State-Aware Contactless Respiratory Volume Monitoring for Breath-Hold–Guided Radiotherapy

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Jiahao Yang, MS[1]; Xinye Ni, PhD[1]
[1]The Second People's Hospital of Changzhou, The Third Affiliated Hospital of Nanjing Medical University, Changzhou 213003, China

ABSTRACT

Purpose: Deep inspiration breath hold (DIBH) can reduce organ-at-risk dose in thoracic and abdominal radiotherapy, but clinical use requires accurate, real-time, and low-burden monitoring of breath-hold stability and tidal volume.

Methods: We developed a contactless tidal-volume estimation framework using a depth camera to capture three-dimensional thoracoabdominal surface motion. A state-aware, surrogate-guided TSMixer model was designed with dynamic state gating, steady-state anchor modeling, adaptive baseline correction, and subject-specific calibration. Synchronized depth-camera data and ABC-derived tidal-volume signals were collected from 22 volunteers and evaluated using subject-independent leave-one-subject-out cross-validation.

Results: The predicted tidal volume strongly correlated with the ABC reference signal (pooled r = 0.968), with an overall MAE of 0.098 L. Phase-specific MAE was 0.1059 L during breath hold and 0.0967 L during normal breathing. GPU-side latency was 12.55 ms per frame.

Conclusion: The proposed framework enables stable, contactless tidal-volume estimation and real-time visual feedback for DIBH-guided radiotherapy.

CONTACT

Xinye Ni
The Second People's Hospital of Changzhou, The Third Affiliated Hospital of Nanjing Medical University, Changzhou, China
No. 68, Gehu Middle Road, Wujin District, Changzhou City, Jiangsu Province, China
nxy@njmu.edu.cn
+86 13776856838

INTRODUCTION

Respiratory motion remains a major source of geometric uncertainty in thoracic and abdominal radiotherapy. In deep inspiration breath hold (DIBH), lung expansion can increase the separation between the heart and chest wall, thereby reducing cardiac and pulmonary dose. However, the clinical benefit of DIBH depends on stable breath-hold execution, reproducible respiratory state control, and reliable real-time monitoring.

Current respiratory monitoring approaches have important practical limitations. The Active Breathing Coordinator (ABC) system can provide tidal-volume information, but it requires a mouthpiece and nose clip and may increase patient burden. Pressure-belt systems require direct contact with the thoracoabdominal surface and are sensitive to placement and tightness. Surface-guided radiotherapy can monitor external surface position, but quantitative estimation of tidal volume from three-dimensional surface deformation remains challenging, especially across alternating free-breathing and breath-hold phases.

Therefore, this study developed a depth-camera-based, contactless respiratory-volume monitoring framework that converts thoracoabdominal surface motion into tidal-volume estimates for DIBH-guided radiotherapy.

METHODS AND MATERIALS

As shown in Fig. 1, synchronized respiratory data were acquired using an Intel RealSense D435i depth camera and an Active Breathing Coordinator (ABC) system during free breathing and DIBH. The depth camera captured three-dimensional thoracoabdominal surface motion, while the ABC-derived tidal-volume signal served as the reference for model development, subject-specific calibration, and performance evaluation. This prospective model-development and pilot internal-validation study included 22 volunteers and yielded 347,963 valid frames.

Depth sequences were aligned, cropped to a fixed thoracoabdominal ROI, temporally synchronized with ABC signals, and converted into a surface-based volumetric respiratory surrogate. As shown in Fig. 2, this surrogate was used as the input to a state-aware, surrogate-guided TSMixer architecture for tidal-volume estimation. The model incorporated dynamic state gating, steady-state anchor modeling, subject-specific calibration, and adaptive baseline correction to improve prediction stability across free-breathing and breath-hold phases.

Performance was evaluated using subject-independent leave-one-subject-out cross-validation. For each held-out subject, the first 20% of the recording was used for lightweight subject-specific calibration, and the remaining 80% was used for testing.

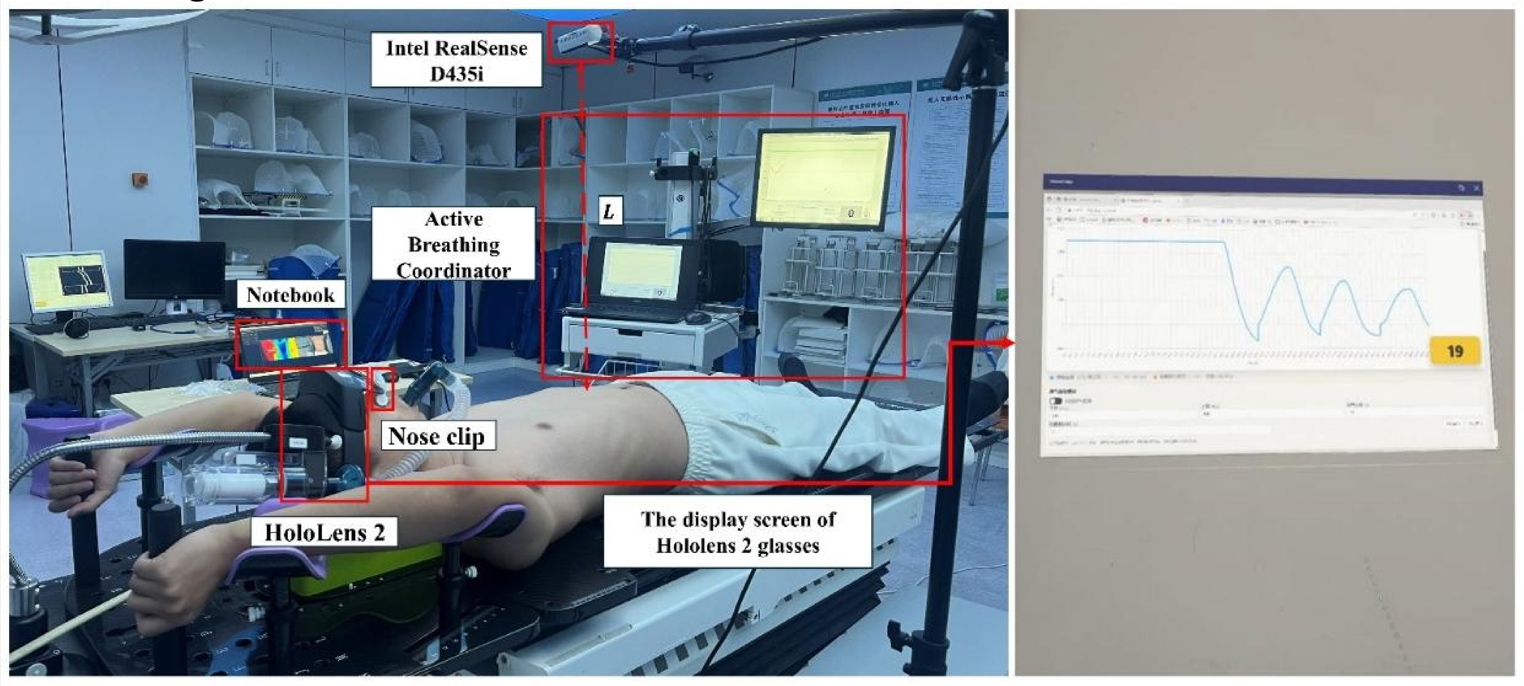

Fig. 1. Experimental setup for contactless respiratory monitoring.

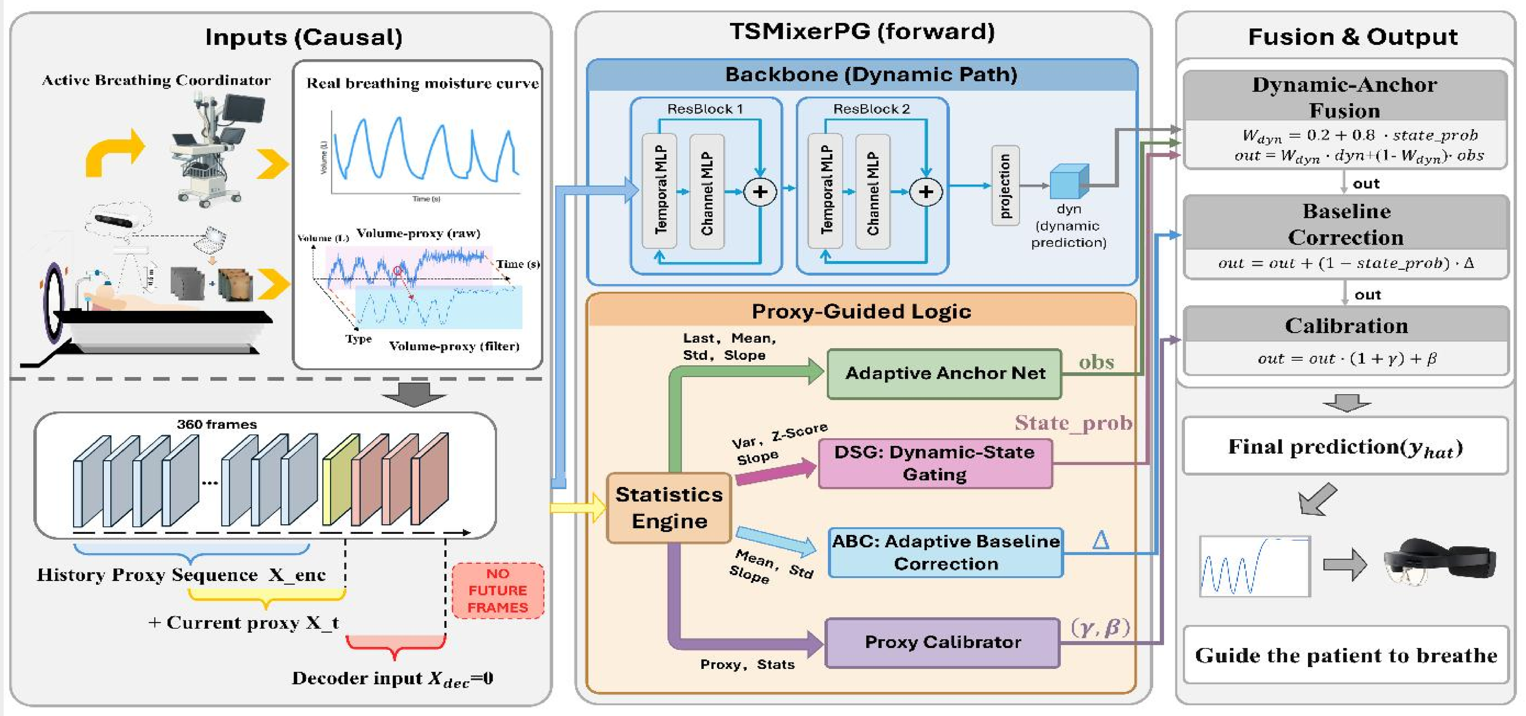

Fig. 2. State-aware, surrogate-guided TSMixer architecture integrating dynamic state gating, steady-state anchor modeling, subject-specific calibration, and adaptive baseline correction.

RESULTS

Table 1 summarizes the overall tidal-volume estimation performance across all test subjects for different models. The proposed model achieved the best overall performance, with RMSE = 0.141 L, MAE = 0.098 L, R^2 = 0.926, and Corr = 0.967. Compared with the baseline time-series models, the proposed framework substantially reduced prediction error and improved temporal agreement with the ABC-derived reference signal.

As shown in Fig. 3, the proposed model closely followed the ABC ground-truth tidal-volume signal during both normal breathing and DIBH. During the breath-hold phase, the proposed model maintained a stable tidal-volume plateau, whereas baseline model predictions showed larger fluctuations or drift. These results indicate that state-aware modeling and steady-state constraints improved prediction stability during the quasi-steady breath-hold state.

Phase-specific analysis further showed comparable performance during breath hold and normal breathing, with MAE/MAPE of 0.1059 L/17.27% and 0.0967 L/16.68%, respectively. The GPU-side algorithmic latency was 12.55 ms per frame, supporting real-time respiratory monitoring.

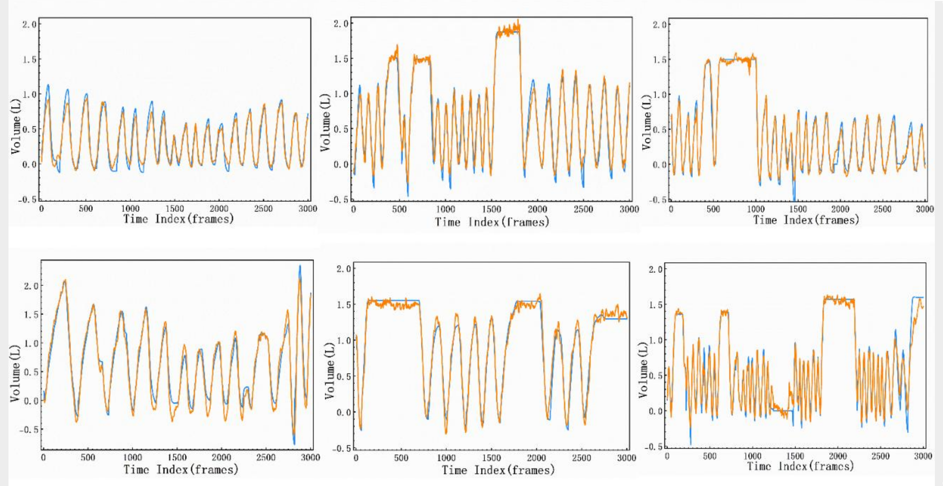

Fig. 3. Representative time-series comparison of tidal volume between the ABC ground truth and model predictions during normal breathing and DIBH.

model	RMSE(L)	MAE(L)	R2	corr
Dlinear[27]	0.346 ± 0.143	0.247 ± 0.067	0.601 ± 0.381	0.837 ± 0.118
LightTS[28]	0.332 ± 0.115	0.223 ± 0.071	0.566 ± 0.366	0.858 ± 0.086
MICN[30]	0.357 ± 0.118	0.268 ± 0.067	0.609 ± 0.239	0.828 ± 0.099
TiDE[31]	0.372 ± 0.142	0.276 ± 0.060	0.581 ± 0.438	0.825 ± 0.119
our model	0.141 ± 0.055	0.098 ± 0.031	0.926 ± 0.070	0.967 ± 0.027

Table 1. Overall tidal-volume estimation performance across all test subjects for different models.

CONCLUSIONS

This study demonstrates the feasibility of contactless tidal-volume estimation during DIBH using depth-camera-derived thoracoabdominal surface motion.

The proposed state-aware, surrogate-guided TSMixer framework achieved high agreement with ABC-derived reference tidal volume while maintaining stable prediction during both normal breathing and breath-hold phases.

By integrating dynamic state gating, steady-state anchoring, subject-specific calibration, and adaptive baseline correction, the framework reduced breath-hold drift and improved prediction stability.

The low GPU-side latency supports real-time respiratory monitoring and visual feedback, suggesting potential value as a low-burden complement to DIBH and surface-guided radiotherapy workflows.

DISCUSSION

The improved performance of the proposed model suggests that DIBH tidal-volume estimation cannot be fully addressed by generic time-series regression alone. Because respiratory signals alternate between dynamic free-breathing phases and quasi-steady breath-hold phases, explicit respiratory-state modeling is important for stable volume estimation.

The representative time-series results indicate that the proposed framework maintained a stable tidal-volume plateau during breath hold, whereas baseline models showed larger fluctuations or drift. This improvement is likely related to the combined effects of dynamic state gating, steady-state anchor modeling, subject-specific calibration, and adaptive baseline correction.

Clinically, the proposed framework may complement existing DIBH and surface-guided radiotherapy workflows by providing a quantitative respiratory-volume signal from contactless surface motion. Once a subject-specific surface-volume mapping is established, online monitoring could rely primarily on depth-camera input while reducing dependence on contact-based respiratory measurement devices.

This study remains a pilot internal-validation study in volunteers. Future work should evaluate the framework in patient cohorts, across treatment fractions, and in relation to internal organ or tumor motion.

REFERENCES

1. Zheng J, Gao Y, Zhang H, et al. OTSU Multi-Threshold Image Segmentation Based on Improved Particle Swarm Algorithm. Applied Sciences. 2022; 12(22):11514.
2. Fromberg L, Das S, Harder Clemmensen LK. Pre-processing Blood-Volume-Pulse for In-the-wild Applications. arXiv:2304.14186 [eess.SP]. 2023.
3. Paszke A, Gross S, Massa F, et al.PyTorch: An Imperative Style, High-Performance Deep Learning Library.Advances in Neural Information Processing Systems (NeurIPS), 2019: 8024–8035.