

Deep neural network for automatic volumetric segmentation of whole-body CT images for body composition assessment

Yoon Seong Lee

Seoul National University Hospital

Namki Hong

Yonsei University College of Medicine <https://orcid.org/0000-0002-8246-1956>

Joseph Nathanael Witanto

MEDICALIP Co. Ltd.

Ye Ra Choi

Seoul Metropolitan Government-Seoul National University Boramae Medical Center

<https://orcid.org/0000-0002-2455-1718>

Junghoan Park

Seoul National University Hospital

Pierre Decazes

Henri Becquerel Cancer Center <https://orcid.org/0000-0001-5323-9910>

Florian Eude

Henri Becquerel Cancer Center

Chang Oh Kim

Yonsei University College of Medicine <https://orcid.org/0000-0002-0773-5443>

Hyeon Chang Kim

Yonsei University College of Medicine <https://orcid.org/0000-0001-7867-1240>

Jin Mo Goo

Seoul National University Hospital <https://orcid.org/0000-0003-1791-7942>

Yumie Rhee

Yonsei University College of Medicine <https://orcid.org/0000-0003-4227-5638>

Soon Ho Yoon (✉ yshoka@gmail.com)

Seoul National University Hospital <https://orcid.org/0000-0002-3700-0165>

Research Article

Keywords: sarcopenia, computed tomography, deep learning, segmentation, body composition

Posted Date: December 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-115444/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Clinical Nutrition on December 16th, 2020.

See the published version at <https://doi.org/10.1016/j.clnu.2021.06.025>.

Deep neural network for automatic volumetric segmentation of whole-body CT images for body composition assessment

Yoon Seong Lee, MD^{1*}, Namki Hong, MD, MPH^{2*}, Joseph Nathanael Witanto³, Ye Ra Choi, MD⁴, Junghoan Park, MD¹, Pierre Decazes, MD⁵, Florian Eude, MD⁵, Chang Oh Kim, MD, PhD⁶, Hyeon Chang Kim, MD, PhD⁷, Jin Mo Goo, MD, PhD^{1,8}, Yumie Rhee, MD, PhD²,
Soon Ho Yoon, MD, PhD^{1,3}

*Y.S.L. and N.H. contributed equally to this work.

¹Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Korea; ²Division of Endocrinology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea; ³MEDICALIP Co. Ltd., Seoul, Korea; ⁴Department of Radiology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Korea; ⁵Department of Nuclear Medicine, Henri Becquerel Cancer Center, Rouen, France; ⁶Division of Geriatrics, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea; ⁷Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Korea; ⁸Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea.

Running title: Automatic segmentation of body CT images

Correspondence to Soon Ho Yoon

Clinical associate professor, Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital. 101 Daehak-ro, Chongno-gu, Seoul 03080, Korea; Tel: 82-2-2072-2584; Fax: 82-2-743-7418; E-mail: yshoka@gmail.com

This study did not have any funding source.

Soon Ho Yoon and Joseph Nathanael Witanto are the chief medical officer and machine learning researcher of MEDICAL IP, Co. Ltd, respectively. Other authors have no conflicts of interest to declare.

First author: Yoon Seong Lee

Resident, Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital. 101 Daehak-ro, Chongno-gu, Seoul 03080, Korea; Tel: 82-2-2072-2518; Fax: 82-2-743-7418; E-mail: yoonseonglee31@gmail.com

Type of Article: Original research

Word count: 3583

Deep neural network for automatic volumetric segmentation of whole-body CT images for body composition assessment

Type of manuscript: Original research

Abstract

Body composition analysis on CT images is a valuable tool for sarcopenia assessment. We aimed to develop and validate a deep neural network applicable to whole-body CT images of PET-CT scan for the automatic volumetric segmentation of body composition.

Methods

For model development, CT images of 100 patients who underwent a whole-body or torso ^{18}F -fluorodeoxyglucose PET-CT scan were retrospectively included. Two radiologists semi-automatically labeled the following seven body components in every CT image slice, providing 39,268 images for training the 3D U-Net: skin, bone, muscle, abdominal visceral fat, subcutaneous fat, internal organs with vessels, and central nervous system. The segmentation accuracy was assessed using reference masks from one internal and three external datasets: two domestic centers (n=20, each) and a French public dataset (n=24). The 3D U-Net-driven values were clinically validated using bioelectrical impedance analysis (BIA) and by assessing the model's diagnostic performance for sarcopenia in a community-based elderly cohort (n=522).

Results

The 3D U-Net achieved accurate body composition segmentation with an average dice similarity coefficient of 96.5% to 98.9% for all masks and 92.3% to 99.3% for muscle, abdominal visceral fat, and subcutaneous fat in the validation datasets. The 3D U-Net-derived torso volume of skeletal muscle and fat tissue and the average area of those tissues in the waist were correlated with BIA-derived appendicular lean mass (correlation coefficients: 0.71

and 0.72, each) and fat mass (correlation coefficients: 0.95 and 0.93, each). The 3D U-Net-derived average areas of skeletal muscle and fat tissue in the waist were independently associated with sarcopenia ($P < .001$, each) with adjustment for age and sex, providing an area under the curve of 0.858 (95% CI, 0.815 to 0.901).

Conclusions

This deep neural network model enabled the automatic volumetric segmentation of body composition on whole-body CT images, potentially expanding adjunctive sarcopenia assessment on PET-CT scan and volumetric assessment of metabolism in whole-body muscle and fat tissues.

Keywords: sarcopenia, computed tomography, deep learning, segmentation, body composition

Abbreviations: BIA, bioelectrical impedance analysis; KURE, Korean Urban Rural Elderly; AVF, abdominal visceral fat; SF, subcutaneous fat; IO, internal organs and vessels; CNS, central nervous system; DSC, dice similarity coefficient; PPV, positive predictive value; BMI, body mass index; SMA, skeletal muscle area; BFA, body fat area; SMI, skeletal muscle index; BFI, body fat index; AUROC, area under the receiver operating characteristic curve

Introduction

Sarcopenia is a progressive and generalized skeletal muscle disorder involving the loss of muscle volume and function (1). Sarcopenia is present in 12.9-40.4% of elderly individuals, and is associated with a variety of adverse health outcomes, including falls, functional decline, impaired quality of life, and mortality, especially in combination with obesity (1). The European Working Group on Sarcopenia in Older People recently updated a practical definition and diagnostic criteria of sarcopenia by including measurements of a combination of muscle mass, strength, and physical performance (2). Dual-energy X-ray absorptiometry and bioelectrical impedance analysis (BIA) are typically used to assess the mass of appendicular skeletal muscle in the general population (3). However, the X-ray absorptiometry cannot provide data on skeletal muscle mass in the trunk or muscle adiposity, while BIA cannot measure visceral fat obesity and is prone to various potential sources of error (4).

CT and MRI are widely used cross-sectional imaging modalities in modern medicine and are considered the gold standard for quantifying the amount of muscle and fat in the body (5). A body composition analysis of a single CT image at the L3 vertebra level can estimate the amount of whole-body muscle and fat (6-8), and this method is widely used to evaluate sarcopenia and visceral obesity (9,10). Furthermore, a deep-learning algorithm enabled the automatic quantification of muscle and fat amount at the L3 vertebra level, which traditionally required considerable time and resources of experienced readers when segmented manually (11,12).

Nevertheless, the distribution of abdominal muscle and fat differs across levels of the lumbar vertebrae in the waist, implying that the distribution could be fully captured by the

volumetric analysis of multiple CT images, rather than a single CT image (11). If volumetric CT analysis of body composition could be applied to whole-body CT images, it would be possible to assess sarcopenia and visceral obesity on CT images beyond the L3 level (i.e., chest CT) without additional radiation exposure. Particularly, as ¹⁸F-fluorodeoxyglucose PET/CT scan is a vital examination for the surveillance of systematic metastasis in cancer patients and covers the whole body (13,14), the volumetric body composition analysis could enable an assessment of whole-body composition at baseline and during follow-up in cancer patients (15,16). In this study, we aimed to develop and validate a deep neural network for the volumetric segmentation of body composition applicable to whole-body CT images of PET-CT scan.

Materials and Methods

The institutional review board of the participating institutions approved this retrospective study for model development, and the requirement for informed patient consent was waived. Analysis of the Korean Urban Rural Elderly (KURE) cohort was approved by the institutional review board after obtaining written consent from all participants prior to the study.

Study population

For the training dataset, we retrospectively collected 100 unenhanced whole-body CT scans of PET/CT examination from 100 patients, preferentially for the purpose of an initial staging workup of early lung cancer or systematic screening at a single tertiary center (Figure 1). We used three datasets from other institutions for external validation to evaluate segmentation performance between the reference masks and the network-derived masks (Table 1): the first dataset contained 20 patients with stage I lung cancer, the second dataset contained 20 participants in a community-based older adults cohort with 20 CT scans, and the third dataset contained 24 CT scans of 18 patients with head and neck cancer. For clinical validation, we used the KURE cohort as a fourth external dataset (17). Details about the validation datasets are summarized in Table 1.

Preparation of training CT data

The whole-body CT images were processed in a commercially available software program (MEDIP PRO v2.0.0.0, MEDICALIP Co. Ltd., Seoul, Korea) for semi-automatic

segmentation of the following seven components of body composition: skin, bone, muscle, abdominal visceral fat (AVF), subcutaneous fat (SF), internal organs and vessels (IO), and central nervous system (CNS). The bone mask extended from the skull to the phalanges and contained the hyoid bone, cricoid, arytenoid, thyroid cartilages, and rib cartilages. The muscle mask included skeletal muscles, including the proximal portion of the thin diaphragm and crus as much as recognizable, resulting in the exclusion of the central tendon of the diaphragm. The AVF mask included fatty tissues of peritoneal and retroperitoneal spaces (18) that were distributed below the xiphoid process and diaphragmatic hiatus to the perirectal fat. In the SF mask, we did not separate SF from intramuscular fat and intrathoracic visceral fat and combined those fatty tissues. The IO mask contained intraabdominal and intrathoracic solid organs, hollow viscera, breast tissue, external genital organs, and the thyroid gland, along with the major arteries and veins. The distal ends of separation between muscle and vessels were the upper end of the thyroid superiorly, the proximal biceps brachii laterally, and the upper one-third of the femur inferiorly. The vessels beyond the distal ends were included in the muscle mask. We excluded parotid glands from any of the masks while including salivary and submandibular glands in the muscle mask, as the latter two glands could not be separated from muscle on noncontrast CT images. The CNS mask comprised the eyeball, intracranial structures, and the spinal canal.

A technician and one radiology resident (Y.S.L. with 3 years of body CT interpretation) initially segmented the CT images by using thresholds of Hounsfield units (HU), providing rough masks of muscle, fat, and bone as follows: -29 to 150 HU for skeletal muscle, -190 to -30 HU for adipose tissue, >250 HU for bone. Then, one body radiologist (S.H.Y., with 15 years of body CT interpretation) adjusted the masks in every axial CT image slice for the training dataset of 100 CT scans by manual correction, morphological closing,

subtraction, and merging. Two body radiologists (J.H.H. and J.H.P. with 6 and 5 years of body CT interpretation, each) prepared the reference masks for the first and second external validation data of the 20 CT scans in a similar way, respectively. Reference masks in the third external validation data were prepared by the French nuclear medicine physician (P.D. with 8 years of experience in PET/CT scan).

Development and materialization of the deep neural network

The 100 CT scans were randomly assigned into one of the three following data sets: 80 cases for the training set, 10 cases for the tuning set, and 10 cases for the internal validation set. Most of the CT scans consisted of 2-mm-section CT images that covered the head to the lower arm and mid-thigh. In total, 39,286 slices of axial data were available in the dataset. The training data of CT images were normalized using the soft tissue window setting. The two-dimensional (2D) U-Net received an input size of $512 \times 512 \times 1$ and consisted of initial convolutions, four encoders, four decoders, and a final convolution. Except for the final convolution (a 1×1 convolution), every convolutional layer consisted of a 3×3 convolution followed by batch normalization (19) and the rectified linear unit activation function (20). For decoders, up-sampling with the bilinear interpolation method was used, followed by concatenation to conserve information before down-sampling. The 3D U-Net received an input size of $512 \times 512 \times 8$ and used three encoders and three decoders. Except for the final convolution (a $1 \times 1 \times 1$ convolution), every convolutional layer consisted of a $3 \times 3 \times 3$ convolution, followed by ReLU and group normalization. The first encoder adopted $1 \times 2 \times 2$ max pooling to preserve data in the z-axis, whereas the second and third encoders used $2 \times 2 \times 2$ max pooling. For decoders, up-sampling with tri-linear interpolation was used. The

Kaiming He initialization method (21) was used for weight initialization. Both models used a softmax function at the final layer and were trained using the stochastic gradient descent algorithm and cross-entropy loss function.

The 2D and 3D U-Net models were packaged as a standalone software (DeepCatch v1.0.0.0; <http://www.medicalip.com/Download>) by a functional modification of the FDA-approved MEDIP PRO v2.0.0.0. The specifications required for running the U-Net models are as follows: operating system, Microsoft Windows 7 (64 bit; IBM Corporation, San Jose, CA) or higher; central processing unit, Intel i5 (Intel Corporation, Santa Clara, CA) or higher; random-access memory, 8 GB or higher; and graphics processing unit, GeForce 1000 series or higher (NVIDIA Corporation, Santa Clara, CA) with the minimum memory of 2 GB or larger for 2D U-Net and 8 GB or larger for 3D-UNet supporting the Compute Unified Device Architecture with the latest drivers. The segmentation was accomplished in around 45 seconds for the 2D U-Net model and in around 75 seconds for the 3D U-Net model with the recommended specifications.

Clinical validation cohort: KURE

The KURE cohort is a government-funded cohort of community-based Korean older adults, in whom cardiovascular, musculoskeletal, and age-related diseases were studied (17). A total of 3517 subjects participated in the study between 2012 and 2015, and a follow-up study was completed between 2016 and 2019. CT images of the torso area (from neck to mid-thigh level) were obtained in the follow-up study through a comprehensive assessment of musculoskeletal function, physical performance, and radiological and biochemical measurements. Because CT images were only available at follow-up, a subset of the data of

the KURE cohort collected in 2016 was analyzed as the clinical validation set. Details of the KURE cohort can be obtained in the cohort profile (22). The KURE population was included in prior studies on study protocol (17,22), vitamin D (23-25), dysmotility (26), and metabolic syndrome (27,28).

All CT scans in KURE participants were performed using SOMATOM Definition CT scanner (Siemens Healthcare, Germany) according to a predefined cohort protocol (noncontrast CT, slice thickness 3 mm, tube voltage 100 kV, tube current 200 mA, reconstruction with standard kernel).

Definition of sarcopenia and dysmobility syndrome

Sarcopenia was defined based on Asian Working Group for Sarcopenia 2019 consensus statement as low appendicular lean mass with low muscle strength or low physical performance (22,29). Appendicular lean mass was assessed using multi-frequency segmental bioimpedance analysis (InBody720, Biospace, Seoul, Korea; appendicular lean mass/height² < 7.0 kg/m² in men; < 5.7 kg/m² in women). Low muscle strength was defined as maximum handgrip strength < 28 kg in men or < 18 kg in women. Low physical performance was measured by the five-time chair stand test (12 seconds or higher). Dysmobility syndrome was defined as having at least three of the following factors using modified definitions: falls in the preceding year, low grip strength (< 28 kg in men or < 18 kg in women), low bone mass (quantitative computed tomography-based volumetric trabecular bone mineral density in the lumbar spine < 80 mg/mm³)(30), low appendicular lean mass (appendicular lean mass/height² < 7.0 kg/m² in men; < 5.7 kg/m² in women), high fat mass (> 30% in men; > 40% in women), and low performance on the timed get-up-and-go test (12 seconds or higher) (31,32).

Statistical analysis

In the development cohorts, we compared the reference masks obtained through semi-automatic segmentation with U-Net-driven masks using the Dice similarity coefficient (DSC), sensitivity, and positive predictive value (PPV)(33). The DSC provided the degree of overlap between U-Net-driven and reference masks. The sensitivity measured the portion of voxels in the reference mask that were also included in the U-Net mask. PPV measured the ratio of pixels in the U-Net-driven mask that was correctly present in the reference mask (Figure S1). The paired t-test was used to compare the segmentation performance between the 2D and 3D U-Net. The independent t-test was used to compare segmentation performance depending upon sex, age (≤ 70 years vs. >70 years), body mass index (BMI) (≤ 23 kg/m² vs. >23 kg/m²), and CT scan range (T12 or above vs. below T12).

In the clinical validation cohort, we calculated the 3D U-Net-driven torso volume (cm³) of skeletal muscle and fat tissue and average skeletal muscle area (SMA, cm²) and body fat area (BFA, cm²) in the waist by dividing the volume of those tissues in the waist by the craniocaudal length of the waist. The waist ranges from the lower margin of the last rib to the upper margin of iliac crest (34), containing the L3 vertebra (35). The average SMA and BFA represented the mean distribution of SMA and BFA in the waist and could be used equivalently to single-slice measures at L3 (2). Skeletal muscle index (SMI, cm²/m²) and body fat index (BFI, cm²/kg) were also calculated by dividing SMA and BFA by height squared and kilograms, respectively. We assessed the correlation between 3D U-Net-driven CT parameters and BIA-based values. Linear regression models were used to estimate the cutoffs of 3D U-Net-derived SMI and BFI that corresponded to the currently endorsed BIA-

based diagnostic thresholds for sarcopenia or dysmobility syndrome (low appendicular lean mass [appendicular lean mass/height² <7.0 kg/m² in men; <5.4 kg/m² in women] and high fat mass [body fat percent >30% in men; >40% in women])(29,31,32,36). Multivariable logistic regression models were used to evaluate the 3D U-Net-driven average SMA and BFA in the waist with sarcopenia and dysmobility syndrome as a continuous or binary variable based on the cutoffs while adjusting for covariates. We compared the area under the receiver operating characteristic curve (AUROC) to evaluate whether the 3D U-Net-derived average SMA and BFA in the waist could improve diagnostic performance when added to age and sex (37). All analyses were performed using STATA version 14.1 (StataCorp, College Station, TX, USA). The statistical significance level was set at a two-sided *P*-value of <0.05.

Results

Segmentation performance in the development set

The mean age and sex distribution of the subjects in the development set (internal set) and the three external validation sets for segmentation accuracy were 62.3 ± 13.3 years (women, 60%), 68.0 ± 10.5 years (women, 45%), 74.0 ± 3.7 years (women, 55%), and 58.5 ± 9.3 years (women, 17%), respectively (Table 1). Both the 2D and 3D U-Net achieved accurate body composition segmentation (Table 2 and Figure 2), and the 3D U-Net performed significantly better than the 2D U-Net for all seven masks (Table S1). The average DSCs between the U-Net results and the reference for all masks were as follows: the internal dataset, $95.3 \pm 3.0\%$ and $96.9 \pm 2.4\%$, each; the first external dataset, $93.3 \pm 1.0\%$ and $96.5 \pm 1.1\%$, each; and the second external dataset, $95.6 \pm 1.1\%$ and $98.9 \pm 0.4\%$, each. When confined to muscle, AVF, and SF masks, the average DSCs between the 2D and 3D U-Net results and the reference masks were as follows: the internal dataset, $94.6 \pm 5.6\%$ and $96.8 \pm 3.5\%$, each; the first external dataset, $94.4 \pm 2.7\%$ and $97.0 \pm 0.9\%$, each; the second validation dataset, $96.2 \pm 2.5\%$ and $99.3 \pm 0.9\%$, each; the third external dataset, $90.3 \pm 3.6\%$ and $92.3 \pm 2.2\%$, each. The segmentation performance of the 3D U-Net was generally not affected by age, sex, BMI, or scan range (Table S2).

Correlation between 3D U-Net parameters and BIA

In the KURE cohort, the Pearson correlation coefficients between the 3D U-Net-driven CT skeletal muscle measurements and BIA-based appendicular lean mass values ranged from 0.59 to 0.72 and from 0.56 to 0.66 in men and women, respectively (Table 3).

Body fat mass and percent measured by BIA showed excellent correlations with CT-based torso fat volume and waist BFA in both men and women (correlation coefficient, 0.82 to 0.95; $P < .001$ for all).

The average SMI values of the waist of $40 \text{ cm}^2/\text{m}^2$ (men) and $35 \text{ cm}^2/\text{m}^2$ (women) corresponded to the BIA-based clinical thresholds for low appendicular lean mass ($7.0 \text{ kg}/\text{m}^2$ in men and $5.7 \text{ kg}/\text{m}^2$ in women) in the linear regression model (Figure 2). The BFI values of the waist of $4.7 \text{ cm}^2/\text{kg}$ (men) and $5.8 \text{ cm}^2/\text{kg}$ (women) corresponded to the BIA-based high fat mass thresholds (30% in men and 40% in women), respectively.

Associations of 3D U-Net body composition assessments with sarcopenia

In the KURE cohort, which was used for clinical validation, study participants with sarcopenia ($n=85$, 16%) had older age, lower BMI, and higher prevalence of low handgrip strength, low lean mass, and low physical performance measured by the timed get-up-and-go test or the five-time chair rise test compared to those without sarcopenia (Table S3, $P < .05$ for all). Individuals with dysmobility syndrome ($n=149$, 29%) showed similar results to those of subjects with sarcopenia except that they had a higher prevalence of high fat mass than those without dysmobility syndrome. Compared to subjects without sarcopenia, individuals with sarcopenia had lower 3D U-Net based CT torso volumetric SMI (3348 vs. 3826 cm^3/m^2), waist areal SMI (38 vs. 43 cm^2/m^2), torso volumetric BFI (197 vs. 233 cm^3/kg), and waist areal BFI (4.0 vs. 4.8 cm^2/kg ; $P < .001$ for all).

Diagnostic value of 3D U-Net assessments for sarcopenia and dysmobility syndrome

A low average SMI in the waist using the cutoffs of 40 cm²/m² in men and 35 cm²/m² in women alone showed 82% and 71% agreement with a clinical diagnosis of sarcopenia (sensitivity, 38%; specificity, 90%; PPV, 43%; negative predictive value, 88%) and dysmobility syndrome (sensitivity, 24%; specificity, 90%; PPV, 49%; negative predictive value, 75%), respectively.

A one-standard deviation decrease of the average SMI in the waist (cm²/m²) was associated with 2.46- and 1.81-fold elevated odds for sarcopenia and dysmobility syndrome, respectively ($P < .001$ for all)(Table 4). Low average SMI (cm²/m²) and high BFI (cm²/kg) in the waist according to the cutoffs were independently associated with sarcopenia (low SMI, adjusted odds ratio [aOR] 5.35; high BFI, aOR 0.51) and dysmobility syndrome (low SMI, aOR 3.48; high BFI, aOR 2.34; $P < .05$ for all) after adjusting for age and sex. In receiver operating characteristic curves (Figure 3), the average SMA and BFA in the waist provided an AUROC of 0.81 (95% CI 0.75 to 0.85) for sarcopenia and an AUROC of 0.66 (95% CI 0.60 to 0.70) for dysmobility syndrome. The average SMA and BFA in the waist improved diagnostic performance for sarcopenia (AUROC 0.66 to 0.86, $P < .001$) and dysmobility syndrome (AUROC 0.70 to 0.76, $P = .006$) when added to age and sex.

Discussion

The 2D and 3D U-Nets developed in this study were validated in several datasets, including the three of the four major clinical CT vendors (38). The correlations observed between the U-Nets and the reference values in the internal validation dataset were relatively well reproduced in the external validation datasets. The 3D U-Net performed better than the 2D U-Net for all masks, and the performance of the 3D U-Net was mostly maintained regardless of age, sex, BMI, and scan range. The magnitude of the correlations was slightly lower in the French public dataset than in the other two external validation datasets. This partly originated from differences in the individual standards for muscle (i.e., whether they contained axillary vessels) and AVF (i.e., whether some feces and outer boundaries outside the pelvic floor muscle were contained) in the French public dataset.

Sarcopenia has long been considered a disease of the elderly. However, recent studies suggest that sarcopenia begins earlier in life (39), underscoring the importance of early detection and intervention. Nevertheless, the sarcopenia or preclinical stage of sarcopenia is under-detected, as clinical suspicion or the SARC-F questionnaire for sarcopenia merely initiates the current diagnostic approach of sarcopenia. More than 100 million diagnostic CT examinations of various body parts are performed annually in the world, and early detection of sarcopenia could be improved if sarcopenia is adjunctively evaluated using diagnostic CT scans with a deep neural network for body composition analysis. Indeed, our deep neural network showed excellent correlations with the results of BIA, and provided a diagnostic accuracy of 0.86 for sarcopenia when combined with age and sex.

The seven classes contained all body areas except for the lung and airway that are segmentable using pre-existing mathematical or deep learning approaches (40). A skin mask

can be used to directly measure body surface area (41) or various body circumference values (42) on CT images. The body surface area is crucial for determining the dosage of chemotherapeutic drugs and assessing renal clearance or the cardiac index. The area is generally estimated well by various formulae, but the estimation is less accurate in underweight or overweight individuals and children (43), who potentially benefit from direct measurements. A bone mask can allow a volumetric assessment of bone mineral density in multiple bones simultaneously. IO and CNS masks can serve as input data that accelerate the development of deep neural networks for segmenting particular organs or intra-organ abnormalities.

Our study has several limitations. First, the amount of validation data for assessing the DSC was relatively small, which was somewhat inevitable as the required time and resources for preparing the reference dataset were substantial. Second, our datasets preferentially included the elderly population. Third, the muscle mask inevitably contained distal peripheral vessels, and intrathoracic visceral fat and intramuscular fat were included in the SF mask. Fourth, we only validated the U-Net-driven muscle and fat volume in the abdominal area in the elderly cohort, although the network was applicable to other body parts (e.g., thoracic area). Fifth, we only presented the results of the U-Net models for noncontrast CT scan due to difficulty in preparing a sufficient number of reference data, while the 3D U-Net for contrast CT scan was also developed. The 3D U-Net showed a similar DSC of 95.4% for all masks (muscle, 96.7%, AVF, 95.1%; SF, 96.2%) in the preliminary analysis of the internal dataset. Sixth, the training dataset lacked any patients with ascites, major anatomical variations, or diffuse subcutaneous edema; therefore, it is uncertain whether this algorithm will also work well in such cases.

Conclusion

A deep neural network enabled the automatic volumetric segmentation of body composition on whole-body CT images, and the 3D U-Net-derived average areas of skeletal muscle and fatty tissue in the waist were independently associated with sarcopenia and dysmobility syndrome in the elderly. The network-driven automatic volumetric segmentation of CT images of PET-CT scan will enable body composition analysis as adjunctive output of PET-CT scan and volumetric analysis of metabolism in whole-body fat and muscle.

Acknowledgement

The authors would like to acknowledge Andrew Dombrowski, PhD (Compecs, Inc.) for his assistance in improving the use of English in this manuscript.

Conflicts of interest

S.H.Y. joined the MEDICAL IP, Co. Ltd. as a chief medical officer since November 2020. J.N.W. is an employer and machine learning researcher of MEDICAL IP, Co. Ltd. Other authors have no conflicts of interest to declare.

Author contributions

Conceptualization, , Y.S.L., N.H., Y.R., S.H.Y., J.M.G.; Data curation, Y.S.L., Y.R.C., J.P., P.D., E.F., C.O.K., H.C.K., Y.R., S.H.Y. ; Formal analysis, Y.S.L., N.H., Y.R., S.H.Y.;

Investigation, Y.S.L., N.H., Y.R., S.H.Y.; Methodology, Y.S.L., N.H., Y.R., S.H.Y.; Project administration, N.H., Y.R., S.H.Y.; Resources, N.H., Y.R., S.H.Y.; Software, J.N.W.; Supervision, Y.R., S.H.Y.; Validation. Y.R.C., J.P., P.D., E.F., N.H., Y.R., S.H.Y. Visualization, Y.S.L., N.H., Y.R., S.H.Y.; Roles/Writing - original draft, Y.S.L., N.H., Y.R., S.H.Y.; Writing - review & editing, all authors.

Financial support

None.

References

1. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. *Lancet*. 2019;393:2636-2646.
2. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing*. 2019;48:16-31.
3. Prado CM, Heymsfield SB. Lean tissue imaging: a new era for nutritional assessment and intervention. *JPEN J Parenter Enteral Nutr*. 2014;38:940-953.
4. Buckinx F, Landi F, Cesari M, et al. Pitfalls in the measurement of muscle mass: a need for a reference standard. *J Cachexia Sarcopenia Muscle*. 2018;9:269-278.
5. Mitsiopoulos N, Baumgartner RN, Heymsfield SB, Lyons W, Gallagher D, Ross R. Cadaver validation of skeletal muscle measurement by magnetic resonance imaging and computerized tomography. *J Appl Physiol (1985)*. 1998;85:115-122.
6. Shen W, Punyanitya M, Wang Z, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. *J Appl Physiol (1985)*. 2004;97:2333-2338.
7. 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;33:997-1006.
8. Shen W, Punyanitya M, Wang ZM, et al. Visceral adipose tissue: relations between single-slice areas and total volume. *American Journal of Clinical Nutrition*. 2004;80:271-278.
9. Boutin RD, Yao L, Canter RJ, Lenchik L. Sarcopenia: Current Concepts and Imaging Implications. *American Journal of Roentgenology*. 2015;205:W255-W266.
10. Baracos VE, Arribas L. Sarcopenic obesity: hidden muscle wasting and its impact for survival and complications of cancer therapy. *Annals of Oncology*. 2018;29:1-9.
11. Weston AD, Korfiatis P, Kline TL, et al. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. *Radiology*. 2019;290:669-679.
12. Nowak S, Faron A, Luetkens JA, et al. Fully Automated Segmentation of Connective Tissue Compartments for CT-Based Body Composition Analysis: A Deep Learning Approach. *Invest Radiol*. 2020;55:357-366.
13. von Schulthess GK, Steinert HC, Hany TF. Integrated PET/CT: current applications and future directions. *Radiology*. 2006;238:405-422.
14. Poeppel TD, Krause BJ, Heusner TA, Boy C, Bockisch A, Antoch G. PET/CT for the staging and follow-up of patients with malignancies. *Eur J Radiol*. 2009;70:382-392.

15. Martin L, Birdsell L, Macdonald N, et al. Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. *J Clin Oncol*. 2013;31:1539-1547.
16. Shachar SS, Williams GR, Muss HB, Nishijima TF. Prognostic value of sarcopenia in adults with solid tumours: A meta-analysis and systematic review. *Eur J Cancer*. 2016;57:58-67.
17. Lee EY, Kim HC, Rhee Y, et al. The Korean urban rural elderly cohort study: study design and protocol. *BMC Geriatr*. 2014;14:33.
18. Hung CS, Lee JK, Yang CY, et al. Measurement of visceral fat: should we include retroperitoneal fat? *PLoS One*. 2014;9:e112355.
19. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:150203167*. 2015.
20. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Paper presented at: Advances in neural information processing systems, 2012.
21. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Paper presented at: Proceedings of the IEEE international conference on computer vision, 2015.
22. Hong N, Kim KJ, Lee SJ, et al. Cohort profile: Korean Urban Rural Elderly (KURE) study, a prospective cohort on ageing and health in Korea. *BMJ Open*. 2019;9:e031018.
23. Song BM, Rhee Y, Kim CO, et al. Urban-rural differences explain the association between serum 25-hydroxyvitamin D level and insulin resistance in Korea. *Nutrients*. 2014;6:5806-5818.
24. Song BM, Kim HC, Rhee Y, Youm Y, Kim CO. Association between serum 25-hydroxyvitamin D concentrations and depressive symptoms in an older Korean population: A cross-sectional study. *J Affect Disord*. 2016;189:357-364.
25. Lee EY, Lee SJ, Kim KM, et al. Association of metabolic syndrome and 25-hydroxyvitamin D with cognitive impairment among elderly Koreans. *Geriatr Gerontol Int*. 2017;17:1069-1075.
26. Hong N, Kim CO, Youm Y, Choi JY, Kim HC, Rhee Y. Dymobility syndrome is associated with prevalent morphometric vertebral fracture in older adults: the Korean Urban-Rural Elderly (KURE) study. *Arch Osteoporos*. 2018;13:86.
27. Choi H, Kim HC, Song BM, et al. Serum uric acid concentration and metabolic syndrome among elderly Koreans: The Korean Urban Rural Elderly (KURE) study. *Arch Gerontol Geriatr*. 2016;64:51-58.
28. Choi H, Kim HC, Stefani KM, et al. Serum high-density lipoprotein cholesterol concentration and functional state: The Korean Urban Rural Elderly (KURE) Study. *Arch Gerontol Geriatr*. 2017;71:115-121.
29. Chen LK, Woo J, Assantachai P, et al. Asian Working Group for Sarcopenia: 2019 Consensus Update on Sarcopenia Diagnosis and Treatment. *J Am Med Dir Assoc*. 2020;21:300-307.e302.

30. Engelke K, Adams JE, Armbrrecht G, et al. Clinical use of quantitative computed tomography and peripheral quantitative computed tomography in the management of osteoporosis in adults: the 2007 ISCD Official Positions. *J Clin Densitom.* 2008;11:123-162.
31. Hong N, Kim CO, Youm Y, Kim HC, Rhee Y. Low peak jump power is associated with elevated odds of dysmobility syndrome in community-dwelling elderly individuals: the Korean Urban Rural Elderly (KURE) study. *Osteoporos Int.* 2018;29:1427-1436.
32. Binkley N, Krueger D, Buehring B. What's in a name revisited: should osteoporosis and sarcopenia be considered components of "dysmobility syndrome?". *Osteoporos Int.* 2013;24:2955-2959.
33. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging.* 2015;15:29.
34. Ross R, Neeland IJ, Yamashita S, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nature Reviews Endocrinology.* 2020;16:177-189.
35. Chakraverty R, Pynsent P, Isaacs K. Which spinal levels are identified by palpation of the iliac crests and the posterior superior iliac spines? *Journal of Anatomy.* 2007;210:232-236.
36. Chen LK, Liu LK, Woo J, et al. Sarcopenia in Asia: consensus report of the Asian Working Group for Sarcopenia. *J Am Med Dir Assoc.* 2014;15:95-101.
37. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837-845.
38. Davidson A. CT Equipment Market: Made in China...and the USA. <https://technology.informa.com/606142/ct-equipment-market-made-in-chinaand-the-usa>. Accessed April 4th, 2020.
39. Sayer AA, Syddall H, Martin H, Patel H, Baylis D, Cooper C. The developmental origins of sarcopenia. *J Nutr Health Aging.* 2008;12:427-432.
40. Mansoor A, Bagci U, Foster B, et al. Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. *Radiographics.* 2015;35:1056-1076.
41. Iannessi A, Beaumont H, Hebert C, Dittlot C, Falewee MN. Computer tomography-based body surface area evaluation for drug dosage: Quantitative radiology versus anthropomorphic evaluation. *PLoS One.* 2018;13:e0192124.
42. Heymsfield SB, Martin-Nguyen A, Fong TM, Gallagher D, Pietrobelli A. Body circumferences: clinical implications emerging from a new geometric model. *Nutr Metab (Lond).* 2008;5:24.
43. Villa C, Primeau C, Hesse U, Hougen HP, Lynnerup N, Hesse B. Body surface area determined by whole-body CT scanning: need for new formulae? *Clin Physiol Funct Imaging.* 2017;37:183-193.

Tables

Table 1. Composition of the internal and external datasets for model development and validation

	Internal dataset	External validation datasets			
		The first dataset	The second dataset	The third dataset	The fourth dataset
Purpose	Model development	Assessing segmentation accuracy			Clinical validation
Country	Korea	Korea	Korea	France	Korea
Number of patients	100	20	20	18	522
Number of CT scans	100	20	20	24	522
Age (years)	62.3±13.3	68.0±10.5	74.0±3.7	58.5±9.3	75.3 ± 4.5
Sex (male:female)	40:60	11:9	9:11	15:3	170:352
Body mass index	23.1±3.9 (13.8-31.2)	23.4±2.5 (17.9-26.4)	23.5±2.6 (16.4-29.4)	Unknown	24.4 ± 3.2 (14.9-36.9)
CT Vendors	Siemens Healthineers; Philips Medical Systems	Philips Medical Systems	Siemens Healthineers	GE Healthcare	Siemens Healthineers
CT machines	mBiograph40, mBiograph64; GEMINI	GENIMI TF TOF 64	SOMATOM Definition AS+	Discovery ST, HR, RX, and STE	SOMATOM Definition AS+
CT scan range	Head, neck, torso, and arm, 64; Whole body, 36	Head, neck, torso, and arm	Torso	Torso	Torso
CT slice thickness	2-6.5mm	4mm	3mm	5mm	3mm
CT slice numbers	145-992	213-255	200-269	117-232	239-268
Tube voltage (kVp)	100-120	120	100	120	100
Tube current (Ref mAs)	40-67	50	200	300	200

Abbreviation: CT, computed tomography; kVp, peak kilovoltage; Ref mAs, Reference milliampere-second. Data in the parenthesis indicates

the range of body mass index. All CT scans were obtained without contrast enhancement.

Table 2. Dice similarity coefficients, sensitivity, and precision between the reference masks and 3D-UNet-derived masks in internal and external validation datasets

	Skin			Bone			Muscle			Abdominal visceral fat			Subcutaneous fat			Internal organs and vessels			Central nervous system		
	DSC	Sens	PPV	DSC	Sens	PPV	DSC	Sens	PPV	DSC	Sens	PPV	DSC	Sens	PPV	DSC	Sens	PPV	DSC	Sens	PPV
2D U-Net																					
Internal dataset	92.2	90.4	94.5	97.6	96.8	98.6	97.2	97.6	96.9	90.5	87.1	94.9	96.1	96.7	95.5	96.3	96.3	96.3	97.5	97.1	98.0
	±5.6	±9.3	±3.2	±2.1	±3.8	±0.4	±1.0	±0.9	±1.4	±7.8	±11.5	±5.0	±3.1	±3.7	±3.0	±1.3	±1.4	±1.4	±3.3	±3.7	±3.1
External dataset 1	80.7	84.0	77.6	96.3	95.3	97.3	95.7	95.4	95.9	93.1	91.7	94.6	94.3	94.7	94.0	94.5	94.5	94.4	98.4	98.5	98.5
	±0.8	±1.0	±1.6	±0.5	±0.7	±0.5	±0.7	±0.9	±0.6	±3.6	±5.0	±2.2	±2.5	±3.1	±2.1	±0.8	±1.1	±1.6	±0.2	±0.4	±0.3
External dataset 2	93.9	93.8	94.1	97.7	96.6	98.8	97.0	96.7	97.3	94.2	94.1	94.2	97.4	99.4	95.4	96.0	95.7	96.4	92.8	90.7	95.2
	±1.8	±1.0	±3.1	±0.6	±0.9	±0.4	±0.8	±1.4	±0.8	±3.5	±4.0	±3.1	±0.8	±0.3	±1.3	±1.5	±1.0	±2.3	±2.8	±4.6	±2.9
External dataset 3							89.9	86.9	93.2	88.3	87.3	90.0	92.7	97.3	88.5						
							±2.2	±2.3	±3.1	±8.2	±11.7	±3.9	±2.9	±0.8	±4.9						
3D U-Net																					
Internal dataset	94.0	92.6	95.9	98.1	98.2	98.0	98.1	98.5	97.8	95.1	94.3	96.0	97.1	96.2	98.0	97.6	97.4	97.8	98.0	97.9	98.2
	±4.5	±7.3	±3.8	±2.0	±3.0	±1.0	±0.6	±0.8	±0.7	±4.9	±7.5	±2.5	±3.1	±4.2	±2.0	±0.8	±0.7	±1.3	±2.9	±3.4	±3.0
External dataset 1	95.4	94.5	96.3	98.2	98.8	97.5	96.8	96.0	97.5	96.3	97.0	95.7	98.0	97.8	98.2	94.2	96.1	92.5	96.8	98.5	95.2
	±0.8	±1.0	±0.9	±0.5	±0.5	±0.8	±1.3	±2.2	±0.5	±2.2	±1.4	±3.6	±1.0	±1.1	±1.0	±2.5	±1.2	±4.1	±2.0	±0.3	±3.8
External dataset 2	98.6	98.0	99.3	99.4	99.7	99.1	99.2	99.5	99.0	98.9	99.1	98.7	99.7	99.6	99.7	98.7	98.4	99.0	98.4	97.2	99.7
	±1.1	±1.5	±0.7	±0.5	±0.3	±0.8	±0.5	±0.5	±0.6	±1.4	±0.9	±2.1	±0.3	±0.4	±0.3	±0.8	±0.8	±1.1	±2.0	±3.6	±0.5
External dataset 3							90.3	86.8	94.1	92.4	91.8	93.2	94.1	96.8	91.6						
							±1.7	±2.0	±2.8	±4.7	±6.8	±3.0	±2.5	±0.9	±4.2						

Abbreviation: DSC=dice similarity coefficient; Sens=sensitivity; PPV=positive predictive value.

Data in columns indicate percentages of mean±standard deviation. The third external dataset, French public dataset, only contained masks for muscle, abdominal visceral, and subcutaneous fat tissues.

Table 3. Correlations between 3D U-Net-derived CT parameters and BIA values for body composition assessment

Comparator (3D U-Net-driven CT parameters)	Reference (BIA)	Men (n=170)		Women (n=352)	
		Correlation coefficients	P-value	Correlation coefficients	P-value
Torso skeletal muscle volume (cm ³)	ALM (kg)	0.71	<.001	0.66	<.001
Average skeletal muscle area in the waist (cm ²)	ALM (kg)	0.72	<.001	0.65	<.001
Torso skeletal muscle index (cm ³ /m ²)	ALMI (kg/m ²)	0.59	<.001	0.56	<.001
Average skeletal muscle index in the waist (cm ² /m ²)	ALMI (kg/m ²)	0.60	<.001	0.58	<.001
Torso fat volume (cm ³)	Fat mass (kg)	0.95	<.001	0.94	<.001
Average fat area in the waist (cm ²)	Fat mass (kg)	0.93	<.001	0.92	<.001
Torso body fat index (cm ³ /kg)	Body fat percent (%)	0.90	<.001	0.86	<.001
Average body fat index in the waist (cm ² /kg)	Body fat percent (%)	0.86	<.001	0.82	<.001

Abbreviation: BIA, body impedance analysis; ALMI, appendicular lean mass index.

Table 4. Logistic regression analyses for sarcopenia and dysmobility syndrome

	Univariate model		Multivariable model 1		Multivariable model 2	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
Sarcopenia						
Average SMA in the waist, per 1 sd decrease*	2.46 (1.80-3.35)	<0.001	4.03 (2.55-6.36)	<0.001	-	-
Average BFA in the waist, per 1 sd increase+	0.39 (0.30-0.51)	<0.001	0.51 (0.38-0.69)	<0.001	-	-
Low average SMI in the waist (vs. high)**	5.67 (3.30-9.76)	<0.001	-	-	5.35 (3.03-9.46)	<0.001
High average BFI in the waist (vs. low)**	0.40 (0.21-0.75)	0.004	-	-	0.43 (0.22-0.84)	0.013
Age, per 1 year increase	1.13 (1.08-1.19)	<0.001	1.10 (1.04-1.17)	0.001	1.14 (1.08-1.20)	<0.001
Women (vs. men)	0.72 (0.44-1.16)	0.180	0.17 (0.07-0.39)	<0.001	0.88 (0.51-1.50)	0.644
Dysmobility syndrome						
Average SMA in the waist, per 1 sd decrease*	1.81 (1.45-2.26)	<0.001	3.47 (1.99-6.04)	<0.001	-	-
Average BFA in the waist, per 1 sd increase+	1.05 (0.88-1.25)	0.573	2.33 (1.48-3.67)	<0.001	-	-
Low average SMI in the waist (vs. high)**	2.81 (1.70-4.65)	<0.001	-	-	3.48 (1.99-6.04)	<0.001
High average BFI in the waist (vs. low)**	1.81 (1.21-2.72)	0.004	-	-	2.34 (1.49-3.67)	<0.001
Age, per 1 year increase	1.15 (1.10-1.21)	<0.001	1.18 (1.12-1.24)	<0.001	1.19 (1.12-1.25)	<0.001
Women (vs. men)	1.60 (1.05-2.45)	0.030	2.56 (1.58-4.16)	<0.001	2.57 (1.58-4.16)	<0.001

Abbreviation: SMA, skeletal muscle area (cm²); BFA, body fat area (cm²); SMI, skeletal muscle index (cm²/m²); BFI, body fat index (cm²/kg).

*1sd=20 cm²; +1sd=100 cm²; **CT body composition parameters were entered into models separately as continuous variables (average SMA and FA in the waist) or as categorical variables (Low average SMI or high average BFI in the waist). 40 cm²/m² and 35 cm²/m² in men and women were used for defining the low average SMI, and 4.7 cm²/kg and 5.8 cm²/kg in men and women were used for the high average

BFI.

Supplemental tables

Table S1. Comparison of segmentation performances of 2D and 3D U-Net in the validation datasets

	Dice similarity coefficient			Sensitivity			Positive predictive value		
	2D U-Net	3D U-Net	P-value	2D U-Net	3D U-Net	P-value	2D U-Net	3D U-Net	P-value
Skin †	88.3±6.9%	96.4±2.8%	<u><0.001</u>	89.2±6.0%	95.5±4.0%	<u><0.001</u>	87.6±8.6%	97.4±2.3%	<u><0.001</u>
Bone †	97.1±1.2%	98.6±1.1%	<u><0.001</u>	96.1±1.9%	99.0±1.5%	<u><0.001</u>	98.1±0.8%	98.3±1.1%	0.160
Muscle ‡	94.3±3.5%	95.5±3.9%	<u><0.001</u>	94.3±3.5%	94.3±5.6%	<u><0.001</u>	95.5±2.6%	96.8±2.6%	<u><0.001</u>
AVF ‡	91.5±6.4%	95.6±4.3%	<u><0.001</u>	91.5±6.4%	95.5±5.6%	<u><0.001</u>	93.0±4.0%	95.7±3.6%	<u><0.001</u>
SF ‡	94.8±3.1%	97.1±2.9%	<u><0.001</u>	94.8±3.1%	97.8±2.1%	<u>0.006</u>	92.8±4.4%	96.4±4.3%	<u><0.001</u>
IO †	95.5±1.5%	96.7±2.7%	<u><0.001</u>	95.3±1.3%	97.3±1.4%	<u><0.001</u>	95.6±2.1%	96.1±4.1%	0.234
CNS †	96.0±3.5%	97.7±2.3%	<u>0.004</u>	95.0±4.9%	97.9±2.7%	<u><0.001</u>	97.1±2.8%	97.6±3.4%	0.432
All masks †	94.6±1.9%	97.6±1.8%	<u><0.001</u>	94.4±2.1%	97.6±1.9%	<u><0.001</u>	94.9±2.0%	97.6±1.8%	<u><0.001</u>

Abbreviation: AVF, abdominal visceral fat; SF, subcutaneous fat; IO, internal organs and vessels; CNS, central nervous system.

Data in the columns indicates a mean value ± standard deviation.

†Statistical comparisons were performed in the pooled data of the internal dataset, the first, and the second external datasets.

‡Statistical comparisons were performed in the pooled data of the internal dataset, the first, second, third external datasets.

P-values underlined in italic type indicate statistical significance.

Table S2. Comparison of segmentation performances of 3D U-Net in the validation datasets according to sex, age, BMI, and scan range

Masks	Sex	Dice score	P-value	Age	Dice score	P-value	BMI	Dice score	P-value	Scan range	Dice score	P-value
Skin	Male	96.6±2.1%	0.725	Age≤70	95.4±3.5%	0.052	BMI≤23	96.2±3.7%	0.637	≥T12	95.9±4.0%	0.418
	Female	96.3±3.3%		Age>70	97.1±1.9%		BMI>23	96.6±1.8%		<T12	96.5±3.2%	
Bone	Male	98.7±0.9%	0.761	Age≤70	98.4±1.5%	0.220	BMI≤23	98.6±1.5%	0.803	≥T12	98.5±1.2%	0.402
	Female	98.6±1.3%		Age>70	98.8±0.8%		BMI>23	98.7±0.8%		<T12	98.7±1.2%	
Muscle	Male	97.7±1.7%	0.120	Age≤70	97.7±1.5%	0.209	BMI≤23	98.3±1.3%	0.305	≥T12	97.8±1.3%	0.306
	Female	98.3±1.0%		Age>70	98.2±1.4%		BMI>23	97.8±1.5%		<T12	98.1±1.8%	
AVF	Male	97.2±2.7%	0.808	Age≤70	96.7±3.5%	0.518	BMI≤23	96.4±4.2%	0.141	≥T12	90.6±17.2%†	<u>0.007</u>
	Female	97.0±3.4%		Age>70	97.3±2.8%		BMI>23	97.7±1.6%		<T12	97.3±2.8%	
SF	Male	98.4±1.3%	0.903	Age≤70	98.0±2.4%	0.145	BMI≤23	98.2±2.4%	0.317	≥T12	97.7±3.9%	0.052
	Female	98.5±2.1%		Age>70	98.8±1.1%		BMI>23	98.7±1.0%		<T12	98.8±1.2%	
IO	Male	95.6±3.4%	<u>0.013</u>	Age≤70	96.2±2.9%	0.250	BMI≤23	97.4±2.0%	0.076	≥T12	97.1±1.9%	0.146
	Female	97.5±1.6%		Age>70	97.1±2.5%		BMI>23	96.1±3.1%		<T12	96.2±3.7%	
CNS	Male	97.4±2.3%	0.456	Age≤70	96.8±2.8%	<u>0.019</u>	BMI≤23	97.4±2.5%	0.368	≥T12	97.8±2.4%	<u><0.001</u>
	Female	97.9±2.3%		Age>70	98.3±1.6%		BMI>23	98.0±2.0%		<T12	95.0±3.7%	
All masks	Male	97.4±1.8%	0.244	Age≤70	97.3±1.6%	0.099	BMI≤23	97.7±1.7%	0.997	≥T12	96.5±3.8%	0.230
	Female	97.9±1.3%		Age>70	98.0±1.4%		BMI>23	97.7±1.5%		<T12	97.2±2.0%	

Abbreviation: AVF, abdominal visceral fat; SF, subcutaneous fat; IO, internal organs and vessels; CNS, central nervous system; BMI=body mass index.

Statistical comparisons were performed in the pooled data of the internal dataset, the first, and the second external datasets.

P-values underlined in italic type indicate statistical significance.

†The AVF mask at the level of T12 or above only contained small areas of upper end of AVF where a small difference could make a significant decrease of dice score.

Table S3. Baseline characteristics and 3D U-Net-derived body composition parameters according to sarcopenia or dysmobility syndrome in the clinical validation cohort

	Sarcopenia			Dysmobility syndrome		
	Present (n=85)	Absent (n=437)	P-value	Present (n=149)	Absent (n=373)	P-value
Age, year	78 ± 5	75 ± 4	<0.001	77 ± 5	74 ± 4	<.001
Women, n (%)	52 (61)	300 (69)	0.179	111 (75)	241 (65)	.030
BMI, kg/m²	21 ± 3	25 ± 3	<0.001	24 ± 4	24 ± 3	.723
Falls within preceding year, n (%)	26 (31)	108 (25)	0.257	69 (46)	65 (17)	<.001
Low handgrip strength, n (%)	58 (68)	84 (19)	<0.001	98 (66)	44 (12)	<.001
Low lean mass, n (%)	85 (100)	38 (9)	<0.001	74 (50)	49 (13)	<.001
High fat mass, n (%)	10 (12)	100 (23)	0.021	61 (41)	49 (13)	<.001
Low bone mass, n (%)	62 (73)	278 (64)	0.099	131 (88)	209 (56)	<.001
Low TUG performance, n (%)	26 (31)	81 (19)	0.012	78 (52)	29 (8)	<.001
Low CRT performance, n (%)	60 (73)	133 (32)	<0.001	88 (65)	105 (29)	<.001
Torso SMI, cm³/m²	3348 ± 514	3826 ± 592	<0.001	3556 ± 570	3824 ± 603	<.001
Average SMI in the abdominal waist, cm²/m²	38 ± 5	43 ± 6	<0.001	40 ± 6	43 ± 6	.006
Torso BFI, cm³/kg	197 ± 70	233 ± 56	<0.001	237 ± 71	223 ± 55	.013
Average BFI in the abdominal waist, cm²/kg	4.0 ± 1.6	4.8 ± 1.2	<0.001	4.8 ± 1.6	4.6 ± 1.2	.083

Abbreviation: SMI, skeletal muscle index; BFI, body fat index; TUG, timed up and go test; CRT, chair rise test.

Figure legends

Figure 1. Study diagram for model development and validation

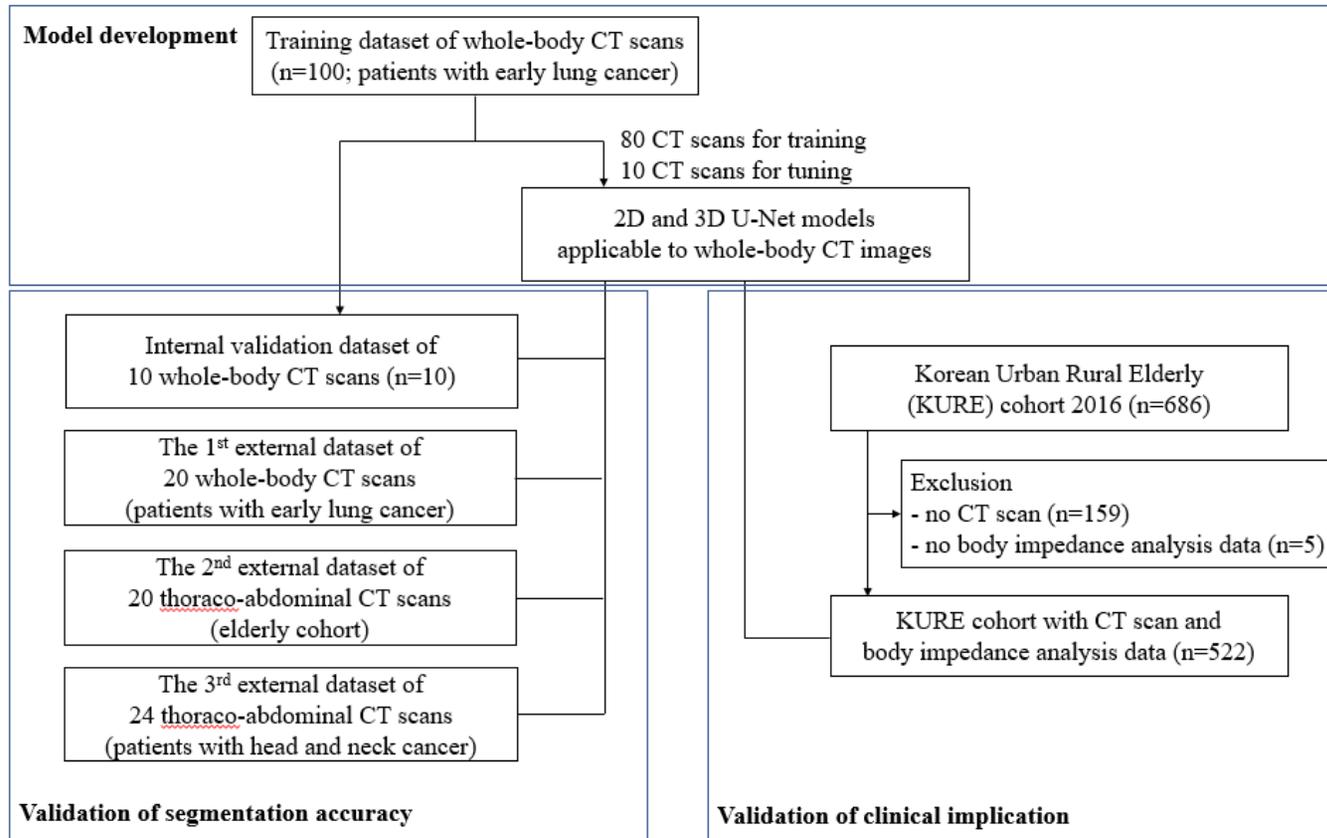
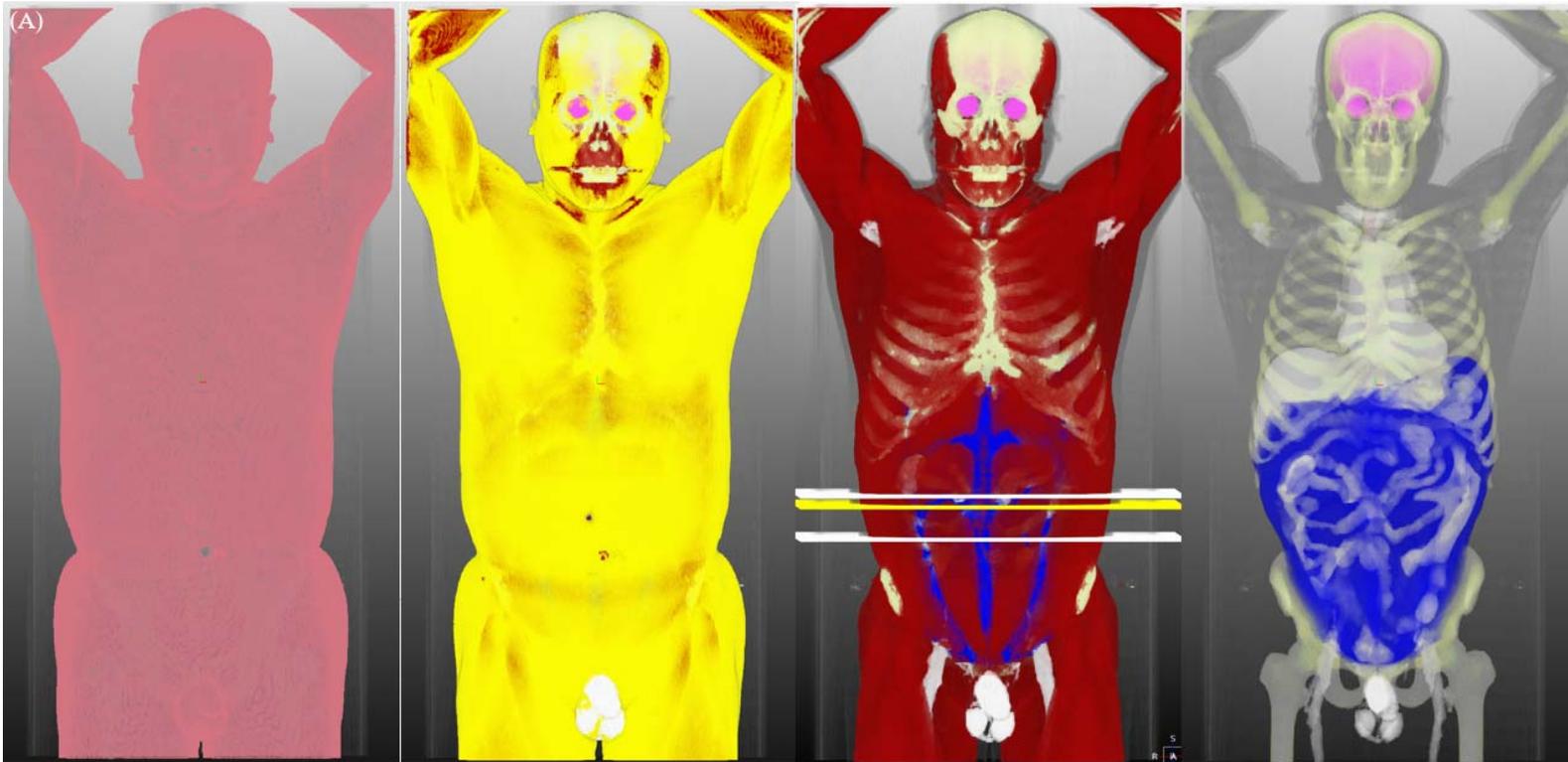
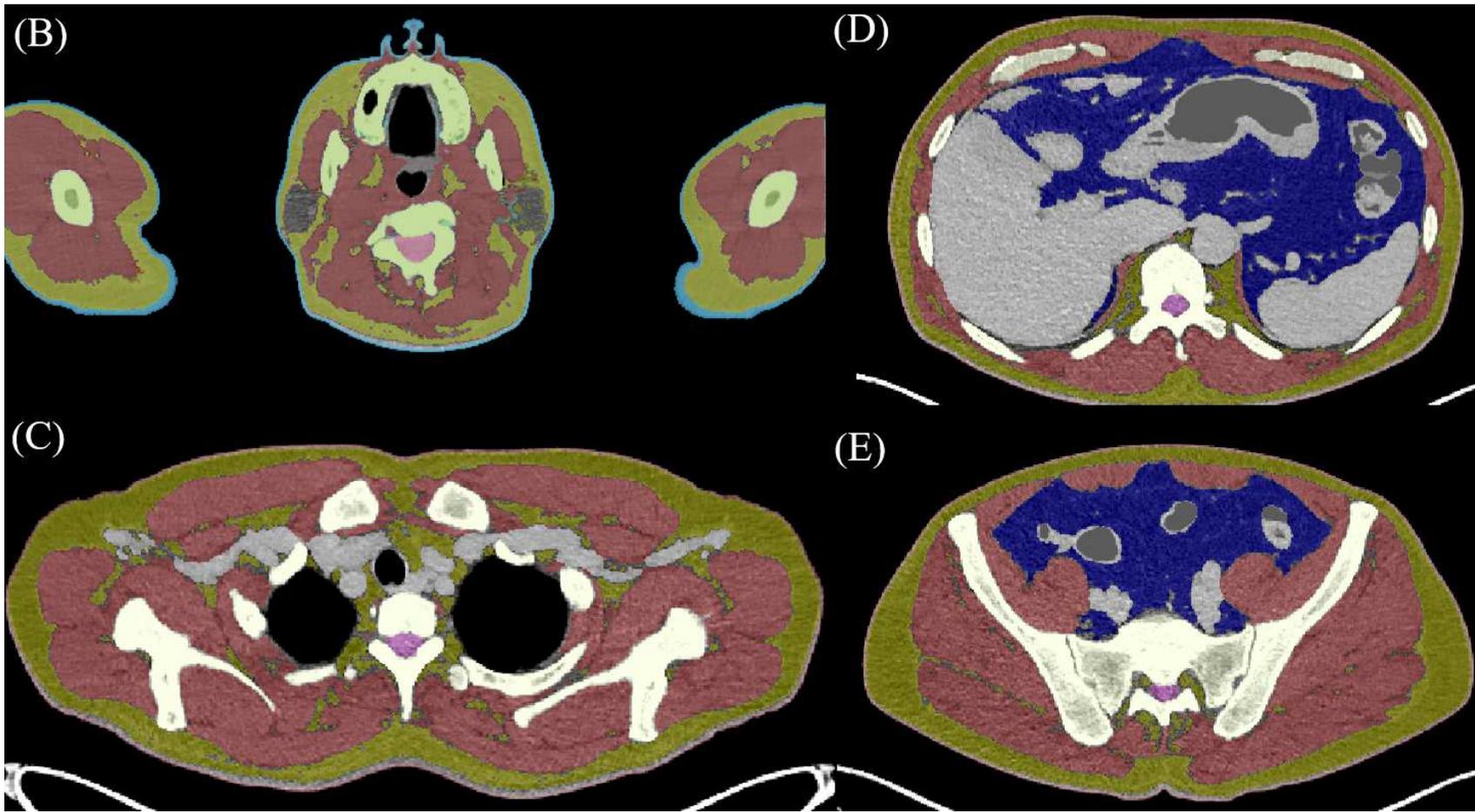


Figure 2. Representative 3D reformatted (A) and cross-sectional (B-E) CT images of 3D U-Net-derived masks in a 67-year-old male patient who underwent surgical resection of early lung cancer (body mass index, 27.0 kg/m²). (A) The four images sequentially show skin, subcutaneous fat, muscle, and abdominal visceral fat, along with bone, internal organs and vessels, and the central nervous system. Overlaid lines indicate the waist (white lines), and the L3 level (yellow line).





(B-E) The images show cross-sectional areas of head and upper arm (B), thoracic inlet (C), upper abdomen (D), and pelvis (E). Note that the

parotid glands are excluded from any masks as intended (A), and the borders of axillary vessels (B), outer margins of liver and spleen (C), and common iliac vessels (D) are well separated from adjacent muscles. The average Dice similarity coefficient for the seven classes was 98.2%, and the Dice similarity coefficients for each class were as follows: skin (pink), 95.7%; subcutaneous fat (yellow), 98.2%; skeletal muscle (brown), 98.4%; abdominal visceral fat (blue), 98.6%; bone (light beige), 98.8%; internal organs and vessels (light gray), 97.6%; and central nervous system (light pink), 99.1%.

Figure 3. Estimated cutoffs of computed tomographic average skeletal muscle index and body fat index in the waist. Cutoffs of 40 cm^2/m^2 and 35 cm^2/m^2 in men and women, respectively, in the average skeletal muscle index of the waist corresponded to bioelectrical impedance analysis-based clinical thresholds for low appendicular lean mass (7.0 kg/m^2 in men and 5.7 kg/m^2 in women) in the linear regression model (Figure 2). Cutoffs of 4.7 cm^2/kg and 5.8 cm^2/kg in men and women, respectively, in the average body fat index of the waist corresponded to bioelectrical impedance analysis-based high fat mass thresholds (30% in men and 40% in women).

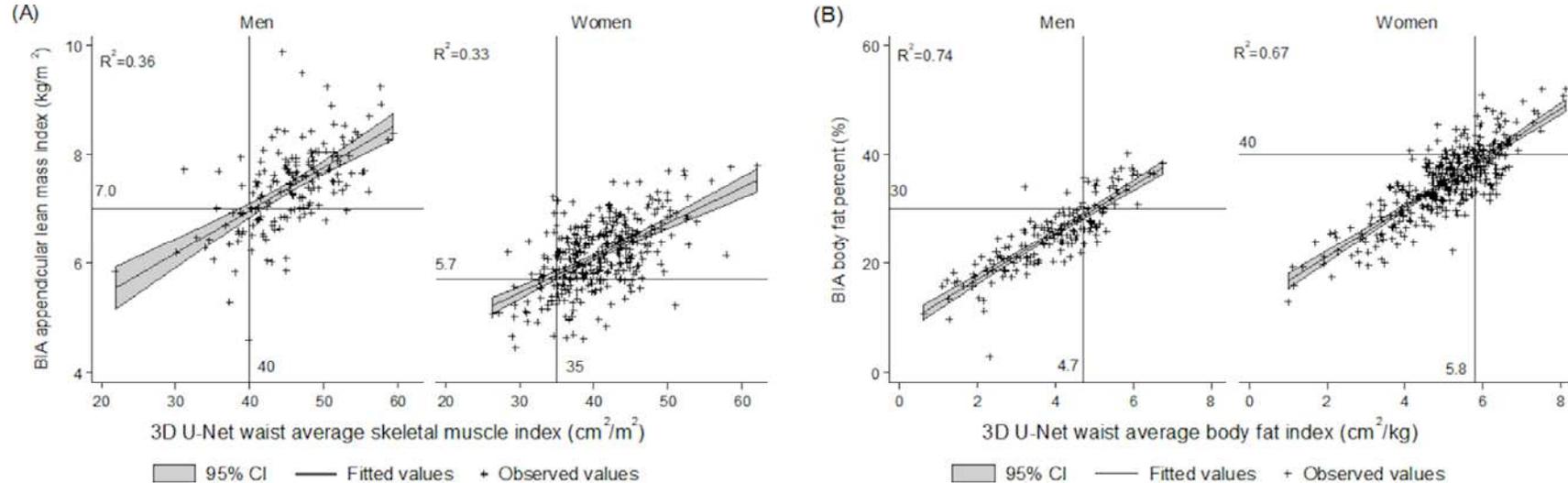
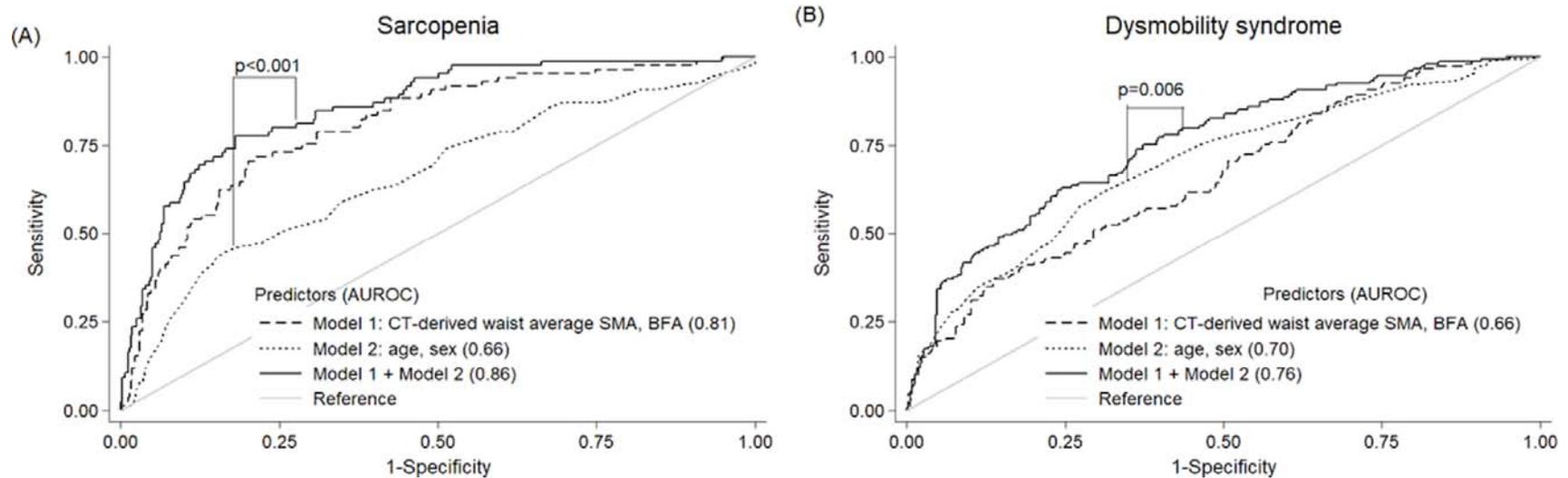


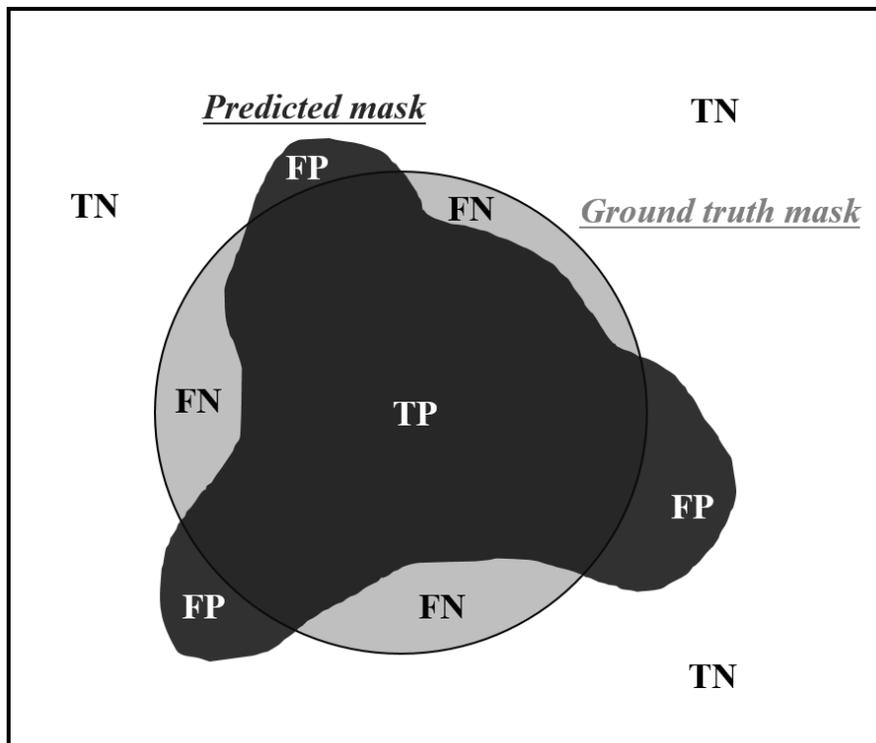
Figure 4. Improved diagnostic performance for sarcopenia and dysmobility syndrome achieved by adding average skeletal muscle area (SMA) and body fat area (BFA) in the waist to age and sex. The average SMA and BFA in the waist showed a modest to good discriminatory performance for sarcopenia (AUROC 0.81, 95% CI 0.75 to 0.85) or dysmobility syndrome (AUROC 0.66, 95% CI 0.60 to 0.70). Adding the average SMA and BFA to age- and sex-based models significantly improved the diagnostic performance for sarcopenia (AUROC 0.66 versus 0.81; $p < 0.001$) and dysmobility syndrome (AUROC, 0.70 versus 0.76; $P = 0.006$).



Supplemental figure legends

Figure S1. Schematic diagram for assessing segmentation accuracy.

A schematic diagram shows the ground truth mask (light gray, circle) and deep learning-predicted mask (dark gray, irregular shape) of the region of interest. A spatial relationship between ground truth and the predicted masks can be categorized into one of four classes: true positivity (TP), positive pixels of ground truth mask that are correctly identified by the predicted mask; true negativity (TN), negative pixels of ground truth mask that are correctly identified by the predicted mask; the incorrectly identified negative pixels of the predicted mask (FN); the incorrectly identified positive pixels of the predicted mask (FP). Equations for calculating dice similarity coefficient, sensitivity, and positive predictive value are as follows: dice similarity coefficient, $2TP/(2TP+FP+FN)$; sensitivity, $TP/(TP+FN)$; positive predictive value, $TP/(TP+FP)$.



Figures

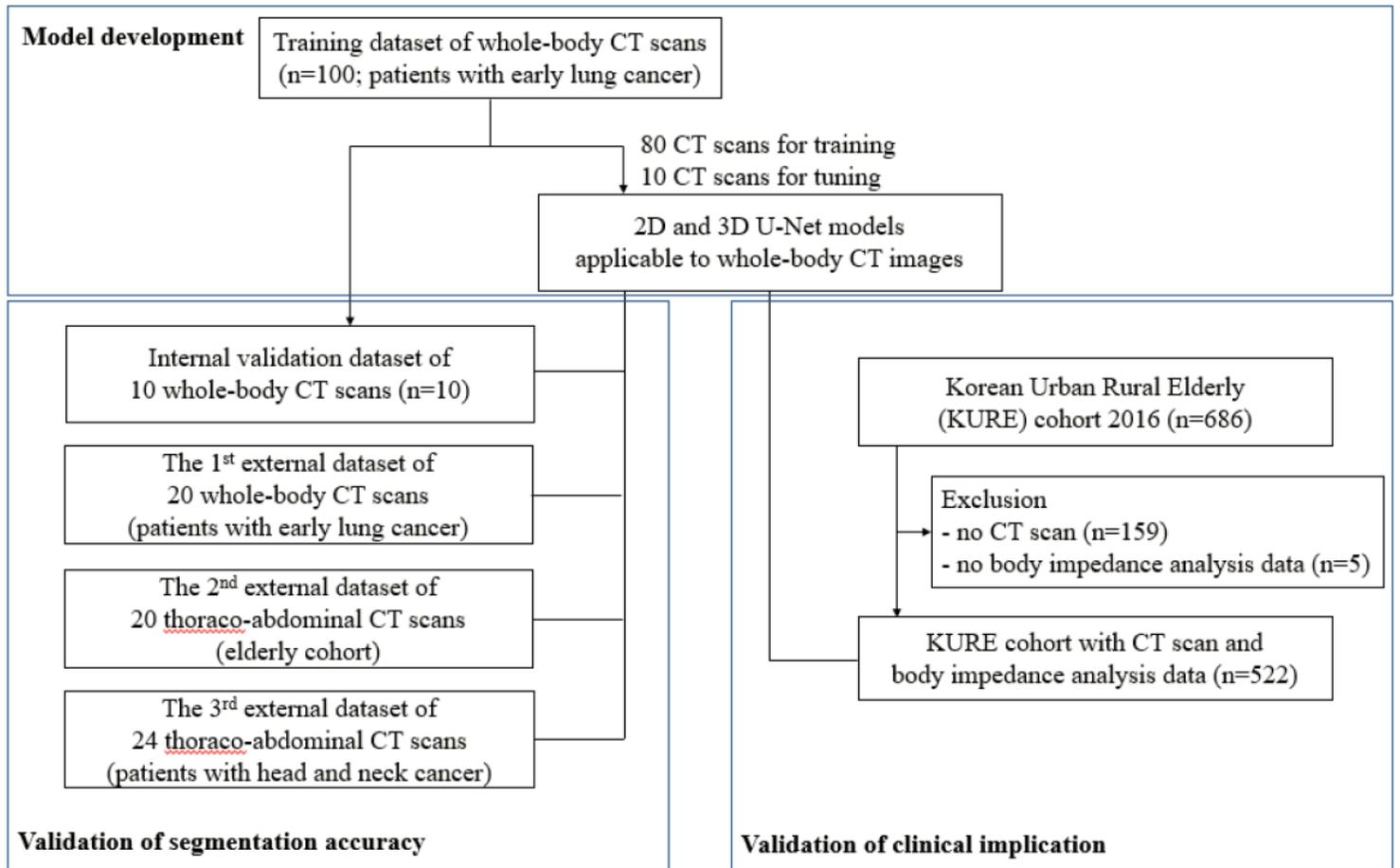


Figure 1

Study diagram for model development and validation

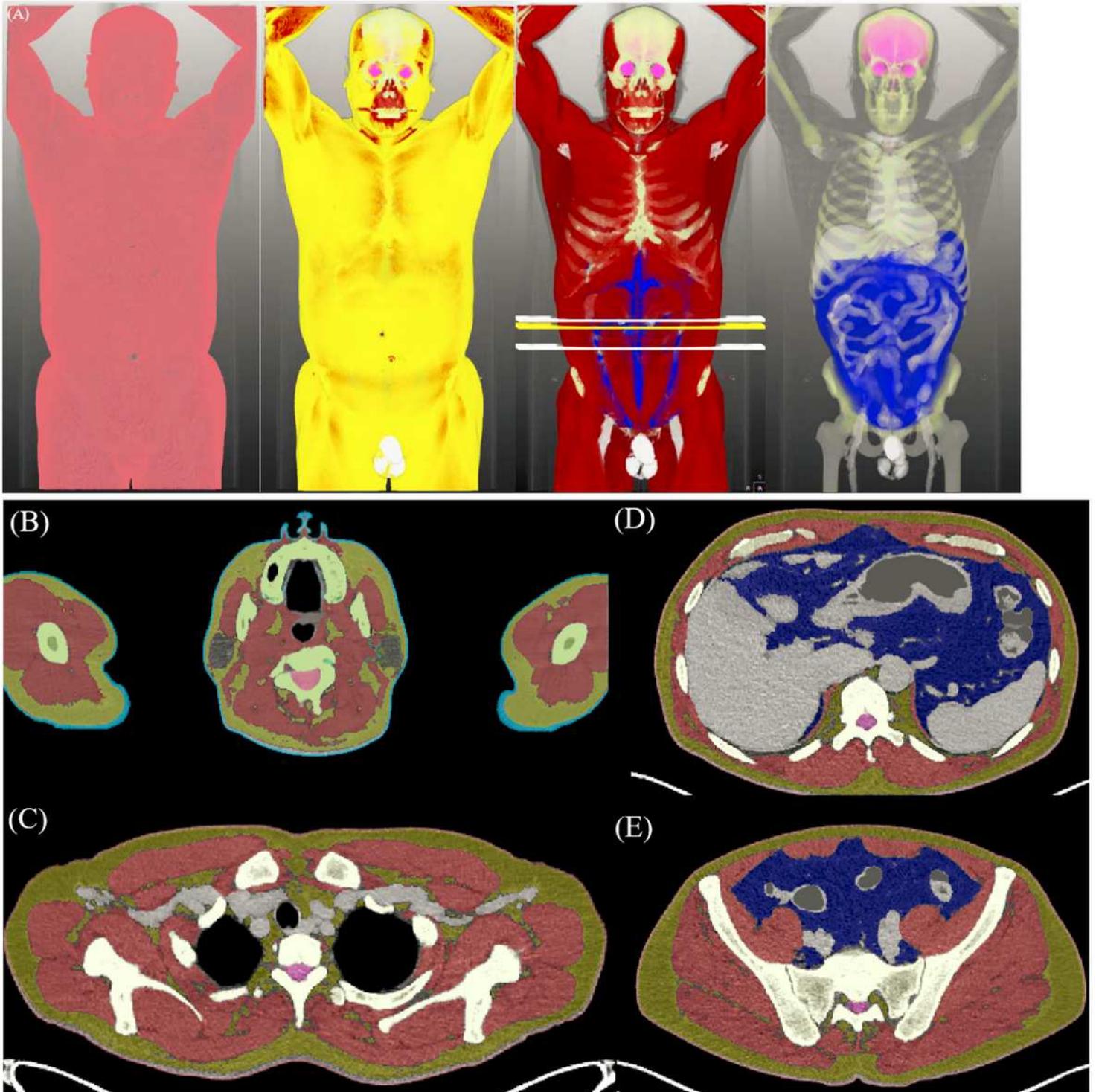


Figure 2

Representative 3D reformatted (A) and cross-sectional (B-E) CT images of 3D U-Net-derived masks in a 67-year-old male patient who underwent surgical resection of early lung cancer (body mass index, 27.0 kg/m²). (A) The four images sequentially show skin, subcutaneous fat, muscle, and abdominal visceral fat, along with bone, internal organs and vessels, and the central nervous system. Overlaid lines indicate the waist (white lines), and the L3 level (yellow line). (B-E) The images show cross-sectional areas of head and upper arm (B), thoracic inlet (C), upper abdomen (D), and pelvis (E). Note that the parotid glands

are excluded from any masks as intended (A), and the borders of axillary vessels (B), outer margins of liver and spleen (C), and common iliac vessels (D) are well separated from adjacent muscles. The average Dice similarity coefficient for the seven classes was 98.2%, and the Dice similarity coefficients for each class were as follows: skin (pink), 95.7%; subcutaneous fat (yellow), 98.2%; skeletal muscle (brown), 98.4%; abdominal visceral fat (blue), 98.6%; bone (light beige), 98.8%; internal organs and vessels (light gray), 97.6%; and central nervous system (light pink), 99.1%.

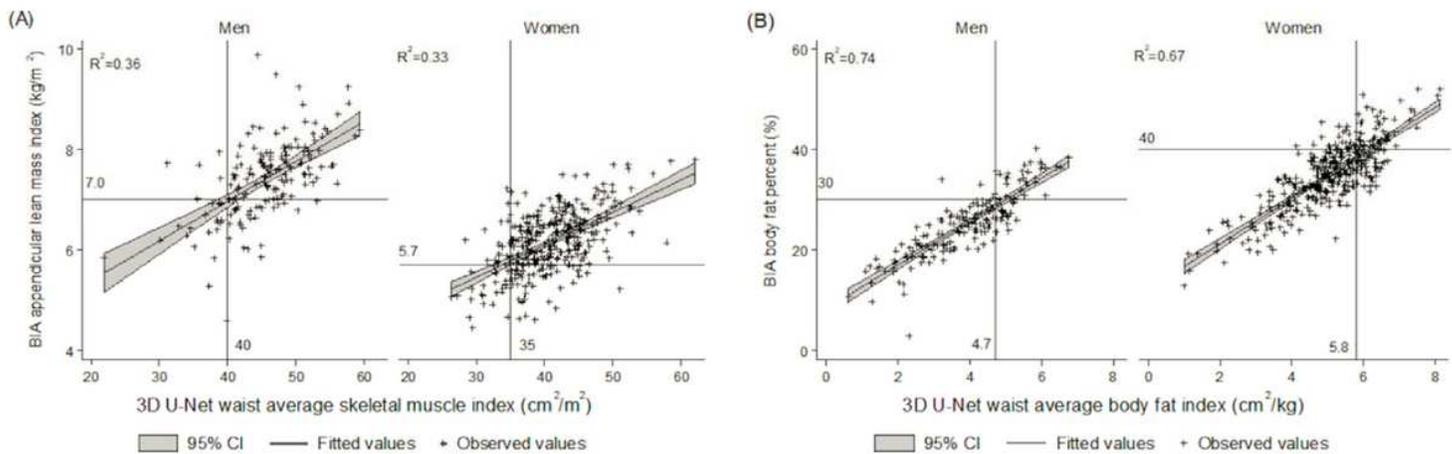


Figure 3

Estimated cutoffs of computed tomographic average skeletal muscle index and body fat index in the waist. Cutoffs of 40 cm²/m² and 35 cm²/m² in men and women, respectively, in the average skeletal muscle index of the waist corresponded to bioelectrical impedance analysis-based clinical thresholds for low appendicular lean mass (7.0 kg/m² in men and 5.7 kg/m² in women) in the linear regression model (Figure 2). Cutoffs of 4.7 cm²/kg and 5.8 cm²/kg in men and women, respectively, in the average body fat index of the waist corresponded to bioelectrical impedance analysis-based high fat mass thresholds (30% in men and 40% in women).

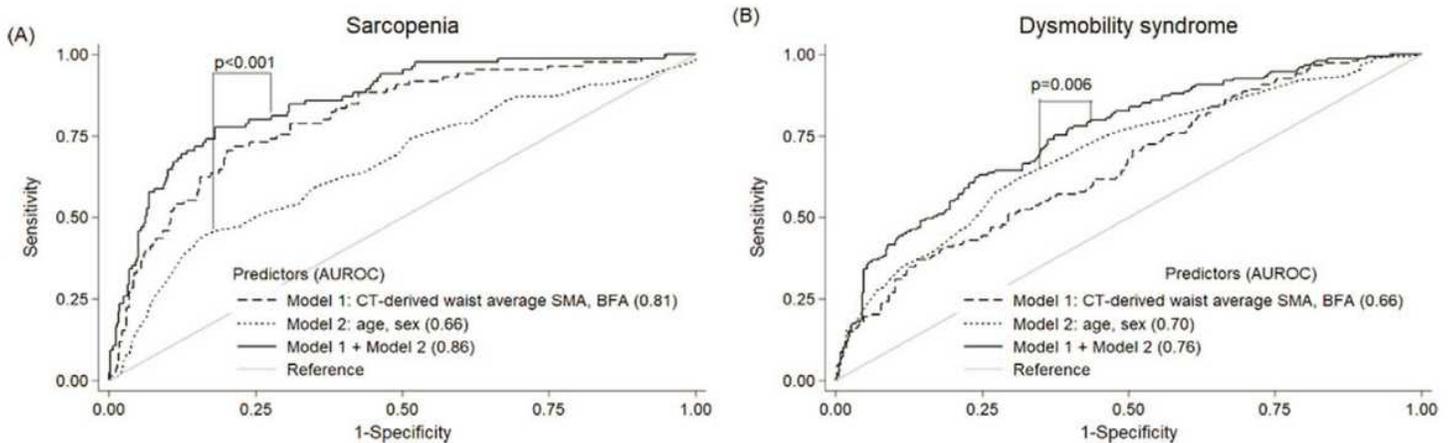


Figure 4

Improved diagnostic performance for sarcopenia and dysmobility syndrome achieved by adding average skeletal muscle area (SMA) and body fat area (BFA) in the waist to age and sex. The average SMA and BFA in the waist showed a modest to good discriminatory performance for sarcopenia (AUROC 0.81, 95% CI 0.75 to 0.85) or dysmobility syndrome (AUROC 0.66, 95% CI 0.60 to 0.70). Adding the average SMA and BFA to age- and sex-based models significantly improved the diagnostic performance for sarcopenia (AUROC 0.66 versus 0.81; $p < 0.001$) and dysmobility syndrome (AUROC, 0.70 versus 0.76; $P = 0.006$).