Supplementary materials

of "Prototype-driven Hard-sample Contrastive Learning for Camera-based Respiratory Imaging Analysis"

---

1 Clinical application examples of respiratory spatial pattern

The observation of regional respiratory patterns is useful for diverse clinical scenarios:

(i) physicians assess the symmetry of respiratory intensity between the left and right lungs through palpation to evaluate the patient's recovery status [1, 2].

(ii) In neonatal intensive care units (NICU), preterm infants infected by respiratory diseases (e.g., pneumonia) will change from abdominal respiration to thoracoabdominal asynchronous respiration. The proposed metric is a clear biomarker that indicates abnormal respiratory function and prompts clinical interventions [3].

(iii) In sleep centers, the presence of thoracoabdominal respiratory effort is a key phenomenon to differentiate between obstructive and central sleep apnea, which guides completely different treatment strategies [4].

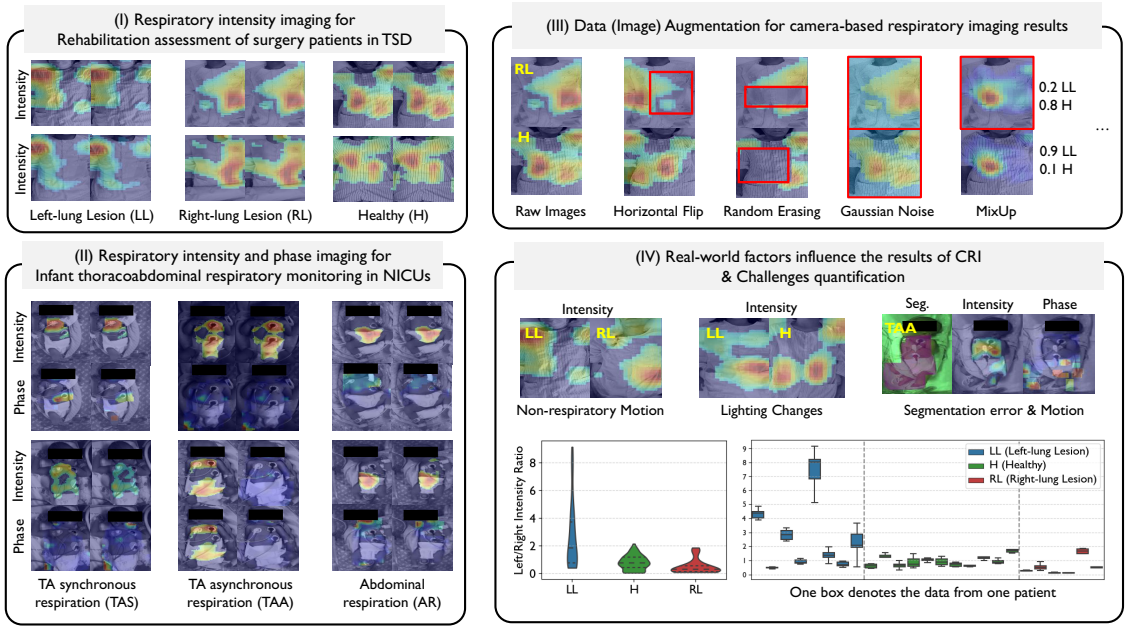


Figure 1: (I) and (II) define the assessment task of camera-based respiratory imaging in two clinical scenarios: thoracic surgery department and neonatal intensive care unit. Here for each state (class), two respiratory imaging results from 2 patients are presented to point out the task-specific challenges, including individual respiratory variations (i.e., different respiratory patterns in the same state) and limited clinical data (i.e., limited available subjects where each one has similar respiratory patterns in a short time). (III) shows that data augmentation destroys the semantic information of respiratory spatial patterns, marked by the red box. (IV) exemplifies hard samples generated with the real-world factors, and the quantification of individual differences and data periodicity, where the left-right lung intensity ratio of multiple continuous respiratory imaging results for each patient is visualized, indicating the high intra-subject consistency but high inter-subject variability of this task.

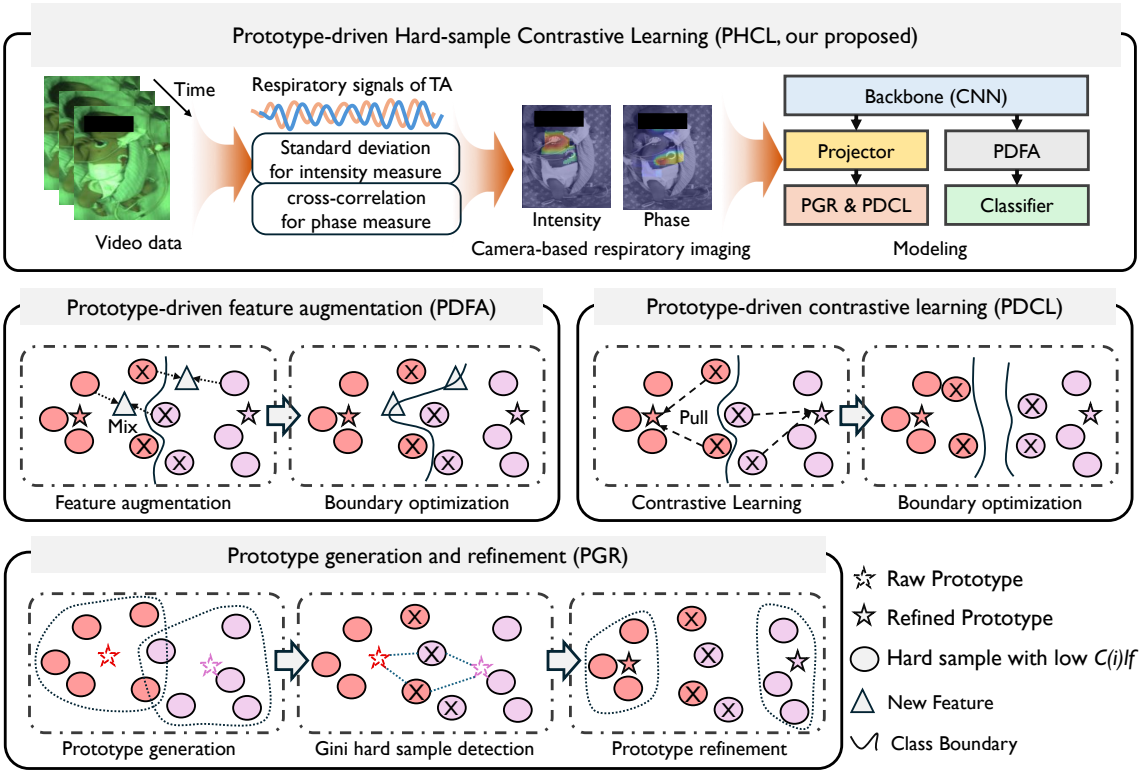


Figure 2: The framework of our proposed prototype-driven hard-sample contrastive learning (PHCL).

2 Motivation of our proposed PHCL

Previous works introduced contrastive learning methods combined with customized image (data) augmentation to enhance model robustness [5, 6]. However, we consider that data augmentation at the input level may disrupt the semantic integrity of irregular or asymmetric physiological activities, potentially hindering the model's performance. As shown in Fig. 1, geometric transformations may distort the underlying physiological structure. For instance, translation, flipping, and scaling can change the spatial location and size of the lesion, leading to a mismatch in critical semantic information. Similarly, random erasing [7] may introduce an artificial asymmetry to the respiratory pattern (see Fig. 1). The reason behind these phenomena is that the common data augmentation offered by the computer vision community does not consider the pathological reality. Such augmentations are especially problematic when analyzing irregular or asymmetric respiratory patterns, as they mislead and undermine the model's ability to interpret the data. In other words, traditional augmentation methods do not reflect the challenges or degradation of data in clinics.

3 Model implementation of our proposed PHCL

Working condition. As shown in Fig. 2, PHCL is working with the following conditions: given the respiratory imaging dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, the i-th sample has the input x_i and label y_i. The model, consisting of a backbone $B(\cdot)$, a classifier $C(\cdot)$, and a projector $P(\cdot)$, aims to predict the $\hat{y}_i$ label of the x_i sample, which can be mathematically expressed.

$$e_i = B(x_i|w_b), \tag{1}$$

$$z_i = P(e_i|w_p), \tag{2}$$

$$\hat{y}_i = C(e_i|w_c), \tag{3}$$

where w denotes the learnable parameter whose subscript is the part to which it belongs. e_i and z_i denote the embedding feature and contrast-space feature of the i-th sample, respectively. This operation follows [8, 9] that maps e onto the contrast space to calculate the contrastive loss, which helps reduce overfitting.

Model implementation. In the thoracic surgery department (TSD), the backbone consists of four convolutional layers with 16, 32, 64, and 128 kernels, utilizing the ReLU activation function and a 2×2 max pooling layer for dimensionality reduction. The backbone extracted feature embeddings from the intensity imaging. Then the embeddings were fed to the classifier that has a dense layer with three units for predicting the health state. The projector has two dense layers, both with 32 units. To train the model by PHCL, α_1 and α_2 are 0.5 and 0.1, and both τ_1 and τ_2 are 0.07 according to the empirical setting.

In the neonatal intensive care units (NICU), the backbone consisted of three convolutional layers with 16, 32, and 64 kernels, using the ReLU activation function and a 2×2 max pooling layer for extracting features from intensity and phase imagers, respectively. The extracted features were concatenated to form the final feature embeddings. The classifier has one dense layer with three units to predict infant respiratory patterns. To allow a fine-grained comparison, two same projectors with one dense layer of 32 units are employed for prototype generation of the intensity and phase imagers. To train the model by PHCL, α_1 and α_2 are set to 0.3, and both τ_1 and τ_2 are 0.1.

For other hyperparameters, all methods used AdamW optimizer with 1e-4 learning rate, 100 epochs, and 64 mini-batch in both clinical scenarios.

Training time cost. Our experiments were conducted on an Ubuntu server equipped with an Intel(R) Xeon(R) Gold 6330 CPU and a single NVIDIA RTX 4090. The software stack included Python 3.8, PyTorch 1.13.1, and CUDA 11.7. In the TSD experiment and NICU experiment, the average training time of one epoch with a 64 mini-batch using PHCL is 1.35 and 1.42 seconds, respectively. It should be noted that the cost of training time only happens once offline, which is not equal to the model's online inference time. In practice, the average time for our method to predict the label of one sample is 0.0014 seconds for the TSD experiment, and 0.0026 seconds for the NICU experiment. The time variation between these two experiments is because the former only uses the intensity images combined with one backbone for feature extraction, while the latter requires the intensity and phase images equipped with two backbones.

4 Clinical trials of camera-based respiratory imaging

We conducted the following clinical trials to evaluate the effectiveness of camera-based respiratory imaging and our proposed PHCL. The detailed subject information is summarized in Table 1, and the clinical acquisition process is introduced below.

Rehabilitation assessment in TSD. The clinical trial was approved by the Medical Ethical Committee of The Third People's Hospital of Shenzhen (No.2022-080-02) and conducted in the Department of Thoracic Surgery of the hospital. Video data of patients were collected using the camera of HUAWEI P20 Pro, capturing multiple 1-minute videos comprising 30 seconds of eupnea and 30 seconds of deep breathing. The videos were recorded at 1280×720 pixels with 30 frames per second (FPS). A total of 45 subjects were recruited, with 17 excluded since they only used abdominal respiration or had overlapping data across two periods; therefore, the clinical dataset

Table 1: Demographic details of patients and infants in our clinical trials.

Characteristic	**Patients from thoracic surgery department**	**Infants from neonatal intensive care units**
Subject	45 (11 LL, 14 RL, 20 H)	44 (17 AR, 12 TAS, 15 TAA)
Gender	25 Male, 20 Female	26 Male, 18 Female
Age (years)	43.68 ± 18.30	–
Gestational age (weeks)	–	36 ± 3
Height (cm)	167.15 ± 7.94	46.77 ± 3.92
Weight (kg)	60.06 ± 11.53	2.58 ± 0.61
Disease types	Tuberculosis-related diseases (e.g., tuberculous pleurisy, cervical tuberculous lymphadenitis), pneumothorax, pulmonary nodules, empyema, lung cancer, and others	Prematurity, low birth weight, neonatal respiratory distress syndrome, perinatal complications (e.g., fetal distress, premature rupture of membranes), and others

LL, left-lung lesion; RL, right-lung lesion; H, healthy; AR, abdominal; TAS, synchronized; TAA, asynchronous.

consists of 28 subjects. To ensure the inclusion of at least one breathing cycle for camera-based respiratory imaging, a sliding window with 6 second length and 3 second stride was employed. For the clinical classification task, the imaging data were labeled into three categories based on the gold standard of clinical diagnosis (X-rays or computerized tomography): healthy, left-lung lesion, and right-lung lesion.

Thoracoabdominal respiratory monitoring in NICUs. This clinical observational trial was approved by the Institutional Review Boards of Nanfang Hospital of Southern Medical University (No. NFEC-2022-100) and The Third People's Hospital of Shenzhen (No. 20220702), and conducted in the NICUs of both hospitals. Video data were collected using an IDS-UI-3860 RGB camera, capturing over 10 minutes at resolutions of 484×274 or 480×300 pixels with 60 FPS. A total of 44 infants were recruited for the experiment. Considering the variability in preterm infant respiration, a sliding window with 15 second length and 8 second stride was employed for camera-based respiratory imaging. Based on clinical knowledge [10], the data were labeled by our medical team into three categories: abdominal respiration, thoracoabdominal synchronous respiration, and thoracoabdominal asynchronous respiration.

5 Ablation analysis of PHCL

In order to analyze the effectiveness of PHCL, the ablation experiment is conducted to train the model by one of the core components, i.e., PDFA and PDCL. We have the following findings.

PDFA evenly enhances the model's ability to identify each class by structuring the transition region between different classes using their hard samples. Table 2 shows that PDFA has a higher score in the composite-index F1 and a more balanced performance between SEN and SPE compared to baseline and other feature augmentations, indicating that PDFA performs well in the overall classification performance. To illustrate this, the confusion matrices of those methods are shown in Fig. 3. Compared to the baseline, PDFA effectively increases most diagonal values, suggesting improved classification accuracy for each class in both tasks, while also reducing deviations across classes, resulting in a more balanced performance. Such improvement is likely because other feature

Table 2: The average performance (%) of subject-wise five-fold cross-validation of data augmentation and feature augmentation. C denotes the consistent performance change in ACC in both tasks compared to the baseline, where ↑ and ↓ denote the ascent and descent, respectively.

Aug.	Method	Rehabilitation assessment in TSD						Thoracoabdominal respiratory monitoring in NICUs						C
		ACC	REC	PRE	F1	SEN	SPE	ACC	REC	PRE	F1	SEN	SPE	
-	Baseline	64.76	64.76	67.98	58.43	65.76	61.13	71.97	66.62	69.58	66.69	61.94	80.14	-
Data	Vertical flip	49.79	50.35	50.10	43.19	38.97	64.25	59.46	56.43	58.84	56.00	52.82	64.58	↓
	Horizontal flip	42.52	45.62	45.25	40.16	46.93	37.79	71.87	67.56	70.28	67.51	60.26	83.70	↓
	Random erasing	56.99	54.47	54.58	49.43	49.45	66.82	70.50	66.49	69.92	66.31	61.37	78.45	↓
	CutMix [11]	57.70	53.59	53.47	46.30	51.40	59.11	71.88	66.93	71.75	66.48	55.47	**91.00**	↓
	Gaussian noise	66.52	66.27	67.64	60.91	67.57	63.75	71.54	66.57	69.93	66.68	59.75	82.67	-
	Rotation	66.04	65.41	74.63	62.04	62.77	69.82	72.50	67.80	70.37	68.00	62.89	80.61	↑
	MixUp [12]	67.47	65.17	65.04	60.99	62.84	73.34	72.59	66.74	70.48	66.79	60.24	84.63	↑
Feature	RSE [13]	67.14	64.71	65.62	63.44	63.32	70.68	72.43	67.20	70.00	67.28	62.61	80.67	↑
	Manifold MixUp [14]	69.97	70.52	72.53	67.81	73.61	64.69	72.59	67.52	70.59	67.72	62.32	81.25	↑
	ProtoMix [15]	71.55	73.73	71.28	68.55	80.21	58.11	73.23	67.82	70.76	67.66	63.51	81.89	↑
	FeatMatch [16]	71.89	69.09	73.28	68.19	66.30	**79.48**	73.31	67.89	70.18	68.00	61.76	84.40	↑
	PDFA (ours)	73.92	70.75	72.19	69.07	71.09	75.63	73.68	69.46	71.04	69.31	67.95	76.94	↑
	PDCL (ours)	69.39	69.93	69.13	65.42	76.67	57.57	73.34	69.72	70.85	69.79	69.47	74.79	↑
	PHCL (ours)	**76.87**	**78.98**	**79.34**	**76.54**	**85.75**	64.90	**77.08**	**73.71**	**75.16**	**73.61**	**73.26**	79.24	↑

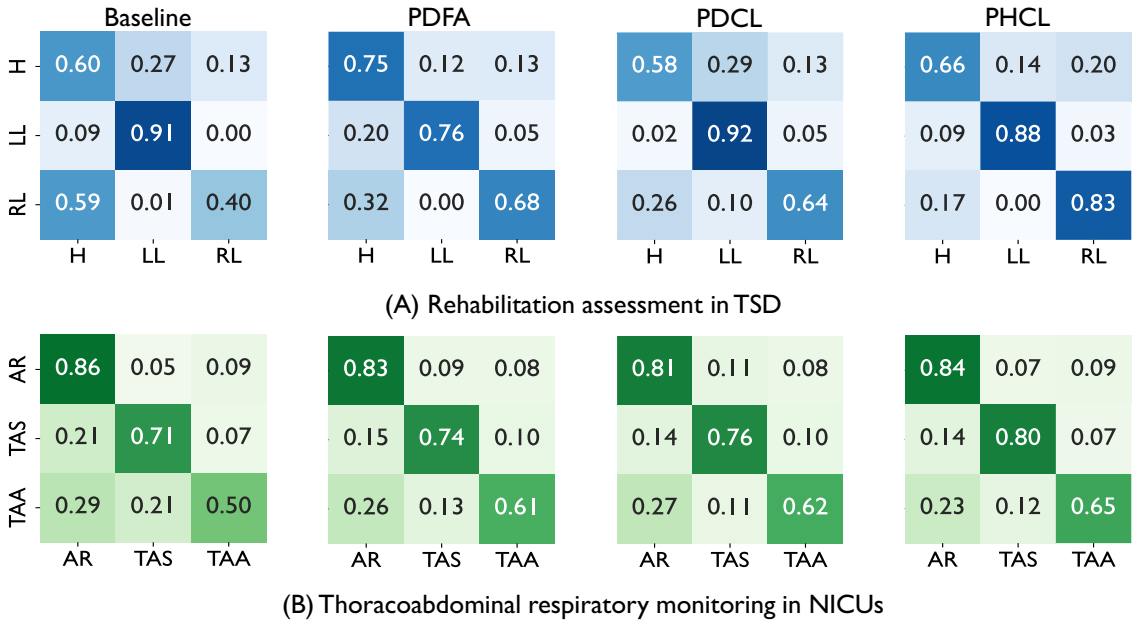
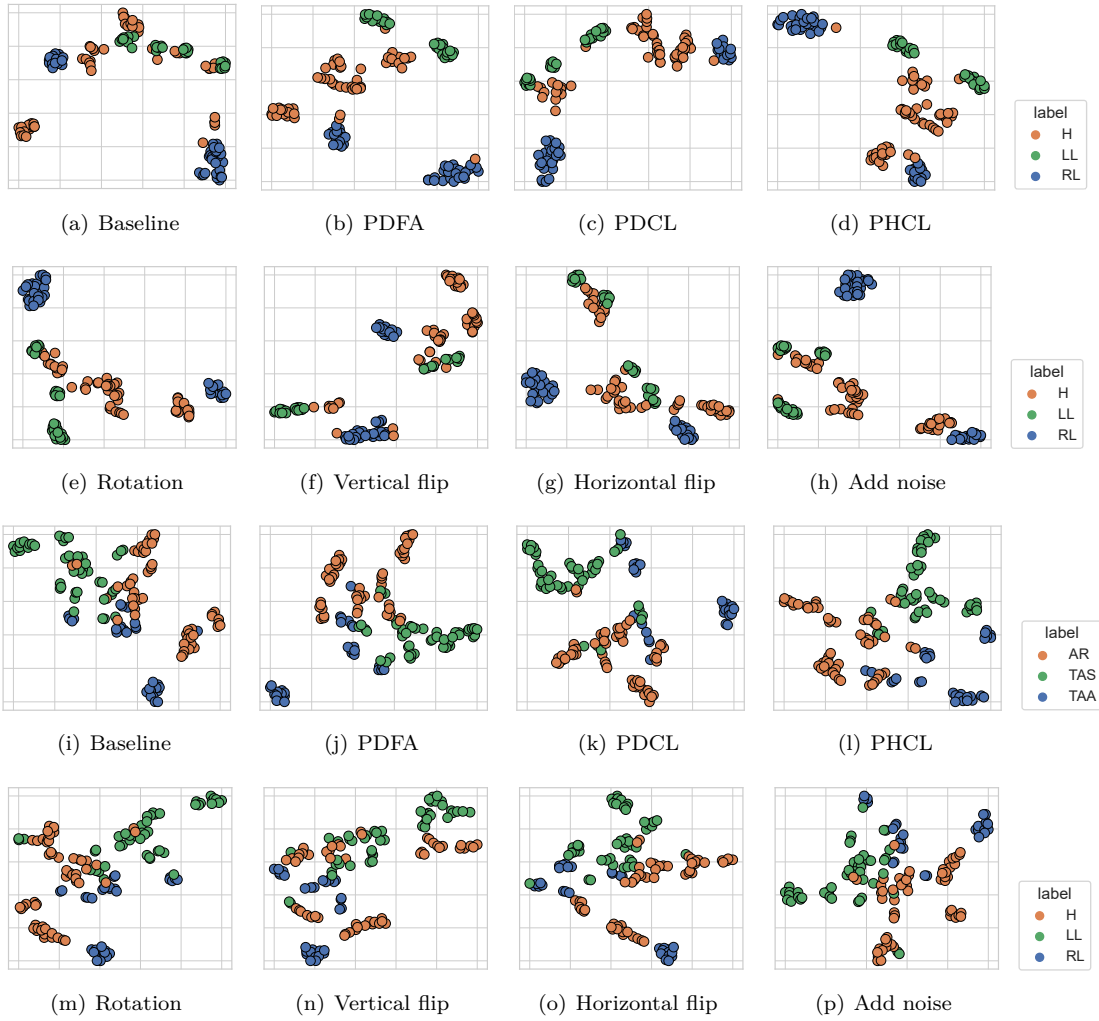


Figure 3: The confusion matrices of baseline, PDFA, PDCL, and PHCL methods in *(A)* rehabilitation assessment of TSD and *(B)* thoracoabdominal respiratory monitoring in NICUs.

augmentation methods did not emphasize the confusing degree of samples, where the features of one mini-batch are randomly taken to generate new features, leading to unbalanced performance similar to the baseline. In contrast, PDFA puts more attention on the hard samples to better structure the transition regions between different classes, thereby improving the model's ability to generalize across all classes.

PDCL enhances the model's capability to recognize abnormal respiratory states by more tightly clustering same-class samples in the representation space. Table 2 shows that PDCL outperforms baseline and other feature augmentation methods in distinguishing abnormal respiratory states in both tasks. Specifically, it obtains higher ACC and SEN while maintaining similar SPE. This conclusion can also be drawn from Fig. 3. This performance likely stems from PDCL's focus on enhancing the similarity of same-class samples extracted from different subjects in the representation



(a) Baseline (b) PDFA (c) PDCL (d) PHCL

(e) Rotation (f) Vertical flip (g) Horizontal flip (h) Add noise

(i) Baseline (j) PDFA (k) PDCL (l) PHCL

(m) Rotation (n) Vertical flip (o) Horizontal flip (p) Add noise

Figure 4: The t-SNE visualization of embedding obtained by the baseline, PDFA, PDCL, and PHCL, as well as the data augmentation methods, using the test set in the rehabilitation assessment task in TSD (*a-h*) and thoracoabdominal respiratory monitoring task in NICUs (*i-p*)

.

space. To validate this hypothesis, Fig. 4 visualizes the embeddings of the test set derived using different methods based on the t-SNE method. The results indicate that compared to the baseline and PDFA, as well as traditional data augmentation methods, PDCL achieves greater aggregation of same-class samples in the embedding space, while enhancing the separability of different-class samples, thus improving the model's accuracy and reliability.

PHCL integrates the characteristics of PDFA and PDCL to enhance the model's generalization capabilities. Table 2 shows that PHCL surpasses other methods across the classification metrics such as ACC, REC, PRE, and F1. Moreover, it maintains a more balanced or higher score in SEN and SPE across the two application scenarios. Fig. 3 and Fig. 4 illustrate how PHCL effectively integrates the advantages of PDFA and PDCL, leading to improvements in diagonal values (accuracy score), better compactness of same-class samples, and enhanced separability between different-class samples.

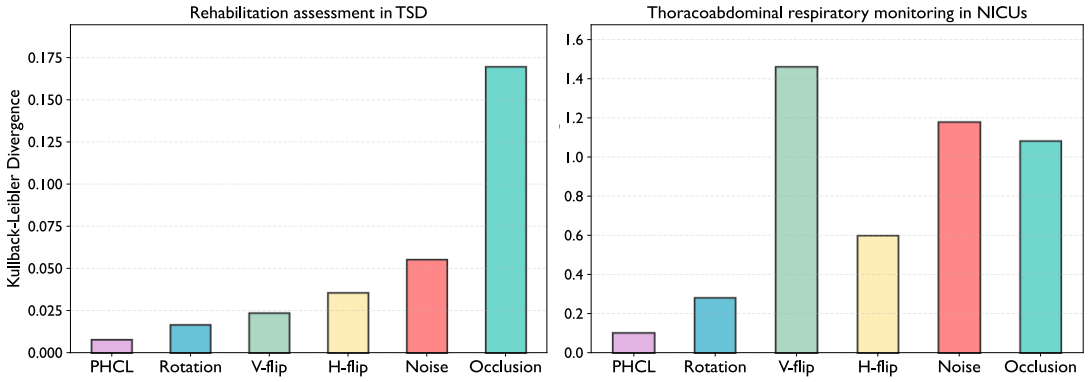


Figure 5: The Kullback-Leibler Divergence of predicted probability distribution between the raw input and its augmented version, including traditional data augmentation and our proposed PHCL (feature augmentation), to measure the semantic distortion.

6 Additional analysis of PHCL

In order to provide a more straightforward comparison (both qualitative and quantitative) between conventional data augmentation and our proposed feature augmentation in preserving semantics, we conducted an extra experiment that visualizes the mean of Kullback-Leibler Divergence (KLD) between the raw input and the augmented version in the test set. Here, KLD can measure the difference between two probability distributions, and the smaller score means a smaller difference; thereby, the semantic distortion between the raw input and the augmented version can be measured by KLD. Fig. 5 shows that using the trained baseline model (without semantic confusion) in two clinical scenarios, our proposed method has a smaller KLD value than conventional data augmentation methods, including Gaussian noise, flip, rotation, and other operations. This indicates that conventional data augmentation, operating at the input level, is more likely to undermine the key semantic information on which the model relies for judgment. In contrast, our proposed method, augmenting at the feature level, can better preserve the core semantic identity of the sample from being damaged.

In order to understand the differences between our approach and data augmentation, Grad-CAM [17] is employed to visualize the focus of the baseline model and PHCL to different-class data with/without data augmentation (horizontal flip) in the rehabilitation assessment task, as shown in Fig. 6. It shows that the baseline model learn the different patterns for three-class data. Here, it focuses on the middle or left-right symmetric regions for the data from the healthy group, and focuses on the non-lesion area for the data from the lesion group, such as focusing on the right lung breathing in the left lesion and vice versa. For data augmentation, Fig. 6 shows that horizontal flip is clearly disruptive to data semantics. There is an exchange of patterns between right and left lesions, i.e., after data augmentation of the left-lung lesion data, the pattern of the right lesion was found. As a result, the model trained with data augmentation is confused for both types of data, accounting for the performance decline. Compared to data augmentation, our proposed method operates in the feature space, maintaining a consistent focus pattern with the baseline while enhancing representation diversity.

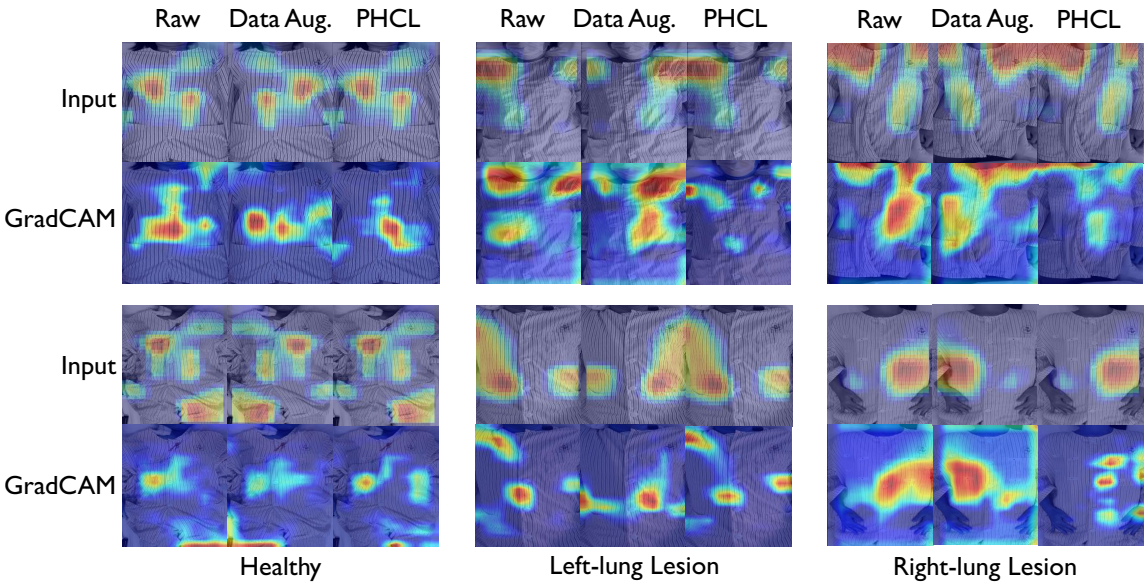


Figure 6: The GradCAM visualization of different-class data in the rehabilitation assessment task of TSD, highlights the decision basis of the baseline and PHCL models. *Raw* denotes the original input, *Data Aug.* denotes the horizontal-flip input.

7 Discussion of camera-based respiratory imaging and PHCL

Our work has some limitations and potentials, as discussed below:

(i) ***Merits and potentials***: (1) *more application scenarios.* We consider that camera-based respiratory imaging (CRI) has promising potential in sleep health monitoring, especially for sleep apnea detection and classification. Specifically, the absence of thoracoabdominal respiratory effort during sleep is a key indicator of central sleep apnea [18], and this phenomenon may be captured by the respiratory intensity imaging. In addition, it has been clinically reported that obstructive sleep apnea leads to the thoracoabdominal asynchronous respiratory [19], which can also be observed from the respiratory phase imaging. These cases indicate the utility of CRI in sleep apnea detection, which should be investigated;

(2) *generalizability to other camera-based tasks.* We consider that our proposed PHCL may have the potential to be generalized to other camera-based physiological monitoring tasks, especially when they face the same challenges, including the issues of data scarcity and individual differences. For example, it had been reported [20, 21] that camera-based blood pressure monitoring has performance degradation in the setting of cross-subject calibration; meanwhile, the stability of blood pressure made it difficult to observe a wider range of blood pressure variations during the data collection process. For camera-based sleep staging, due to the accessibility of polysomnography (PSG, gold standard) and privacy issues, it is difficult to collect large amounts of sleep video data; meanwhile, the patients' basic physiological parameters (e.g., heart rate and respiratory rate) vary under the same sleep stages [22, 23]. Therefore, since these camera-based monitoring tasks also face the issues of data scarcity and individual differences, our proposed PHCL may improve their generalizability by feature augmentation and contrastive learning.

(ii) ***Limitations of current work***: (1) *posture condition.* Our work was conducted under the condition that the patient/infant lies supine, and it remains unclear whether posture affects the stability of camera-based respiratory imaging results. We consider that posture changes may lead to significant alterations, such as a shift in the optical flow direction of respiratory signals. For instance, when the infant lies supine, the major respiratory motion component is physiologically

concentrated in the vertical direction, whereas it shifts to the horizontal direction when the infant lies on their side [24, 25];

(2) *surface texture dependency.* The thoracoabdominal respiratory motion is measured by the spatial and temporal gradients of a sequence of images, which is essentially dependent on the texture of the surface (chest and abdomen) [26]. It means that when the surface texture is not significant (i.e., lacks distinguishable features), the camera-based respiratory imaging method may struggle to accurately capture motion, leading to potential failure or reduced reliability. Here we consider that the structured light may solve this issue by projecting a known pattern (e.g., points, grids, or stripes) onto the surface, which creates artificial texture that can be reliably detected by the imaging system;

(3) *lack of motion robustness.* It should be noted that our work was conducted under the condition without large body motions since the system is limited by the assumption of optical flow-based motion measurement. Although this is acceptable for clinical spot-checks, the system may introduce a new module that can perform continuous respiratory spatio-temporal monitoring, such as motion separation. The above directions will be left as our future work to promote the clinical applications of camera-based respiratory imaging.

(iii) ***extra computational cost of PHCL***. Our method employed the embedding features of the training set to generate the prototypes at each epoch. When dealing with large clinical datasets, the computing costs will further increase. To reduce the computational costs, we may consider introducing the exponential moving average [27] in the future to update the prototype in each mini-batch. However, the compatibility of EMA and PHCL needs to be investigated.

References

[1] Malay Sarkar et al. "Auscultation of the respiratory system". In: *Annals of thoracic medicine* 10.3 (2015), pp. 158–168.

[2] Heidi Simpson. "Respiratory assessment". In: *British journal of nursing* 15.9 (2006), pp. 484–488.

[3] Deborah C Givan. "Physiology of breathing and related pathological processes in infants". In: *Seminars in Pediatric Neurology*. Vol. 10. 4. Elsevier. 2003, pp. 271–280.

[4] RB Berry et al. "Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events". In: *J Clin Sleep Med* 8.5 (2012), pp. 597–619.

[5] Dongmin Huang et al. "Camera-based respiratory imaging for intelligent rehabilitation assessment of thoracic surgery patients". In: *IEEE Internet of Things Journal* (2024).

[6] D. Huang et al. "Camera-Based Respiratory Imaging System for Monitoring Infant Thoracoabdominal Patterns of Respiration". In: *IEEE Journal of Biomedical and Health Informatics* (2024).

[7] Zhun Zhong et al. "Random erasing data augmentation". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. 07. 2020, pp. 13001–13008.

[8] Dongmin Huang et al. "A contrastive embedding-based domain adaptation method for lung sound recognition in children community-acquired pneumonia". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2023, pp. 1–5.

[9] Ting Chen et al. "A simple framework for contrastive learning of visual representations". In: *International conference on machine learning*. PmLR. 2020, pp. 1597–1607.

[10] Herbert C Miller and Franklin C Behrle. “Changing patterns of respiration in newborn infants”. In: *Pediatrics* 12.2 (1953), pp. 141–150.

[11] Sangdoo Yun et al. “Cutmix: Regularization strategy to train strong classifiers with localizable features”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 6023–6032.

[12] Hongyi Zhang et al. “mixup: Beyond empirical risk minimization”. In: *arXiv preprint arXiv:1710.09412* (2017).

[13] Xuanqing Liu et al. “Towards robust neural networks via random self-ensemble”. In: *Proceedings of the european conference on computer vision*. 2018, pp. 369–385.

[14] Vikas Verma et al. “Manifold mixup: Better representations by interpolating hidden states”. In: *International conference on machine learning*. PMLR. 2019, pp. 6438–6447.

[15] Yongxin Xu et al. “Protomix: Augmenting health status representation learning via prototype-based mixup”. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, pp. 3633–3644.

[16] Chia-Wen Kuo et al. “Featmatch: Feature-based augmentation for semi-supervised learning”. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer. 2020, pp. 479–495.

[17] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[18] SHAHROKH Javaheri and JA Dempsey. “Central sleep apnea”. In: *Comprehensive Physiology* 3.1 (2013), pp. 141–163.

[19] J Hammer and CJL Newth. “Assessment of thoraco-abdominal asynchrony”. In: *Paediatric respiratory reviews* 10.2 (2009), pp. 75–80.

[20] Yukai Huang et al. “Camera-based blood pressure monitoring based on multi-site and multi-wavelength pulse transit time features”. In: *IEEE Transactions on Instrumentation and Measurement* (2024).

[21] Zongshen Hou et al. “Exploiting Multi-wavelength Morphological Features of Camera-PPG for Blood Pressure Estimation”. In: *IEEE Transactions on Instrumentation and Measurement* (2025).

[22] Mathias Perslev et al. “U-Sleep: resilient high-frequency sleep staging”. In: *NPJ digital medicine* 4.1 (2021), p. 72.

[23] Qiongyan Wang, Hanrong Cheng, and Wenjin Wang. “Video-psg: An intelligent contactless monitoring system for sleep staging”. In: *IEEE Transactions on Biomedical Engineering* (2024).

[24] Yongshen Zeng et al. “A multi-modal clinical dataset for critically-ill and premature infant monitoring: Eeg and videos”. In: *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics*. IEEE. 2022, pp. 1–5.

[25] Anouk WJ Scholten et al. “Diaphragmatic electromyography in infants: an overview of possible clinical applications”. In: *Pediatric Research* 95.1 (2024), pp. 52–58.

[26] W. Wang et al. “Algorithmic insights of camera-based respiratory motion extraction”. In: *Physiological measurement* 43.7 (2022), p. 075004.

[27] Yuhang Zang, Chen Huang, and Chen Change Loy. “Fasa: Feature augmentation and sampling adaptation for long-tailed instance segmentation”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 3457–3466.