Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated 3D CT Segmentation of Intercostal Muscles Using Bayesian U-Net: a retrospective technical validation study with exploratory pulmonary function analysis

Yoko Murakami

monn@belle.shiga-med.ac.jp

Shiga University of Medical Science

Yukihiro Nagatani

Shiga University of Medical Science

Yuto Masaki

Nara Institute of Science and Technology Ikoma

Yoshito Otake

Nara Institute of Science and Technology Ikoma

Yoshinobu Sato

Nara Institute of Science and Technology Ikoma

Satoshi Ishimoto

Shiga University of Medical Science

Makoto Wakamiya

Nagahama City Hospital Nagahama

Yoshiyuki Watanabe

Shiga University of Medical Science

Research Article

Keywords: computed tomography, intercostal muscle, segmentation, deep learning, pulmonary function, Bayesian U-Net

Posted Date: August 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10641800/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Quantitative evaluation of respiratory muscles is increasingly important in thoracic imaging, yet intercostal muscles remain difficult to assess on computed tomography (CT) because of their thin, curved anatomy. We developed a Bayesian U-Net-based method for automated 3D intercostal muscle segmentation on thin-slice chest CT and explored associations between CT-derived muscle metrics and pulmonary function.

Methods

This retrospective technical validation study included 145 chest CT examinations: 129 from patients undergoing pulmonary tumor surgery for model development/internal validation and 16 from another institution for external validation. An uncertainty-based active-learning framework produced 60 manually corrected reference examinations. Segmentation performance was assessed using fivefold cross-validation and independent external validation against two reader-generated ground truth labels, including an additional full-volume validation of the bilateral sixth intercostal muscles. Associations with pulmonary function were evaluated in 95 examinations using Pearson correlation and exploratory multivariable linear regression adjusted for height, body mass index, age, and sex.

Results

Internal validation showed a Dice similarity coefficient (DSC) of 0.84 ± 0.08 and an average symmetric surface distance (ASSD) of 0.66 +/- 0.47 mm. In external validation, DSC values were 0.774 ± 0.043 and 0.796 ± 0.045 versus the two ground truth labels, comparable to the interobserver DSC of 0.775 ± 0.042. For the bilateral sixth intercostal muscles, DSC was 0.738 ± 0.025, ASSD was 0.921 ± 0.161 mm, and 95-percentile Housdorff distance (HD95) was 2.673 ± 0.220 mm. Intercostal muscle volume correlated positively with forced vital capacity (FVC; r = 0.568) and forced expiratory volume in 1 second (FEV1; r = 0.429). In exploratory adjusted analyses, total and lean intercostal muscle volumes remained positively associated with FVC, whereas the FEV1 association was attenuated.

Conclusions

Automated 3D segmentation of intercostal muscles on thin-slice chest CT is technically feasible. CT-derived intercostal muscle metrics showed biologically plausible exploratory associations with pulmonary function, particularly FVC, supporting future disease-focused respiratory muscle imaging studies.

Background

Pneumonia is a major cause of morbidity and mortality in older adults, particularly aspiration pneumonia, which requires sufficient cough strength for prevention. Coughing relies heavily on the respiratory muscles, including the intercostal muscles (IMs) as well as the diaphragm, which play essential roles in respiration and airway clearance (1, 2). Their clinical relevance is supported by evidence showing increased IM activity during exercise in COPD (3), FDG uptake associated with coughing (4), and associations between reduced IM cross sectional area and COPD severity or exacerbation risk in ultrasound CT or based studies (5–8).

Intercostal muscles are important but difficult to measure because they are thin, curved, segmental structures between ribs, and computed tomography (CT) attenuation is influenced by partial-volume effects and adjacent fat, pleura, and vessels. Accordingly, prior imaging studies have largely quantified its thickness by ultrasound or its area on CT at any dimension (5–11).

Three-dimensional (3D) quantification using CT images could provide a more comprehensive assessment, but manual whole-muscle segmentation is impractical because the muscles are small, numerous, and anatomically complex. Despite these findings, assessing IMs has been limited by technical challenges, including their small size and complex anatomy. Traditional imaging techniques such as computed tomography (CT) and ultrasound provide two-dimensional measurements (3, 5–7, 9–11) but lack comprehensive three-dimensional analysis because manual three-dimensional segmentation of IMs is problematic, inaccurate, time-consuming, and labor-intensive. Advances in artificial intelligence, particularly deep learning, offer promising solutions (12, 13). Notably, the Attention U-Net model has shown efficacy in segmenting small anatomical structures (14–17). Feasible 3D segmentation models were reported for, paraspinal muscles (12, 14), pectoralis muscle (13–15), and individual muscles around the hip joint (16–18). However, its application to IM analysis remains limited. In a previous study, Choi et al. reported that entire three-dimensional IMs were accurately segmented, and their measured volumes correlated with pulmonary function using a dataset from patients with asthma or chronic obstructive pulmonary diseases (COPD) but no severe underlying diseases(14).

This study aims to address this gap by developing an automatic segmentation method using a Bayesian U-Net model for IMs on CT images. We assessed its feasibility by evaluating its performance in clinical cases with lung parenchymal or pleural lesions, and the correlation between segmented muscle volumes and pulmonary function tests.

Methods

Study design, setting, and ethics

This retrospective technical validation study used chest CT examinations from three participating institutions. The study was approved by the institutional review boards of Shiga University of Medical Science, Nara Institute of Science and Technology, and Nagahama Municipal Hospital. (approval number

R2016-209). The requirement for informed consent was waived because of the retrospective design with opt-out disclosure.

Study datasets

Two datasets were used: a primary dataset for model development/internal validation and an external dataset for independent validation.

The primary dataset comprised 129 thin-slice chest CT examinations obtained at one institution between January 2015 and August 2020 in patients undergoing surgery for pulmonary tumors (110 non-enhanced and 19 contrast-enhanced examinations). Both preoperative and postoperative studies were included to test technical robustness in heterogeneous thoracic CT examinations containing masses, pleural effusions, pneumonia, and postoperative changes.

The external dataset consisted of 16 thin-slice chest CT examinations from another institution acquired between January and November 2019 (8 non-enhanced and 8 contrast-enhanced examinations). These cases were selected as reference cases because, while they were anatomically standardized, showing no significant asymmetry of the ribs or muscles or obvious abnormalities of the chest wall, they contained lesions in the lung fields.

Clinical characteristics and CT acquisition parameters are summarized in Tables 1 and 2. Reconstructed images had slice thicknesses of 0.5-1.0 mm and increments of 0.5–0.8 mm and were transferred in Digital Imaging and Communications in Medicine (DICOM) format to a dedicated workstation (Ziostation 2, Ziosoft, Tokyo, Japan) for manual segmentation, correction, and visual assessment.

Table 1
Clinical characteristics of study cohorts.

Number	Overall	Training and internal validation	External validation	*p*-value
	145	**129**	**16**	**N/A**
Age	65.7 (32–88)	67.2 (32–88)	59.6 (37–88)	0.033*
Gender (F:M)	56:89	48:81	8:8	0.325
Height (m)	1.616 ± 0.084	1.615 ± 0.082	1.623 ± 0.101	0.750
Body weight (kg)	59.7 ± 11.7	60.0 ± 11.7	57.4 ± 12.3	0.424
BMI (kg/m^2)	22.7 ± 3.4	22.9 ± 3.2	21.8 ± 4.7	0.394

Demographic data of patients for training/internal validation and external validation

BMI, body mass index.

Height, body weight, and BMI are reported as the mean value and standard deviation.

* Significant.

Table 2
CT acquisition and reconstruction parameters.

Dataset	Scanner(s)	Tube Voltage	Tube Current	Slice Thickness / Increment	Matrix / Field of View (FOV)
Primary (Training & Internal Validation)	- Canon Aquilion 64 (63 patients) - Canon Aquilion ONE (50 patients) - Siemens Sensation 16 (16 patients)	120 kV	Automatic exposure control	0.5 mm / 0.5 mm	512 × 512 / 350 mm
External Validation	Siemens SOMATOM Sensation Cardiac 64	120 kV	180 mAs (fixed) or CareDose4D (modulated)	0.5–1.0 mm / 0.5–0.8 mm	512 × 512 / 350 mm

Development of the automated segmentation model

A Bayesian U-Net was used for automated intercostal muscle segmentation (16). Dropout layers estimated predictive uncertainty by Monte Carlo dropout (19). Following an uncertainty-based active-learning strategy (20), the model was trained with Adam optimization (100,000 iterations; learning rate, 0.0001)(21) and iteratively refined using cases with high structure-wise predictive variance.

Reference labels were generated in six steps. Observer A (18 years of thoracic imaging experience) manually segmented the entire intercostal muscle on 10 non-enhanced CT examinations. We identified 10 cases in order of decreasing PSV (high-uncertainty cases), as measured by predictive structure-wise variance (PSV). Observer A manually corrected the automated segmentations for these 10 high-uncertainty cases, creating a refined dataset. This process was repeated three times. One additional set of 10 high-uncertainty non-enhanced cases was corrected by observer B (20 years of thoracic imaging experience). After training on all non-enhanced label sets, the model was applied to the 19 contrast-enhanced examinations, and 10 high-uncertainty enhanced cases were manually corrected by observer A. In total, 60 manually corrected examinations were used for model development (Figs. 1 and 2).

Internal validation

Internal performance was assessed by fivefold cross-validation using the manually corrected reference set. In each fold, four subsets were used for training and one for testing. Dice similarity coefficient (DSC) and average symmetric surface distance (ASSD)(22, 23) were calculated.

Because many reference labels were created through model-assisted correction during active learning, internal validation was interpreted as development-set performance rather than as a fully independent test.

External validation

External validation used an independent dataset in another institution with manually-generated reference labels. Because to generate an annotation of the whole intercostal muscles is extremely labor-intensive, a reproducible representative 3D slab was selected. For each of the cases, 10 consecutive mid-coronal reconstructed images with 1-mm slices through the lateral arches of both first ribs were extracted. In these coronal planes, all intercostal muscles, including the transverse arch of the first rib, could be consistently visualized. Two radiologists independently segmented the right intercostal muscles with 10-mm thickness to create reference label 1 (RL1) and reference label 2 (RL2). In addition, bilateral sixth intercostal muscles were manually segmented in their entirety based on detailed observation in coronal, sagittal, and trans-axial planes and used as additional reference labels (RL3). Model predictions were compared with each of the reference labels (RL1, RL2, and RL3) using DSC, ASSD, 95-percentile Housdorff distance (HD95)(24), and Bland-Altman analysis.

Visual assessment

Two radiologists independently reviewed all segmentation results in consensus. Segmentation accuracy was evaluated across six anatomical regions of the intercostal muscles: right 1st–3rd, right 4th–6th, right 7th–11th, left 1st–3rd, left 4th–6th, and left 7th–11th intercostal muscles. Inter-regional differences were analyzed using ANOVA.

Segmentation errors were categorized as excessive errors (inclusion of non-muscle structures such as adjacent muscles, pleura, or lesions) and insufficient errors (under-segmentation of true intercostal muscle areas). Each error type was scored on a 4-point scale (0–3). A visual accuracy score was calculated as the sum of both error scores, representing the overall clinical acceptability of the segmentation. For example, a score of 1 for excessive inclusion of two non-IM structures and a score of 1 for missing three IM regions resulted in a total score of 2.

Exploratory analysis of intercostal muscle metrics and pulmonary function

To explore clinical plausibility, we analyzed 95 examinations from the primary dataset in which PFT had been performed within 1 month of CT. For examinations obtained within 1 month after surgery, the preoperative PFT result was used; for examinations obtained later, the postoperative result was used.

Two CT attenuation-based intercostal muscle volume metrics were calculated: total intercostal muscle volume (-190 to 150 HU), including muscle and intramuscular fat, and lean intercostal muscle volume (-30 to 150 HU), preferentially reflecting lean muscle tissue. Whole-volume measurements are reported in cm3. Height-normalized volume, body mass index (BMI)-normalized volume, and mean CT density were also calculated. These variables were correlated with height, weight, BMI, forced vital capacity (FVC), forced expiratory volume in 1 second (FEV1), FEV1/FVC, and percent vital capacity (%VC).

Statistical analysis

Continuous variables are presented as mean ± standard deviation. Pearson correlation coefficients were used for exploratory association analyses, with $p < 0.05$ considered statistically significant. Additional exploratory multivariable linear regression models for FVC and FEV1 included total or lean intercostal muscle volume separately, with height, BMI, age, and sex as covariates. No formal sample size calculation was performed because this was an exploratory technical validation study. p values for multivariable analyses are presented descriptively. Statistical analyses were performed using SPSS software (IBM Corp., Armonk, NY, USA; version to be specified).

Use of generative artificial intelligence

During manuscript preparation, the authors used ChatGPT-5 and Gemini 2.5 Flash to assist with grammar, clarity, and structural refinement. The authors reviewed and edited all AI-assisted output and take full responsibility for the content of the manuscript.

Results

Segmentation performance

The automated segmentation process required approximately 5 minutes per examination.

In development-set internal validation, the Bayesian U-Net achieved a DSC of 0.84 ± 0.08 and an ASSD of 0.66 ± 0.47 mm. Visual review showed that excessive segmentation errors were uncommon and were mainly limited to structures close to the chest wall, such as the quadratus lumborum muscle, apical pleural or inflammatory lesions, and intercostal vessels (Tables 3 and 4; Fig. 3). Major thoracic abnormalities within the lung fields, including pneumonia, pleural effusion, and pulmonary masses, were not mislabeled as intercostal muscle. Although portions of the intercostal vessels were occasionally included as segmentation errors, vessels that were visually separable from the intercostal muscles were adequately distinguished in the other anatomical sections (supplemental file 1,2). Insufficient segmentation errors were more frequent in lower intercostal spaces, particularly in regions with marked fatty change, but regional differences were not statistically significant.

Table 3
Visual evaluation of segmentation accuracy for internal validation

Section of intercostal muscles	Excessive errors (0–3)	Insufficient errors (0–3)	Visual segmentation accuracy (0–6)
Right 1st to 3rd	0.14 ± 0.354	0.18 ± 0.441	0.33 ± 0.591
Right 4th to 6th	0.16 ± 0.373	0.08 ± 0.449	0.25 ± 0.560
Right 7th to 11th	0.16 ± 0.373	0.27 ± 0.569	0.43 ± 0.645
Left 1st to 3rd	0.18 ± 0.391	0.15 ± 0.500	0.33 ± 0.625
Left 4th to 6th	0.18 ± 0.391	0.14 ± 0.500	0.33 ± 0.591
Left 7th to 11th	0.18 ± 0.391	0.29 ± 0.540	0.47 ± 0.649
p-value	0.976	0.234	0.372

Excessive error, insufficient error, and visual segmentation accuracy in each of the 6 sections are expressed as the mean value and standard deviation. For individual section of intercostal muscles, both excessive and insufficient errors were scored in a 4-grade scale as follows; 0 point/0 error, 1 point/1–3 errors, 2 points/4–6 errors, 3 points/7 or more errors.

Table 4
Visual assessment of segmentation errors.

Regions with excessive errors	Number
Quadratus lumborum muscles	5
Pulmonary apical lesions	5
Pleura	3
Diaphragm	1
Intercostal arteries and veins	4
Rectus abdominis muscles	2
Shoulder skin	1
The numbers of regions where excessive errors were identified are indicated.	

In independent external validation, interobserver agreement between RL1 and RL 2 yielded a DSC of 0.775 ± 0.042. DSC between the predictive label and RL1 was 0.774 ± 0.043, and that between the predictive label and RL2 was 0.796 ± 0.045, indicating model performance comparable to human interobserver agreement at the selected anatomical level. Bland-Altman analysis showed mean differences of -1178 mm^3 (limits of agreement, -2614 to 258 mm^3) between the predictive label and RL1 and − 1666 mm^3 (limits of agreement, -3691 to 358 mm^3) between the predictive label and RL2, indicating a mild tendency toward smaller model-derived slab volumes than manual slab volumes. The mean volume of the bilateral 6th intercostal muscle was 37.37cm^3 with manual segmentation and 37.41 cm^3 with the predictive label. For the bilateral 6th intercostal muscles, the model achieved a DSC of 0.738 ± 0.025, an ASSD of 0.921 ± 0.161 mm, and HD95 of 2.673 ± 0.220 mm compared with RL3. Bland–Altman analysis showed a mean difference of − 0.04 cm^3 (limits of agreement, − 1.96 to 11.05 cm^3), providing validation of the model's reproducible performance at this intercostal level (Fig. 4).

Exploratory association with body habitus and pulmonary function

The exploratory PFT analysis included 95 examinations from the primary dataset. Whole intercostal muscle volume was higher in men than in women (378.7 ± 66.9 cm^3 vs 246.7 ± 41.2 cm^3, $p < 0.001$) and correlated positively with height ($r = 0.690$, $p < 0.001$), weight ($r = 0.663$, $p < 0.001$), and BMI ($r = 0.435$, $p < 0.001$) (Table 5, Fig. 5).

Table 5
Correlations with body habitus and pulmonary function.

Comparison between genders					
	IMV	**Lean IMV[1]**	**IMV/height**	**IMV/BMI**	**MCTD (HU)[2]**
Male	378.4 ± 66.4	246.7 ± 59.6	138.4 ± 21.4	16.4 ± 2.8	15.2 ± 5.2
Female	243.4 ± 36.8	169.7 ± 36.1	101.4 ± 15.3	11.5 ± 1.8	14.6 ± 5.2
p-value	< 0.001	< 0.001	< 0.001	< 0.001	0.559
Correlation with pulmonary function tests (*r*-value)					
	IMV	Lean IMV[1]	IMV/height	IMV/BMI	MCTD(HU) [2]
FEV_1	0.429*	0.412*	0.295*	0.231*	0.292*
FVC	0.568*	0.513*	0.431*	0.392*	0.275*
FEV_1/FVC	-0.019	-0.016	-0.074	-0.052	-0.129
%VC	-0.007	0.070	0.009	0.073	0.157
Comparison of intercostal muscles between genders and correlation of intercostal muscles with pulmonary function test parameters in 95 patients who underwent non-enhanced CT					
BMI, body mass index; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; HU, Hounsfield unit; IMV, intercostal muscle volume; MCTD, mean computed tomography density.					
[1]Lean IMV was defined as the total number of pixels with CT density more than − 30 HU multiplied by a single pixel dimension and automatically measured.					
[2]MCTD stands for the mean CT density of IMV.					
*Significant(*P*<0.001).					

Intercostal muscle volume also correlated positively with FVC (r = 0.568, p < 0.001) and FEV1 (r = 0.429, p < 0.001), whereas no significant correlation was found with FEV1/FVC or %VC. Mean CT density and height-normalized intercostal muscle volume showed similar positive relationships with FVC and FEV1. Lean intercostal muscle volume demonstrated stronger correlations with both measures than total intercostal muscle volume. In exploratory multivariable analyses adjusted for height, BMI, age, and sex, lean intercostal muscle volume remained positively associated with FVC (beta = 0.204, p = 0.027), and total intercostal muscle volume showed a similar association (beta = 0.253, p = 0.038). The adjusted association with FEV1 was attenuated (beta = 0.082, p = 0.534).

Discussion

This study shows that automated 3D segmentation of intercostal muscles on thin-slice chest CT is technically feasible using a Bayesian U-Net framework. The model achieved acceptable development-set performance, external agreement comparable to human interobserver agreement at a predefined anatomical level, and CT-derived muscle metrics that showed exploratory positive associations with FVC and FEV1.

Intercostal muscle segmentation is more challenging than segmentation of larger skeletal muscles such as the psoas, gluteal, or thigh muscles (12, 13, 15–18, 20, 23, 25–28). The muscles are thin, curved, and adjacent to fat, pleura, vessels, and other chest-wall structures, which is why prior imaging studies have usually evaluated thickness or cross-sectional area at selected levels (5–11). Choi et al. developed an empirically modified Attention U-Net model for 3D intercostal muscle segmentation but reported significantly smaller muscle volumes compared to our finding (14). This discrepancy likely results from differences in CT slice thickness (3 mm in Choi's study vs. ≤1 mm in ours). Furthermore, our study included patients with pleural abnormalities and pulmonary diseases—conditions explicitly excluded by Choi et al.—demonstrating the clinical robustness of our model in a diverse, real-world dataset.

Building on prior work, our study extends this field by targeting whole-muscle 3D segmentation on routine thin-slice chest CT. We deliberately used a heterogeneous development cohort that included postoperative changes, pleural abnormalities, and pulmonary diseases, because a model that performs only in ideal anatomy would have limited clinical value.

Internal validation should be interpreted cautiously because many reference labels were created within an active-learning workflow and some were derived from model-generated segmentations that were subsequently corrected. We therefore consider the external validation to provide the more important evidence of generalizability. Notably, the model-to-reader DSC values were close to the reader-to-reader DSC, suggesting that, at the tested anatomical level, algorithm performance was within the range of human variability.

DSC is known to be considerably affected by the area/volume of the measurement, and smaller anatomical structures, with lesser pixel/voxel number, generally yield lower DSC values (29–32). This underscores the importance of boundary-based metrics such as HD95 and ASSD when assessing geometric accuracy. In the head and neck region, the DSC for small and complex organs or structures typically ranges from 0.70 to 0.80, which could be considered acceptable (24, 30, 33). The DSC obtained in this study (0.738 ± 0.025) aligns with the results of the previous studies and indicates appropriate segmentation accuracy for small muscles. Moreover, the ASSD (< 1 mm) and HD95 (2.673 ± 0.220 mm) could reflect sufficient boundary precision. A previous study has also shown that, for small structures, feasible segmentation performance could be achieved even when Dice scores are modest, provided that boundary-based metrics are favorable (34). Thus, we believe that the segmentation performance of the model developed in this study can be feasible based on the quantitative measurements. Nonetheless, underestimation tendency of the ICM segmentation in limited regions including the mid-coronal plane was shown by the model. On the other hand, considering that the estimated volume of the single ICM by

the model was almost comparable to that by the human observer, the model might recognize structures or regions passing diagonally more appropriately compared with the human observer.

Visual assessment revealed minimal gross errors. Our model minimized over-segmentation by effectively excluding adjacent structures such as pleura, blood vessels, serratus anterior, and vastus lateralis muscles. Twenty-one excessive errors were noted among 49 cases, primarily involving non-muscle structures such as quadratus lumborum and apical lung lesions, but no thoracic diseases (e.g., pleural effusion, pneumonia, or lung cancer) were misclassified as IMs, and such errors were easily rectified during visual review. Insufficient errors, primarily observed in the lower intercostal regions, may result from fat tissue being misclassified as extramuscular structures. These errors, however, can often be visually corrected.

The external validation also has important limitations. It was performed in a relatively small dataset from a second institution and evaluated only partial intercostal muscle regions, namely a representative coronal slab and the full volume of the sixth intercostal muscles, rather than the entire thorax. This design was chosen because independent whole-thorax manual labeling of all intercostal muscles is extremely time-consuming. Accordingly, the external validation should be viewed as a rigorous site-specific benchmark rather than definitive proof of full-volume performance across all disease states. Larger multi-institutional studies, ideally including COPD, emphysema, interstitial lung disease, neuromuscular disease, and intensive care populations, are needed.

We showed that automatic segmentation based on the thickness of the coronal cross-section tends to yield slightly smaller muscle volumes compared to manual contouring. A similar trend has been reported in the segmentation of muscle and adipose tissue in cross-sections of the trunk(35, 36). This likely reflects a conservative boundary definition by the model: manual tracing may include a small amount of boundary-adjacent tissue to avoid missing thin muscle fibers, whereas the model may more strictly follow the visible muscular margin. Importantly, this bias was modest and occurred in the setting of model-reader agreement similar to interobserver agreement. In contrast, for the full volume of the 6th intercostal muscle, the mean difference was nearly eliminated (− 0.04 cm^3). This is likely because manual full-muscle tracing allows observers to inspect the muscle from oblique views and more broadly assess its boundaries, making it easier to exclude adjacent tissues.

The observed correlations between intercostal muscle metrics and pulmonary function support the biological plausibility of the segmentation outputs. Respiratory muscle mass and quality are expected to relate to ventilatory performance, and prior ultrasound- and CT-based studies have reported associations between intercostal muscle measurements and COPD severity or respiratory status (5–8). In our study, total volume, lean volume, and density-based metrics were positively related to FVC and FEV1, whereas FEV1/FVC and %VC were not. This pattern suggests that segmented intercostal muscle metrics may reflect body-size-related ventilatory capacity more strongly than airflow obstruction. In exploratory multivariable analyses adjusted for age, sex, height, and BMI, positive associations with FVC persisted for both total and lean intercostal muscle volume, whereas the adjusted association with FEV1 did not.

These analyses do not establish an independent biomarker, but they suggest that the segmentation output is physiologically meaningful enough to justify future disease-focused studies.

This work should be viewed as an initial technical validation of deep-learning-based intercostal muscle quantification across diverse thoracic CT examinations. Future studies should examine clinical applications in chronic obstructive pulmonary disease, interstitial lung disease, neuromuscular disorders, postoperative respiratory impairment, and respiratory failure. Longitudinal evaluation of intercostal muscle volume and density may also clarify relationships with functional decline, exacerbation risk, postoperative recovery, ventilator dependence, and response to rehabilitation.

Our study has several limitations. First, the development cohort consisted of patients undergoing surgery for pulmonary tumors rather than a disease-specific respiratory cohort. Second, internal validation was not fully independent because of the active-learning process. Third, the external validation dataset was small and anatomically standardized. Fourth, the pulmonary function analysis was exploratory and performed in a limited subset. Fifth, we did not compare the Bayesian U-Net directly with other architectures or formally test 2D versus 3D endpoints. Finally, internal and external intercostal muscle layers were not segmented separately.

Conclusions

Bayesian U-Net-based automated segmentation enables quantitative 3D assessment of intercostal muscles on thin-slice chest CT, including technically challenging examinations. External validation showed agreement comparable to human interobserver agreement at a predefined anatomical level, and an additional full-volume assessment of the bilateral sixth intercostal muscles supported performance at a reproducible intercostal level. The resulting metrics showed exploratory associations with pulmonary function, particularly FVC. This method may provide a practical platform for future respiratory muscle imaging studies.

Abbreviations

AI
artificial intelligence
ASSD
average symmetric surface distance
BMI
body mass index
COPD
chronic obstructive pulmonary disease
CT
computed tomography
DICOM

Digital Imaging and Communications in Medicine
DSC
Dice similarity coefficient
FEV1
forced expiratory volume in 1 second
FVC
forced vital capacity
GT
ground truth
HD95
95-percentile Housdorff distance
HU
Hounsfield unit
IM
intercostal muscle
IMV
intercostal muscle volume
PFT
pulmonary function test
RL
reference label

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the principles of the Helsinki Declaration. This retrospective study involving human imaging data was approved by the institutional review boards of Shiga University of medical science (approval number R2016-209). The requirement for informed consent was waived because of the retrospective design with opt-out disclosure.

Consent for publication

Not applicable. This manuscript does not contain identifying individual participant information.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Author Contribution

YS conducted the investigation and contributed to the methodology. YM administered the project and drafted the manuscript. YM, YMk, and MW provided resources. YMk developed the software and performed validation with YN. YMk and YM prepared the visualizations. SI and MW collected and managed the external validation data. YW supervised the study and, together with YN, revised the manuscript. All authors approved the final version.

Acknowledgements

Not applicable.

Data Availability

The datasets used and/or analyzed during the current study are not publicly available because they contain institutional clinical imaging data and patient information covered by privacy and ethics restrictions but may be available from the corresponding author on reasonable request and with permission from the relevant participating institutions and ethics committees.

References

1. Okazaki T, Suzukamo Y, Miyatake M, Komatsu R, Yaekashiwa M, Nihei M, et al. Respiratory Muscle Weakness as a Risk Factor for Pneumonia in Older People. Gerontology. 2021;67(5):581–90.
2. Okazaki T, Ebihara S, Mori T, Izumi S, Ebihara T. Association between sarcopenia and pneumonia in older people. Geriatr Gerontol Int. 2020;20(1):7–13.
3. Duiverman ML, de Boer EW, van Eykern LA, de Greef MH, Jansen DF, Wempe JB, et al. Respiratory muscle activity and dyspnea during exercise in chronic obstructive pulmonary disease. Respir Physiol Neurobiol. 2009;167(2):195–200.
4. Wang Y, Shao F, Zhang L, Luo X, Chen Y. Increased 18F-FDG Uptake in Multiple Muscles in a Patient With Violent Cough. Clin Nucl Med. 2017;42(6):451–3.
5. Wallbridge P, Parry SM, Das S, Law C, Hammerschlag G, Irving L, et al. Parasternal intercostal muscle ultrasound in chronic obstructive pulmonary disease correlates with spirometric severity. Sci Rep. 2018;8(1):15274.
6. Park MJ, Cho JM, Jeon KN, Bae KS, Kim HC, Choi DS, et al. Mass and fat infiltration of intercostal muscles measured by CT histogram analysis and their correlations with COPD severity. Acad Radiol. 2014;21(6):711–7.

7. Ju S, Lee SJ, Park MJ, Cho YJ, Jeong YY, Jeon KN, et al. Clinical importance of cross-sectional area of intercostal muscles in patients with chronic obstructive pulmonary disease. Clin Respir J. 2018;12(3):939–47.
8. Güerri R, Gayete A, Balcells E, Ramirez-Sarmiento A, Vollmer I, Garcia-Aymerich J, et al. Mass of intercostal muscles associates with risk of multiple exacerbations in COPD. Respir Med. 2010;104(3):378–88.
9. Nakanishi N, Oto J, Ueno Y, Nakataki E, Itagaki T, Nishimura M. Change in diaphragm and intercostal muscle thickness in mechanically ventilated patients: a prospective observational ultrasonography study. J Intensive Care. 2019;7:56.
10. Mourtzakis M, Prado CM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Appl Physiol Nutr Metab. 2008;33(5):997–1006.
11. Wang K, Wu F, He H, Hu C, Chen X, Chen J, et al. Association between computed tomography-quantified respiratory muscles and chronic obstructive pulmonary disease: a retrospective study. BMC Pulm Med. 2024;24(1):150.
12. Kamiya N, Li J, Kume M, Fujita H, Shen D, Zheng G. Fully automatic segmentation of paraspinal muscles from 3D torso CT images via multi-scale iterative random forest classifications. Int J Comput Assist Radiol Surg. 2018;13(11):1697–706.
13. Ibad HA, Hathaway QA, Bluemke DA, Kasaeian A, Klein JG, Budoff MJ, et al. CT-derived pectoralis composition and incident pneumonia hospitalization using fully automated deep-learning algorithm: multi-ethnic study of atherosclerosis. Eur Radiol. 2024;34(6):4163–75.
14. Choi I, Choi J, Yong HS, Yang Z. Deep learning-based respiratory muscle segmentation as a potential imaging biomarker for respiratory function assessment. PLoS ONE. 2024;19(7):e0306789.
15. Yang Z, Choi I, Choi J, Jung J, Ryu M, Yong HS. Deep learning-based pectoralis muscle volume segmentation method from chest computed tomography image using sagittal range detection and axial slice-based segmentation. PLoS ONE. 2023;18(9):e0290950.
16. Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N, Sato Y. Automated Muscle Segmentation from Clinical CT Using Bayesian U-Net for Personalized Musculoskeletal Modeling. IEEE Trans Med Imaging. 2020;39(4):1030–40.
17. Kim HS, Kim H, Kim S, Cha Y, Kim JT, Kim JW, et al. Precise individual muscle segmentation in whole thigh CT scans for sarcopenia assessment using U-net transformer. Sci Rep. 2024;14(1):3301.
18. Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, et al. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiol Artif Intell. 2023;5(5):e230024.
19. Steinbrener J, Posch K, Pilz J. Measuring the Uncertainty of Predictions in Deep Neural Networks with Variational Inference. Sens (Basel). 2020;20(21).
20. Li G, Otake Y, Soufi M, Taniguchi M, Yagi M, Ichihashi N, et al. Hybrid representation-enhanced sampling for Bayesian active learning in musculoskeletal segmentation of lower extremities. Int J Comput Assist Radiol Surg. 2024;19(11):2177–86.

21. Diederik P, Kingma. Jimmy Ba. Adam: A Method for Stochastic Optimization. In: Ba J, editor. ArXiv:1412.69802014.
22. Jafari-Khouzani K, Elisevich KV, Patel S, Soltanian-Zadeh H. Dataset of magnetic resonance images of nonepileptic subjects and temporal lobe epilepsy patients for validation of hippocampal segmentation techniques. Neuroinformatics. 2011;9(4):335–46.
23. Gaj S, Eck BL, Xie D, Lartey R, Lo C, Zaylor W et al. Deep learning-based automatic pipeline for quantitative assessment of thigh muscle morphology and fatty infiltration. Magn Reson Med. 2023.
24. Choi B, Beltran CJ, Yoo SK, Kwon NH, Kim JS, Park JC. The InterVision Framework: An Enhanced Fine-Tuning Deep Learning Strategy for Auto-Segmentation in Head and Neck. J Pers Med. 2024;14(9).
25. Ding J, Cao P, Chang HC, Gao Y, Chan SHS, Vardhanabhuti V. Deep learning-based thigh muscle segmentation for reproducible fat fraction quantification using fat-water decomposition MRI. Insights Imaging. 2020;11(1):128.
26. Amarasinghe KC, Lopes J, Beraldo J, Kiss N, Bucknell N, Everitt S, et al. A Deep Learning Model to Automate Skeletal Muscle Area Measurement on Computed Tomography Images. Front Oncol. 2021;11:580806.
27. Ha J, Park T, Kim HK, Shin Y, Ko Y, Kim DW, et al. Development of a fully automatic deep learning system for L3 selection and body composition assessment on computed tomography. Sci Rep. 2021;11(1):21656.
28. Duong F, Gadermayr M, Merhof D, Kuhl C, Bruners P, Loosen SH, et al. Automated major psoas muscle volumetry in computed tomography using machine learning algorithms. Int J Comput Assist Radiol Surg. 2022;17(2):355–61.
29. Doolan PJ, Charalambous S, Roussakis Y, Leczynski A, Peratikou M, Benjamin M, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. Front Oncol. 2023;13:1213068.
30. Podobnik G, Ibragimov B, Tappeiner E, Lee C, Kim JS, Mesbah Z, et al. HaN-Seg: The head and neck organ-at-risk CT and MR segmentation challenge. Radiother Oncol. 2024;198:110410.
31. Maier-Hein L, Reinke A, Godau P, Tizabi MD, Buettner F, Christodoulou E, et al. Metrics reloaded: recommendations for image analysis validation. Nat Methods. 2024;21(2):195–212.
32. Chen X, Sun S, Bai N, Han K, Liu Q, Yao S, et al. A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. Radiother Oncol. 2021;160:175–84.
33. Heilemann G, Buschmann M, Lechner W, Dick V, Eckert F, Heilmann M, et al. Clinical Implementation and Evaluation of Auto-Segmentation Tools for Multi-Site Contouring in Radiotherapy. Phys Imaging Radiat Oncol. 2023;28:100515.
34. Mackay K, Bernstein D, Glocker B, Kamnitsas K, Taylor A. A Review of the Metrics Used to Assess Auto-Contouring Systems in Radiotherapy. Clin Oncol (R Coll Radiol). 2023;35(6):354–69.

35. Esparza JP, Ahmed MM, Shrestha U, Dayal B, Makhani S, Nanda D et al. A dual-contrast deep learning model for automated assessment of sarcopenia using MRI. Abdom Radiol (NY). 2026.
36. Kuss CP, Ulbricht L, Schumacher K, Ripka WL. Comparison between automated and manual segmentation in computed tomography for body composition analysis. Biomed Eng Online. 2026;25(1).

Figures

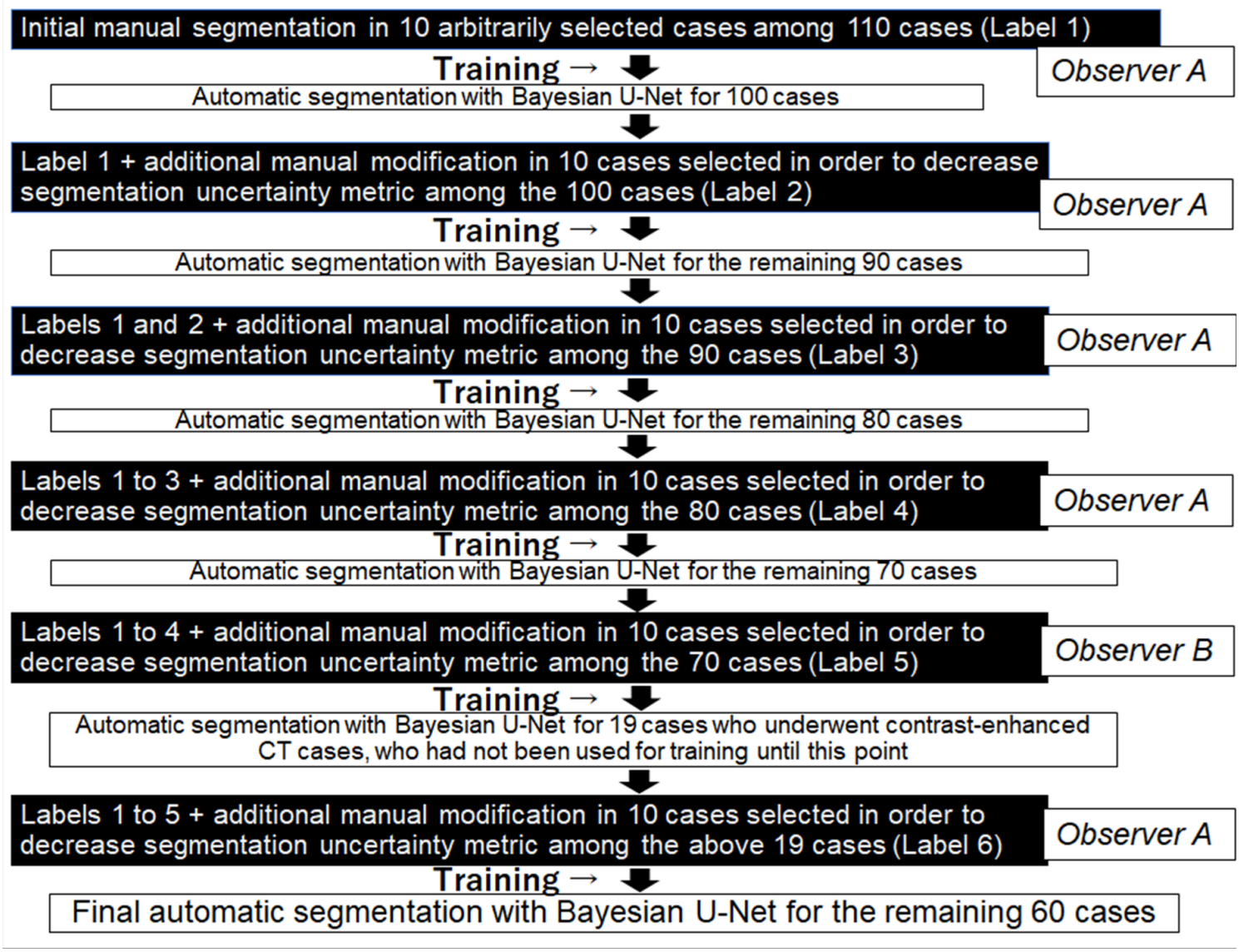


Figure 1

Active-learning framework for intercostal muscle segmentation. Overview of the uncertainty-based active-learning workflow used to select high-uncertainty cases and generate manually corrected reference labels.

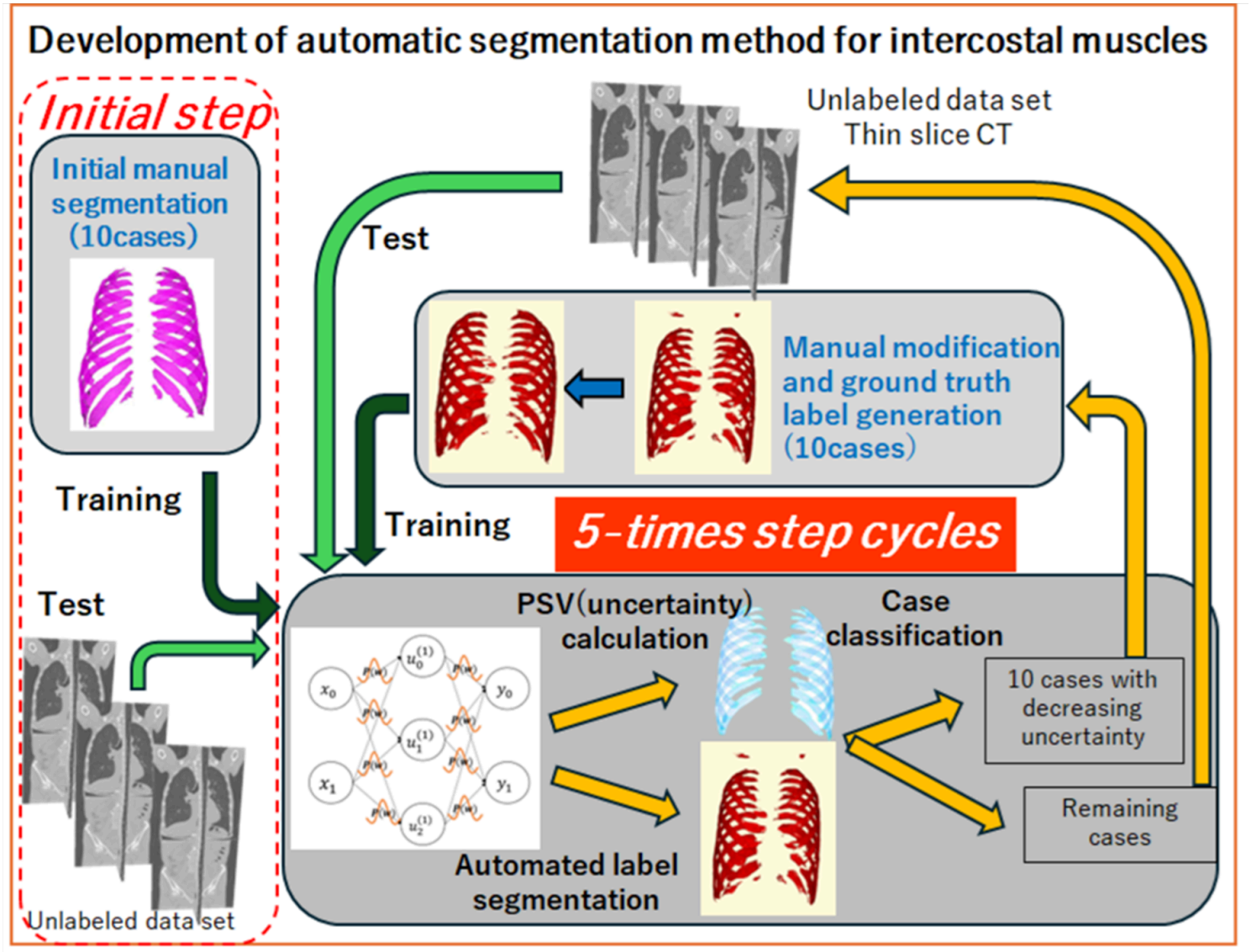


Figure 2

Overview of automated intercostal muscle segmentation methods

Bayesian U-Net development and manual correction process. Stepwise generation of non-enhanced and contrast-enhanced reference labels for model training and validation.

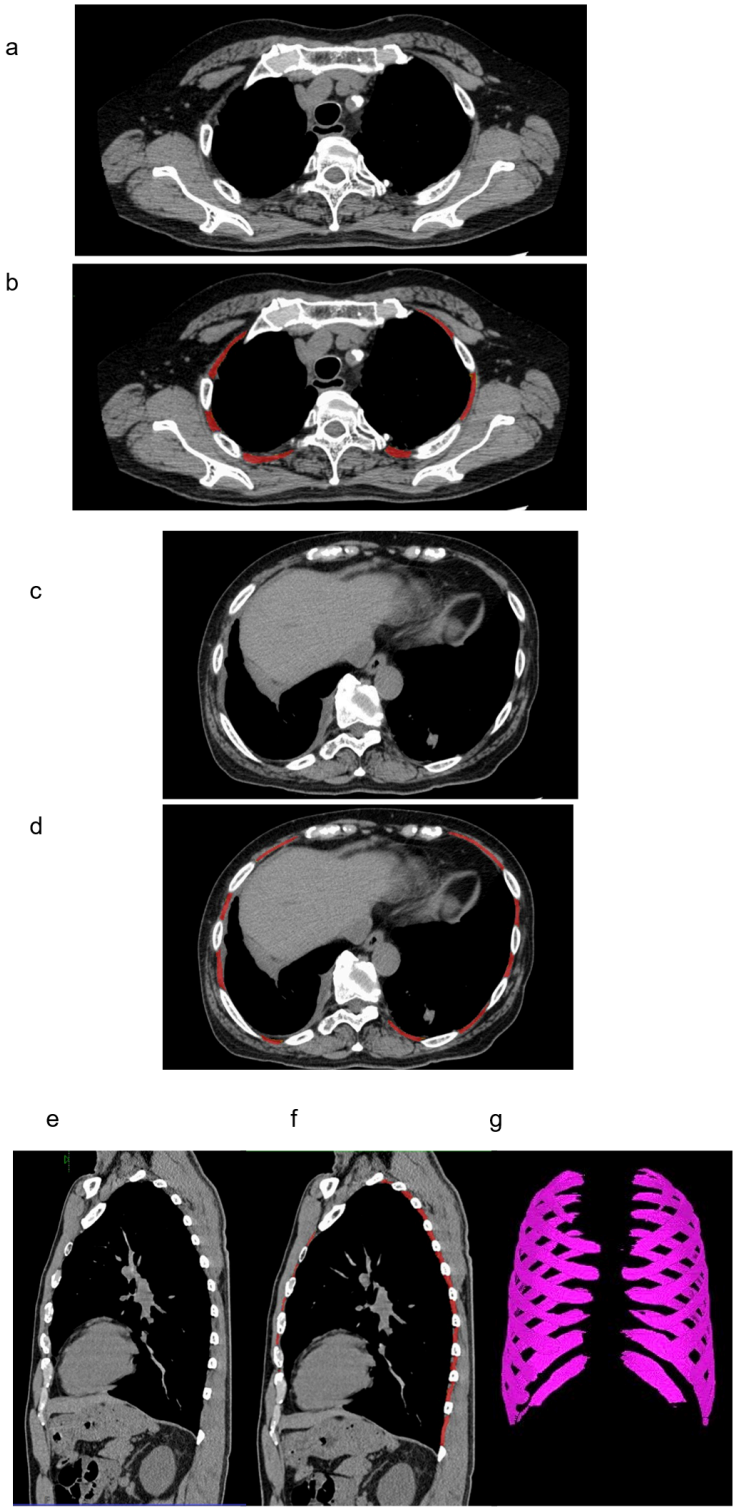


Figure 3

A case with a Dice similarity coefficient of 0.917 and a normal BMI of 18.6. Trans-axial images in a level just above the aortic arch (a, b) and in another level at the diaphragmatic crus (c, d), a multiplanar reconstructed image in the mid-sagittal plane of the left lung (e, f) are demonstrated. Highlighted areas in red stand for labelled ones by the Bayesian U-Net-based automated segmentation algorithm (b, d, and f). Compared with original images (a, c, and e), it is obvious that most intercostal muscles are correctly

segmented, and neither pleural thickening/effusion, nor the diaphragm adjacent to the intercostal muscles are included in the labelled areas. A volume rendered image colored in a pink (g) corresponding to the highlighted areas in red (f) shows three-dimensional appearance of the segmented intercostal muscles.

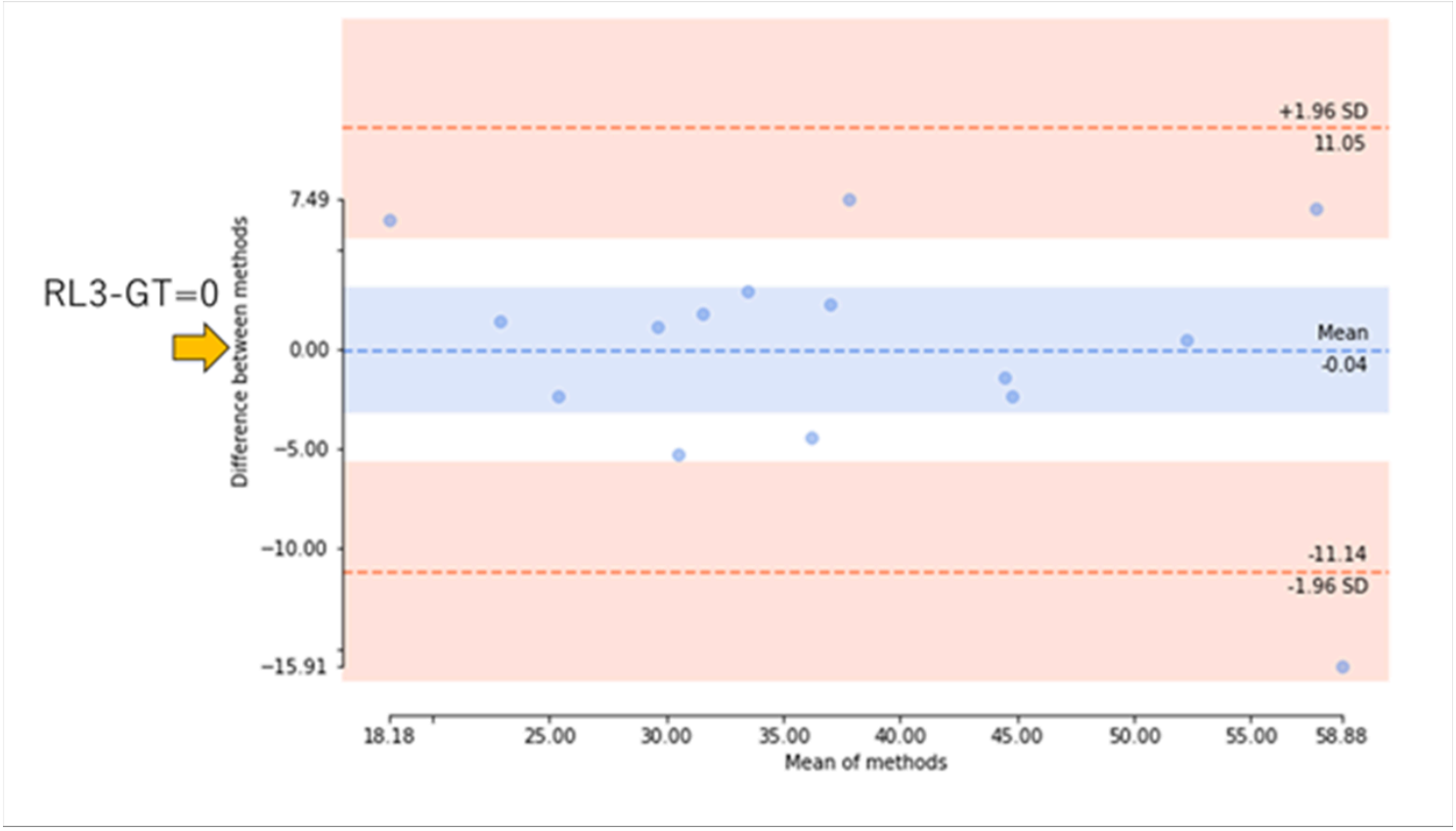


Figure 4

Bland–Altman analysis comparing 3D segmentation of the bilateral sixth intercostal muscles between the reference model 3 (RL3) and manual ground-truth (GT) annotations. The Bland–Altman plot shows a mean difference of −0.04 cm³, with limits of agreement ranging from −1.96 to 11.05 cm³, demonstrating reproducible agreement between RL3 and manual segmentation at this intercostal level.

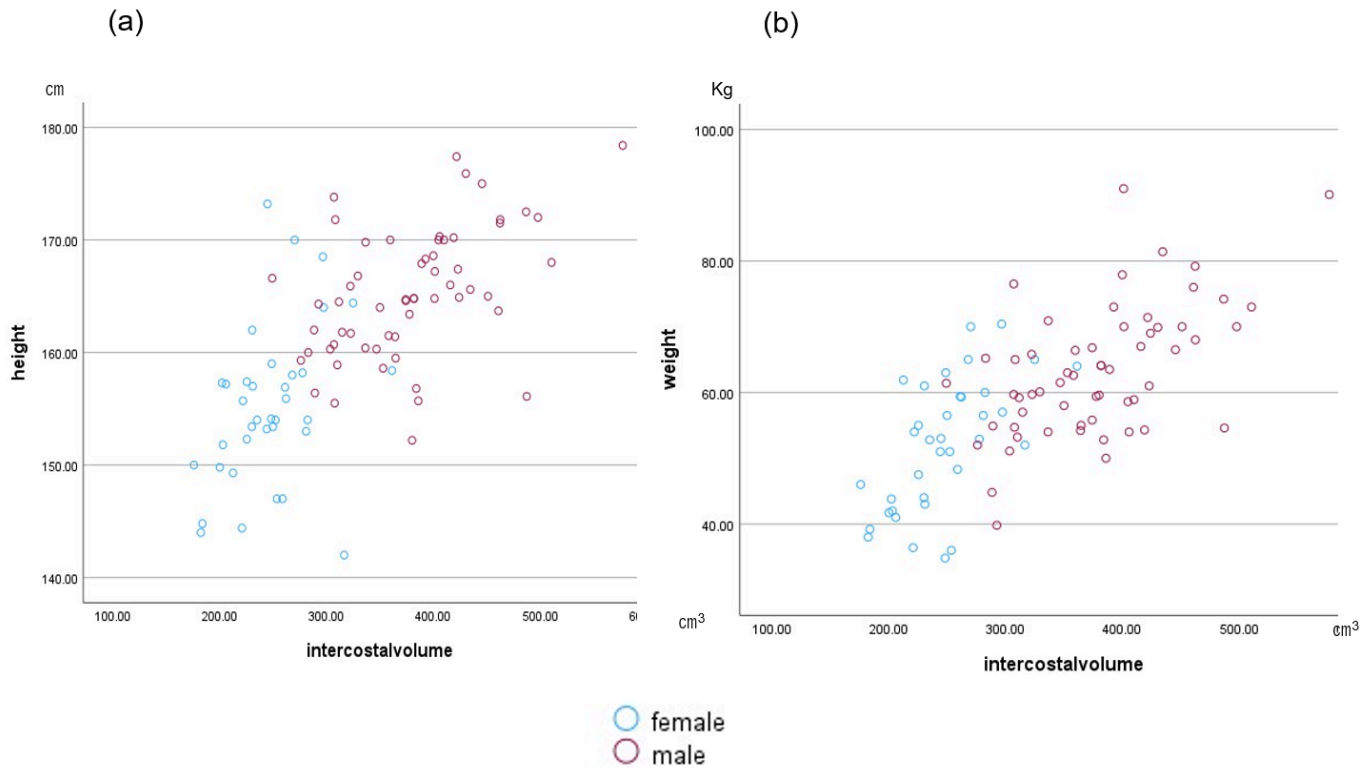


Figure 5

Scatter plots demonstrate the relationship between intercostal muscle volume and height (a) and between intercostal muscle volume and weight (b). Blue and red circle plots indicate for male and female patients, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalfile1.pptx
- sipplementalfile2.pptx