
[28] Tian Hao, Chongguang Bi, Guoliang Xing, Roxane Chan, and Linlin Tu. 2017. MindfulWatch: A Smartwatch-Based System For Real-Time Respiration Monitoring During Meditation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3, Article 57 (Sept. 2017), 19 pages.
[29] James M. Harte, Christopher K. Golby, Johanna Acosta, Edward F. Nash, Ercihan Kiraci, Mark A. Williams, Theodoros N. Arvanitis, and Babu Naidu. 2016. Chest wall motion analysis in healthy volunteers and adults with cystic fibrosis using a novel Kinect-based motion tracking system. *Medical & Biological Engineering & Computing* 54, 11 (2016), 1631–1640.
[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[31] Anne E Holland, Catherine J Hill, Alice Y Jones, and Christine F McDonald. 2012. Breathing exercises for chronic obstructive pulmonary disease. *Cochrane Database of Systematic Reviews* 10 (2012).
[32] Wakana Imano, Kenichi Kameyama, Malene Hollingdal, Jens Refsgaard, Knud Larsen, Cecilie Topp, Sissel Højsted Kronborg, Josefine Dam Gade, and Birthe Dinesen. 2020. Non-Contact Respiratory Measurement Using a Depth Camera for Elderly People. *Sensors* 20, 23 (Jan. 2020), 6901. Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.
[33] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su. 2018. Towards Environment Independent Device Free Human Activity Recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, New Delhi India, 289–304.
[34] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643* (2023).
[35] Marek Kowalski, Jacek Naruniec, and Michal Daniluk. 2015. Livescan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors. In *2015 International Conference on 3D Vision*. 318–325.

[36] Gregorij Kurillo, Evan Hemingway, Mu-Lin Cheng, and Louis Cheng. 2022. Evaluating the accuracy of the azure kinect and kinect v2. *Sensors* 22, 7 (2022), 2469.

[37] Yi Li, Hongyu Qian, Kewei Yu, Ying Huang, et al. 2020. Nonadherence in home-based pulmonary rehabilitation program for COPD patients. *Canadian respiratory journal* 2020 (2020).

[38] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 56 (jun 2019), 22 pages.

[39] Yue-Der Lin, Ya-Hsueh Chien, and Yi-Sheng Chen. 2017. Wavelet-based embedded algorithm for respiratory rate estimation from PPG signal. *Biomedical Signal Processing and Control* 36 (2017), 138–145.

[40] Haipeng Liu, John Allen, Dingchang Zheng, and Fei Chen. 2019. Recent development of respiratory rate measurement technologies. *Physiological measurement* 40, 7 (2019), 07TR01.

[41] Alvaro Lopez Paredes, Qiang Song, and Miguel Heredia Conde. 2023. Performance Evaluation of State-of-the-Art High-Resolution Time-of-Flight Cameras. *IEEE Sensors Journal* 23, 12 (2023), 13711–13727.

[42] Ilya Loshchilov and Frank Hutter. 2017. Fixing Weight Decay Regularization in Adam. *ArXiv* abs/1711.05101 (2017).

[43] Carlo Massaroni, Daniel Simões Lopes, Daniela Lo Presti, Emiliano Schena, and Sergio Silvestri. 2018. Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: A single camera approach. *Journal of Sensors* 2018 (2018).

[44] David J Meredith, D Clifton, Peter Charlton, J Brooks, CW Pugh, and L Tarassenko. 2012. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. *J Med Eng Technol* 36, 1 (Jan 2012), 1–7.

[45] Phuc Nguyen, Xinyu Zhang, Ann Halbower, and Tam Vu. 2016. Continuous and fine-grained breathing volume monitoring from afar using wireless signals. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. 1–9.

[46] American Association of Cardiovascular & Pulmonary Rehabilitation. 2011. *Guidelines for pulmonary rehabilitation programs*. Human Kinetics.

[47] KyeongTaek Oh, Cheung Soo Shin, Jeongmin Kim, and Sun K. Yoo. 2019. Level-Set Segmentation-Based Respiratory Volume Estimation Using a Depth Camera. *IEEE Journal of Biomedical and Health Informatics* 23, 4 (July 2019), 1674–1682. Conference Name: IEEE Journal of Biomedical and Health Informatics.

[48] Yulia Orlova, Alexander Gorobtsov, Oleg Sychev, Vladimir Rozaliev, Alexander Zubkov, and Anastasia Donsckaia. 2023. Method for Determining the Dominant Type of Human Breathing Using Motion Capture and Machine Learning. *Algorithms* 16, 5 (2023), 249.

[49] Sarah Ostadabbas, Nordine Sebkhi, Mingxi Zhang, Salman Rahim, Larry J. Anderson, Frances Eun-Hyung Lee, and Maysam Ghovanloo. 2016. A Vision-Based Respiration Monitoring System for Passive Airway Resistance Estimation. *IEEE Transactions on Biomedical Engineering* 63, 9 (Sept. 2016), 1904–1913. Conference Name: IEEE Transactions on Biomedical Engineering.

[50] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. 2011. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Transactions on Biomedical Engineering* 58, 1 (2011), 7–11.

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 234–241.

[52] Thayla A Santino, Gabriela SS Chaves, Diana A Freitas, Guilherme AF Fregonezi, and Karla MPP Mendonça. 2020. Breathing exercises for adults with asthma. *Cochrane Database Syst Rev* 3, 3 (Mar 2020), CD001277.

[53] Christopher G. Scully, Jinseok Lee, Joseph Meyer, Alexander M. Gorbach, Domhnull Granquist-Fraser, Yitzhak Mendelson, and Ki H. Chon. 2012. Physiological Parameter Monitoring from Optical Recordings With a Mobile Phone. *IEEE Transactions on Biomedical Engineering* 59, 2 (2012), 303–306.

[54] Pragya Sharma, Xiaonan Hui, Jianlin Zhou, Thomas B Conroy, and Edwin C Kan. 2020. Wearable radio-frequency sensing of respiratory rate, respiratory volume, and heart rate. *NPJ Digit Med* 3 (2020), 98.

[55] Charles Sharp, Vahid Soleimani, Sion Hannuna, Massimo Camplani, Dima Damen, Jason Viner, Majid Mirmehdi, and James W. Dodd. 2017. Toward Respiratory Assessment Using Depth Measurements from a Time-of-Flight Sensor. *Frontiers in Physiology* 8 (2017).

[56] Chen-Hsuan (Iris) 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (Dec. 2019), 1–30.

[57] Vahid Soleimani, Majid Mirmehdi, Dima Damen, Massimo Camplani, Sion Hannuna, Charles Sharp, and James Dodd. 2018. Depth-Based Whole Body Photoplethysmography in Remote Pulmonary Function Testing. *IEEE Transactions on Biomedical Engineering* 65, 6 (June 2018), 1421–1431. Conference Name: IEEE Transactions on Biomedical Engineering.

[58] Vahid Soleimani, Majid Mirmehdi, Dima Damen, James Dodd, Sion Hannuna, Charles Sharp, Massimo Camplani, and Jason Viner. 2017. Remote, Depth-Based Lung Function Assessment. *IEEE Transactions on Biomedical Engineering* 64, 8 (Aug. 2017), 1943–1958. Conference Name: IEEE Transactions on Biomedical Engineering.

[59] Vahid Soleimani, Majid Mirmehdi, Dima Damen, Sion Hannuna, Massimo Camplani, Jason Viner, and James Dodd. 2015. Remote pulmonary function testing using a depth sensor. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 1–4.

[60] Xiao Sun, Li Qiu, Yibo Wu, Yeming Tang, and Guohong Cao. 2017. SleepMonitor: Monitoring Respiratory Rate and Body Position During Sleep Using Smartwatch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 104 (sep 2017), 22 pages.

[61] Hiroki Takamoto, Hiroki Nishine, Shohei Sato, Guanghao Sun, Sadao Watanabe, Kim Seokjin, Masahito Asai, Masamichi Mineshita, and Takemi Matsui. 2020. Development and Clinical Application of a Novel Non-contact Early Airflow Limitation Screening System Using an Infrared Time-of-Flight Depth Image Sensor. *Frontiers in Physiology* 11 (2020).

[62] Michal Tölgyessy, Martin Dekan, L'uboš Chovanec, and Peter Hubinskỳ. 2021. Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors* 21, 2 (2021), 413.

[63] Luigi Vetrugno, Giovanni Maria Guadagnin, Federico Barbariol, Nicola Langiano, Alberto Zangrillo, and Tiziana Bove. 2019. Ultrasound Imaging for Diaphragm Dysfunction: A Narrative Literature Review. *J Cardiothorac Vasc Anesth* 33, 9 (Sep 2019), 2525–2536.

[64] Tianben Wang, Zhishen Wang, Xiantao Liu, Wenbo Liu, Leye Wang, Yuanqing Zheng, Jin Hu, Tao Gu, and Daqing Zhang. 2023. OmniResMonitor: Omnimonitoring of Human Respiration using Acoustic Multipath Reflection. *IEEE Transactions on Mobile Computing* (2023), 1–14.

[65] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 170 (jan 2018), 20 pages.

[66] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.

[67] Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2023. EarSpiro: Earphone-based Spirometry for Lung Function Assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (Jan. 2023), 188:1–188:27.

[68] Wentao Xie, Runxin Tian, Jin Zhang, and Qian Zhang. 2021. Noncontact Respiration Detection Leveraging Music and Broadcast Signals. *IEEE Internet of Things Journal* 8, 4 (Feb. 2021), 2931–2942. Conference Name: IEEE Internet of Things Journal.

[69] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-Grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services* (Seoul, Republic of Korea) *(MobiSys '19)*. Association for Computing Machinery, New York, NY, USA, 54–66.

[70] Yangfan Xu, Meiqinzi Tong, Wai-Kit Ming, Yangyang Lin, Wangxiang Mai, Weixin Huang, and Zhuoming Chen. 2021. A Depth Camera-Based, Task-Specific Virtual Reality Rehabilitation Game for Patients With Stroke: Pilot Usability Study. . *JMIR Serious Games* 9, 1 (Mar 2021), e20916.

[71] Wellington P Yamaguti, Renata C Claudino, Alberto P Neto, Maria C Chammas, Andrea C Gomes, João M Salge, Henrique T Moriya, Alberto Cukier, and Celso R Carvalho. 2012. Diaphragmatic breathing training program improves abdominal motion during natural breathing in patients with chronic obstructive pulmonary disease: a randomized controlled trial. *Arch Phys Med Rehabil* 93, 4 (Apr 2012), 571–577.

[72] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* 10, 2, Article 12 (jan 2019), 19 pages.

[73] Zhicheng Yang, Parth H. Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2016. Monitoring Vital Signs Using Millimeter Wave. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (Paderborn, Germany) *(MobiHoc '16)*. Association for Computing Machinery, New York, NY, USA, 211–220.

[74] Katherine Ka-Yin Yau and Alice Yuen Loke. 2021. Effects of diaphragmatic deep breathing exercises on prehypertensive or hypertensive adults: A literature review. *Complementary Therapies in Clinical Practice* 43 (May 2021), 101315.

[75] Masami Yokogawa, Tomoyo Kurebayashi, Toshikazu Ichimura, Manabu Nishino, Hiroichi Miaki, and Takao Nakagawa. 2018. Comparison of two instructions for deep breathing exercise: non-specific and diaphragmatic breathing. *Journal of Physical Therapy Science* 30, 4 (2018), 614–618.

[76] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. 2021. MoRe-Fi: Motion-Robust and Fine-Grained Respiration Monitoring via Deep-Learning UWB Radar. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems* (Coimbra, Portugal) *(SenSys '21)*. Association for Computing Machinery, New York, NY, USA, 111–124.