

Automated Muscle and Fat Segmentation in Computed Tomography for Comprehensive Body Composition Analysis

Yaqian Chen ¹, Hanxue Gu ¹, Yuwen Chen ¹, Jicheng Yang ¹, Haoyu Dong ¹, Joseph Y. Cao ², Adrian Camarena ³, Christopher Mantyh ³, Roy Colglazier ², Maciej A. Mazurowski ^{1,2,4,5},

1 Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708

2 Department of Radiology, Duke University, Durham, NC 27708

3 Department of Surgery Duke University School of Medicine, Durham, NC 27708

4 Department of Biostatistics & Bioinformatics, Duke University, Durham, NC 27708

5 Department of Computer Science, Duke University, Durham, NC 27708

Abstract

Body composition assessment using CT images can potentially be used for a number of clinical applications, including the prognostication of cardiovascular outcomes, evaluation of metabolic health, monitoring of disease progression, assessment of nutritional status, prediction of treatment response in oncology, and risk stratification for surgical and critical care outcomes. While multiple groups have developed in-house segmentation tools for this analysis, there are very limited publicly available tools that could be consistently used across different applications. To mitigate this gap, we present a publicly accessible, end-to-end segmentation and feature calculation model specifically for CT body composition analysis. Our model performs segmentation of skeletal muscle, subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT) across the chest, abdomen, and pelvis area in axial CT images. It also provides various body composition metrics, including muscle density, visceral-to-subcutaneous fat (VAT/SAT) ratio, muscle area/volume, and skeletal muscle index (SMI), supporting both 2D and 3D assessments. The model is shared for public use. To evaluate the model, the segmentation was applied to both internal and external datasets, with body composition metrics analyzed across different age, sex, and race groups. The model achieved high dice coefficients on both internal and external datasets, exceeding 89% for skeletal muscle, SAT, and VAT segmentation. The model outperforms the benchmark by 2.40% on skeletal muscle and 10.26% on SAT compared to the manual annotations given by the publicly available dataset. Body composition metrics show mean relative absolute errors (MRAEs) under 10% for all measures. Furthermore, the model provided muscular fat segmentation with a Dice coefficient of 56.27%, which can be utilized for additional analyses as needed. Our code is available at <https://github.com/mazurowski-lab/CT-Muscle-and-Fat-Segmentation.git>.

Keywords

Deep learning, Segmentation, Muscles, Subcutaneous Fat, Visceral Fat, Body Composition

Article informations

Corresponding author: author@institute.tld

©YYYY Name1 and Name2. License: CC-BY 4.0

1. Introduction

Correlating body composition metrics based on computed tomography (CT) images with disease and clinical variables, such as cancer (Rutten et al., 2016; Kumar et al., 2016; Deluche et al., 2018; Yoshikawa et al., 2020; Iwase et al., 2016; Boer et al., 2020), cachexia (Ali and Garcia, 2014; Fearon and Preston, 1990; Al-Sawaf et al., 2023; Baracos et al., 2010) and frailty (Falsarella et al., 2015; Reinders et al., 2017; Villareal et al., 2004), is becoming a widely adopted approach to leverage medical imaging data

for real-world clinical applications (Tolonen et al., 2021). By measuring body composition, such as quantity and location of fat as well as quantity and quality of muscle, clinicians are able to gain valuable insights into a patient's physiological status (Prado and Heymsfield, 2014). This information enables them to assess disease progression (Baracos and Kazemi-Bajestani, 2013), evaluate treatment efficacy (Bates and Pickhardt, 2022), and predict clinical outcomes (Weston et al., 2019).

Several key metrics are frequently utilized in body composition analysis, including muscle density, the visceral-to-

subcutaneous fat (VAT/SAT) ratio, muscle area or volume, and the skeletal muscle index (SMI). Most studies in this field measure these metrics from a single slice, most commonly at the third lumbar vertebra (L3) (Arayne et al., 2023), while others employ volumetric analysis (Connelly et al., 2013). However, regardless of the approach, extracting these metrics relies on effective segmentation models to accurately identify and quantify various tissues within the body. Traditional methods, such as pixel thresholding based on Hounsfield units (HU) (Wang and Torriani, 2020) and fuzzy c-means clustering (Wang and Torriani, 2020; Christ and Parvathi, 2011), often require manual adjustments and are time-intensive (Wang and Torriani, 2020). Furthermore, pixel thresholding algorithms cannot differentiate between visceral fat, subcutaneous fat, and intramuscular fat—an essential distinction when measuring the VAT/SAT ratio. Deep learning segmentation is a response to these limitations, and some groups have developed in-house segmentation models customized for their private datasets (Fu et al., 2020; Lee et al., 2017; Weston et al., 2019; Wang et al., 2017; Hemke et al., 2020; Koitka et al., 2021). These models typically lack public accessibility and are designed for specific tasks. Furthermore, we observed inconsistency in how muscular fat (both intra-muscular and inter-muscular fat) is utilized in research. While some studies include muscular fat as part of skeletal muscle measurements (Hou et al., 2024; Blanc-Durand et al., 2020; Weston et al., 2019), others classify it under VAT (Camus et al., 2014; Wirtz et al., 2021; Connelly et al., 2013), and a smaller subset considers it part of SAT (Ozturk et al., 2020).

In order to advance the research on the relationship of imaging-based body composition with disease and clinical variables, a robust, thoroughly validated, and publicly available tissue segmentation model and body composition variable calculation is needed. This model will allow different research groups to test the model with their data and correlate the unified body composition measurements with the clinical outcomes of their interest, building consistent scientific evidence of the importance of body composition in human health.

Toward this goal, we developed a segmentation model using nnU-Net (Isensee et al., 2021, 2024) which identifies the areas of skeletal muscle, SAT, VAT, and muscular fat. For the model training and evaluation, we collected 813 CT volumes of chest, abdomen, and pelvis for 483 patients from Duke University Health System. In both training and test datasets, we included volumes from different years (from 2016 to 2019), various scanners, and diverse patient demographics to ensure the model’s generalizability. Additionally, we incorporated the publicly available Sparsely Annotated Region and Organ Segmentation (SAROS) dataset (Koitka et al., 2023; Clark et al., 2013) for segmentation

evaluation, demonstrating the generalizability of our proposed model. Furthermore, we also analyze the relationships between the four body composition metrics (muscle density, VAT/SAT ratio, muscle area/volume, and SMI) with respect to three key demographic variables: age, sex, and race (shown in Section 5). Notably, to facilitate wide use of the model, we have made it publicly available.

The main contributions of this work are summarized as follows:

- To the best of our knowledge, this is the first publicly available deep-learning model designed to segment skeletal muscle, SAT, VAT, and muscular fat across the chest, abdomen, and pelvis on CT.
- We provide end-to-end standardized and publicly available measurements for four common body composition metrics, including muscle density, visceral-to-subcutaneous fat (VAT/SAT) ratio, muscle area/volume, and SMI on both L3 for 2D and T12 to L4 for 3D measurement.
- Our model outperforms TotalSegmentator (Wasserthal et al., 2023) and Enhanced segmentation (Hou et al., 2024) by 2.40% on skeletal muscle and 10.26% on SAT compared to the manual annotations given by publicly available dataset SAROS.
- We perform the statistical analysis to correlate four metrics (muscle density, VAT/SAT ratio, muscle area/volume, and SMI) with three patient demographic variables: age, sex, and race in both 2D and 3D settings.

2. Related Works

2.1 Body composition analysis using CT

Body composition plays a crucial role in influencing physical performance (Hernandez-Martinez et al., 2024; Falsarella et al., 2015), metabolic health (Trouwborst et al., 2024; Kakinami et al., 2022), and disease outcomes (Rutten et al., 2016). Imaging offers an objective, quantitative approach to its analysis through various techniques, including CT, magnetic resonance imaging (MRI), and ultrasonography (Hou et al., 2024; Tan et al., 2024; Sharifi et al., 2024; Xu et al., 2024). Among all the modalities, CT offers high spatial resolution, faster acquisition times, and superior contrast between tissues (Zhang et al., 2021), making it particularly suitable for assessing visceral and subcutaneous fat, skeletal muscle, and organ-specific fat deposits (Wathen et al., 2013).

During body composition calculation on CT, several key metrics are frequently utilized, including muscle density, the VAT/SAT ratio, muscle area/volume, and the SMI. Muscle density in CT provides insights into muscle

quality, which is linearly influenced by muscular fat content (Engelke et al., 2018). A reduction in muscle density is often associated with increased fat infiltration within the muscle, known as myosteatosis (Chang and Cheng, 2024), which compromises muscle function and structural integrity. This reduction serves as a critical indicator of sarcopenia, a condition characterized by the progressive loss of skeletal muscle mass and strength, as well as frailty and diminished physical performance, particularly in aging populations (Cawthon, 2015).

The VAT/SAT ratio, on the other hand, is a key metric for assessing metabolic risk (Kaess et al., 2012; Oh et al., 2017). While visceral adipose tissue (VAT) is strongly associated with metabolic disturbances and cardiovascular risk (Vasamsetti et al., 2023), its volume alone may reflect both overall fat mass and an individual's tendency to store fat viscerally (Kaess et al., 2012). In contrast, the VAT/SAT ratio offers a more precise assessment, as it accounts for the balance between visceral and subcutaneous fat, providing insight that is independent of total body fat percentage (Kaess et al., 2012).

Muscle area/volume and SMI are essential measurements of total muscle quantity and its proportionality to body size. These metrics provide critical information about an individual's muscle reserves, which are vital for mobility, metabolic function, and overall health status (Chen et al., 2023). Studies highlight them as significant markers of nutritional status (Risch et al., 2022), which are crucial for recovery from illness, mortality, and treatment-related complications, such as the length of hospital stays and the rate of readmissions (Schuetz et al., 2021; Kaegi-Braun et al., 2021; Guenter et al., 2021). Furthermore, they also serve as important factors in assessing metabolic health (Cruz-Jentoft et al., 2019; Prado and Heymsfield, 2014; Martin et al., 2013; Dodds et al., 2015), as lower muscle mass is associated with insulin resistance and impaired glucose metabolism.

The collection of these metrics pictures the clear body condition of patients, showcasing a comprehensive overview of their muscle composition, fat distribution, and overall physiological status.

2.2 Traditional methods for body composition segmentation

Most early studies on body composition analysis rely on semi-automated threshold-based segmentation using pre-defined Hounsfield unit (HU) ranges to differentiate lean muscle mass from adipose tissue (Lee et al., 2017; Ji et al., 2022). Despite its simplicity, threshold-based segmentation presents significant challenges due to the overlapping HU values between different tissue types, such as SAT and skin, as well as muscle and adjacent organs (Lee et al., 2017).

The method is also highly susceptible to image noise (Seghal and Kaushik, 2022; Diwakar et al., 2020), which can significantly compromise tissue classification accuracy, particularly in low-quality or artifact-prone scans. As a result, the method typically requires manual correction based on visual analysis by highly skilled radiologists and is impractical on large datasets due to the expense and time required.

To overcome these limitations, researchers have developed various advanced segmentation algorithms, including rule-based (Kamiya et al., 2009, 2011), clustering-based (Positano et al., 2009, 2004; Christ and Parvathi, 2011), and finite-element-method-based (Popuri et al., 2015) approaches. Kamiya et al. proposed a rule-based expert system for segmenting the psoas major and rectus abdominis muscles from CT images, approximating muscle shapes with simple quadratic functions (Kamiya et al., 2009, 2011). Positano et al. utilize a fuzzy c-mean algorithm to make unsupervised classification of image pixels on MRI (Positano et al., 2009, 2004). Karteek and the team developed a novel FEM deformable model for muscle and fat segmentation from CT (Popuri et al., 2015).

However, these methods primarily focus on extracting specific muscle groups from CT or MRI scans and are unable to differentiate between visceral fat, subcutaneous fat, and intramuscular fat—an essential distinction in many body composition analysis tasks (Staley et al., 2019; Torres et al., 2013; Iwase et al., 2016). A potential approach to address these challenges is the use of deep learning-based segmentation algorithms.

2.3 Deep learning-based models for body composition segmentation

Deep learning-based segmentation has been proven to be a reliable technique in various clinical applications (Gu et al., 2024; Dong et al., 2024; Wasserthal et al., 2023; Mazurowski et al., 2023). While networks offer high accuracy, reduce human labor, and provide greater generalizability compared to traditional segmentation algorithms, it is straightforward to apply deep learning-based segmentation algorithms for body composition analysis.

The majority of current deep learning-based segmentation models for body composition are still based on convolutional neural networks (CNNs) (Nowak et al., 2020). U-Net and its variants are among the most widely used architectures in this domain, providing precise segmentation of body composition components such as skeletal muscle, SAT, and VAT (Paris, 2020; Weston et al., 2019). However, these models are typically not publicly accessible and are often designed for specific tasks (Mai et al., 2023). While a few commercial models are available (Cespedes Feliciano et al., 2020; Mai et al., 2023; Lee et al., 2021), they are often associated with high costs and lim-

ited customization options. TotalSegmentator (Wasserthal et al., 2023), a recently published general CT segmentation model based on nnU-Net, also supports muscle and fat segmentation. However, studies have shown that its performance in segmenting muscle, SAT, and VAT can be further improved, and its non-commercial license restricts broader usage. Therefore, there remains a significant need for publicly accessible, transparent, and generalizable segmentation models for body composition analysis.

3. Methods

3.1 Datasets

In this study, we utilize two datasets: an internal dataset collected from Duke Hospital and the publicly available SAROS dataset (Koitka et al., 2023; Clark et al., 2013) that integrates four publicly available CT datasets from TCIA with sparse annotations. The internal dataset is exclusively used for model implementation to ensure flexible and permissive licensing requirements. Both datasets are utilized for segmentation evaluation. The datasets are diverse in institutions, years, and patient demographics, providing the general and reliable evaluation and analysis results in this study.

3.1.1 Dataset 1: Internal dataset

For this project, we collected 8948 CT volumes from Duke University Health System, spanning January 2016 to November 2019, including chest, abdomen, and pelvis exams. From the initial collection, we further identified 1927 volumes from 854 patients based on two criteria: (1) axial view exam was available (2) the volumes were original axial acquisitions, not derived from multiplanar reconstructions (MPR) or reformatted from other planes. These criteria were selected to align the model with real-world clinical scenarios. Among the identified studies, 483 patients were designated for segmentation model development, with 453 patients used for training and 30 patients used for testing, while the remaining 371 patients were allocated for further analysis of the relationship between body composition measurements and demographics. Noticeably, for 371 patients that were assigned for demographic analysis, only non-contrast-enhanced volumes were selected to ensure the analysis consistency.

To mitigate the potential bias in the testing and analysis process, only a single volume was randomly selected if patients had multiple eligible CT volumes. This approach ensured that the testing and analysis datasets provided an unbiased representation of the patient population.

Moreover, to best utilize our limited annotation resources and ensure data variability and model generalization, the training dataset was constructed by randomly

sampling slices from the volumes. The selected slices were annotated by four Duke students under the guidance of experienced radiologists. To ensure the model accuracy, all the slices in the testing dataset are modified and approved by the radiologists.

3.1.2 Dataset 2: SAROS dataset

The SAROS dataset (Koitka et al., 2023; Clark et al., 2013) is a comprehensive collection of CT imaging volumes available on TCIA, featuring sparse annotations for 13 body region labels and six body part labels. The 13 annotations for body regions include the abdominal cavity, thoracic cavity, bones, brain, breast implants, mediastinum, muscles, parotid glands, submandibular glands, thyroid glands, pericardium, spinal cord, and subcutaneous tissue. The six body parts are the left arm, right arm, left leg, right leg, head, and torso, comprising a total of 900 CT volumes from 882 unique patients.

Given the torso label overlap with chest, abdomen, and pelvis regions, which are the focus body regions for our study, we utilize the slices with torso labels for our model evaluation by comparing the model segmented results with annotated skeletal muscle and SAT annotations. Furthermore, to ensure the flexible use of our model, we only selected a subset of the SAROS dataset covered under a commercial license for evaluation, more details for collection selection are shown in Appendix Section A. As a result, in total, 650 CT volumes from 632 unique patients with CT slices are selected for segmentation model evaluation.

3.1.3 Patient demographics

The below table provides a demographic overview of patients from the five dataset collections, derived from two sources: the internal dataset (including internal training, internal testing, and demographic analysis collections, shown in Section 3.1.1) and the external dataset (SAROS collections, in Section 3.1.2). Notably, the training, testing, and demographic analysis collections from the internal dataset are entirely separate from one another, ensuring a clean environment for segmentation model development. Patients' ages and races are unknown for the SAROS dataset due to de-identification.

3.2 Segmentation algorithm

While the nnU-Net (Isensee et al., 2021, 2024) has been proven to be one of the most powerful segmentation models in many bodies' regions on medical imaging, in this work, we utilize the 2D nnU-Net with ResEnc presets on skeletal muscle, SAT, VAT, and muscular fat segmentation. Five-fold cross-validation is utilized to select the best-performance model based on the average Dice coefficient across all four

		Internal Training	Internal Testing	Demographic Analysis	SAROS Dataset
Number of patients		453	30	371	632
Number of studies		783	30	371	650
Number of slices		1863	636	183972	10038
Demographics	Age	60.1 (3 - 89)	58.5 (5 - 84)	58.3 (0.25-7)	-
	Sex	Female	51.88% (235)	53.33% (16)	52.02% (193)
		Male	48.12% (218)	46.67% (14)	47.98% (178)
	Race	White	67.11% (304)	63.33% (19)	68.19% (253)
		Black/ African American	25.83% (117)	30.00% (9)	23.99% (89)
		Asian	1.99% (9)	0% (0)	1.62% (6)
		American Indian	0.66% (3)	0% (0)	1.08% (4)
		Other	4.42% (20)	3.33% (1)	5.12% (19)

Table 1: Patient demographics for four collections: Table presents demographic details for four collections, including Internal Training, Internal Testing, Demographic Analysis, and SAROS Dataset Testing. For sex and race, the absolute number of patients is shown in parentheses alongside the percentages. For age, the mean values for ages are provided in years along with the minimum and maximum age in parentheses. Notably, the youngest patient, recorded as 3 months old, is consistently represented as 0.25 years.

labels. In the subsections below, we introduce the overall architecture of the 2D nnU-Net and the Dice coefficient measurement. In this work, the algorithm was trained and evaluated on an NVIDIA RTX 3090 GPU, ensuring efficient computation and high performance.

3.2.1 nnU-Net

nnU-Net (Isensee et al., 2021) is a highly adaptable semantic segmentation method designed to automatically configure an optimized U-Net-based pipeline for any given dataset. Recent updates to the nnU-Net methodology have introduced enhancements to the U-Net baseline, emphasizing the importance of using advanced CNN architectures like ResNet and ConvNeXt variants, leveraging the robust nnU-Net framework, and employing model scaling for improved performance. By analyzing the nnU-Net’s performance on multiple medical imaging datasets, the nnU-Net ResEnc XL has been shown to surpass the vanilla nnU-Net by an average of 0.93% (Isensee et al., 2024). Therefore, in this work, we follow their findings and adopt the newly published nnU-Net ResEnc XL for our model development.

3.2.2 Dice coefficient

Dice coefficient is commonly used for image segmentation tasks to evaluate segmentation accuracy by measuring the overlap between predicted and ground truth regions. It ranges from 0 to 1, where a value of 1 indicates perfect overlap and 0 signifies no overlap. The mathematic formula for the Dice coefficient is Equation (1)

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

where A represents the model predicted mask and the B represents the set of pixels in the ground truth.

3.2.3 Mean relative absolute error

MRAE measures the absolute mean for relative errors across all data points (shown in Equation (2), which is normally applied to measure the relative error between prediction and ground truth. Lower MRAE values indicate better model performance, reflecting smaller deviations between predicted and actual values.

$$\text{MRAE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - B_i}{A_i} \right| \quad (2)$$

where A_i represents the ground truth values, B_i represents the predicted values, and n is the total number of data points.

3.3 Body composition metrics

Our model is capable of measuring four commonly used body composition metrics: muscle density, VAT/SAT ratio, muscle area/volume, and SMI in both 2D and 3D settings. By analyzing previous studies on body composition analysis, we selected the third lumbar vertebral level (L3) for 2D body composition measurements and the region spanning the twelfth thoracic vertebral level (T12) to the fourth lumbar vertebral level (L4) for 3D measurements. L3 is considered the most commonly used standard for body composition assessment in multiple clinical applications, including rectal cancer assessment (Han et al., 2020; Arayne et al., 2023), sarcopenia evaluation (Amini et al., 2019; Pickhardt et al., 2020b,a), and obesity research (Liu et al., 2023; Malietzis et al., 2015). For the 3D body composition measurement, T12 is selected as the beginning of the

3D measurement region following the approach of previous studies (Demerath et al., 2007; Tong et al., 2014). This focus is particularly relevant for assessing visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT); however, recent studies related to sarcopenia and rectal cancer also pay increasing attention to T12 (Fernández-Jiménez et al., 2024; Arayne et al., 2023; Soh et al., 2024). L4 is selected as the ending point since 92.62% of our internal abdominal CT volumes include L4, while only 78.07% include L5.

TotalSegmentator (Wasserthal et al., 2023) is utilized for automatically extracting, T12, L3, and L4. We select the slice with the largest L3 label among all slices with L3 mask for our 2D body composition measurement. For 3D measurement, we extract the portion between the first slice with T12 label and the last slice with L4. In the subsequent sections, we detail the calculations for muscle density, VAT/SAT ratio, muscle area/volume, and SMI.

Muscle density measures the average Hounsfield Unit (HU) values within the segmented skeletal muscle area (SMA) with higher values indicating leaner muscle and lower values (typically from -29 to 29 HU (Salam et al., 2023)) suggesting fat infiltration. Muscle density is the crucial biomarker for muscle quality (Looijaard et al., 2016; Cleary et al., 2015; Wang et al., 2021) and is frequently associated with evaluations of sarcopenia and myosteatosis (Cawthon, 2015; Sergi et al., 2016; Tagliafico et al., 2022).

VAT/SAT ratio is more commonly related to obesity-related health risks, such as diabetes, cardiovascular disease, and metabolic syndrome (Piché et al., 2018; Frank et al., 2019; Goossens, 2017; Ladeiras-Lopes et al., 2017; Tanaka et al., 2021). A higher ratio indicates a predominance of visceral adipose tissue (VAT) over subcutaneous adipose tissue (SAT), reflecting an unfavorable fat distribution pattern (Ladeiras-Lopes et al., 2017). Visceral fat is metabolically active and associated with chronic inflammation, insulin resistance, and dyslipidemia, which contribute to the development and progression of these conditions (Hardy et al., 2012; Chait and Den Hartigh, 2020; Bansal et al., 2023).

Muscle area/volume assesses the total skeletal muscle within the region of interest (ROI). Specifically, for 2D measurement, this metric, also referred to as skeletal muscle area (SMA), is calculated by multiplying the number of pixels within the segmented skeletal muscle mask by the area (m^2) represented by each pixel. For 3D measurement, the segmented skeletal muscle area/volume is determined by multiplying both the pixel size and the slice thickness (m^3). The SMA is one of the standard metrics for muscle quantity evaluation (Goodpaster et al., 2000; Sinelnikov et al., 2016; Vella et al., 2020) and has been demonstrated to be highly correlated with patients' post-operative recovery and survival rates (Antoniou et al., 2019; Bradley et al.,

2022; Antoniou et al., 2019). 3D muscle area/volume provides better representation of the entire muscle (Momose et al., 2017).

SMI is another commonly used metrics for muscle quantity measurement. This metrics normalizes the muscle cross-sectional area (CSA) by dividing it by the individual's height squared (m^2).

4. Segmentation results

This section presents the segmentation performance of our model on three selected labels: skeletal muscle, SAT, and VAT, as well as four key body composition metrics: muscle density, muscle area/volume, SMI, VAT/SAT ratio. The evaluation is provided both qualitatively, through visual comparisons, and quantitatively, using the Dice coefficient and MRAE to measure the overlap between the manually annotated ground truth and the model's segmentation. The quantitative analysis highlights the performance of our model on both our internal dataset and the publicly available SAROS dataset (Koitka et al., 2023; Clark et al., 2013), benchmarking it against TotalSegmentator (Wasserthal et al., 2023) and the internal tool (Hou et al., 2024).

4.1 Qualitative evaluation

Figure 1 presents the L3 segmentation results and their corresponding body composition metrics (muscle density, VAT/SAT ratio, muscle area/volume, and SMI) for selected patients. The samples are categorized based on their body composition metric values into five groups: Low, Moderately Low, Moderate, Moderately High, and High, with cut-off points set at the 20th, 40th, 60th, and 80th percentiles of the entire population. For each body composition metric, one sample is randomly selected from each category for visualization. Each column in Figure 1 corresponds to a specific category, with patients arranged from low to high values across the columns, ensuring a consistent representation of the metric's progression. Each row, in turn, highlights a specific body composition metric. The first row illustrates muscle density, the second depicts the VAT/SAT ratio, the third represents muscle area/volume, and the fourth corresponds to the SMI. Notably, as shown in figure 1 result, there's no simple correlation between the four body composition metrics. For example, a patient with the highest muscle density in the first row does not exhibit the highest muscle area/volume. More precise relationship analysis based on Pearson Correlation coefficient for each body composition metrics pair is shown in Appendix C.

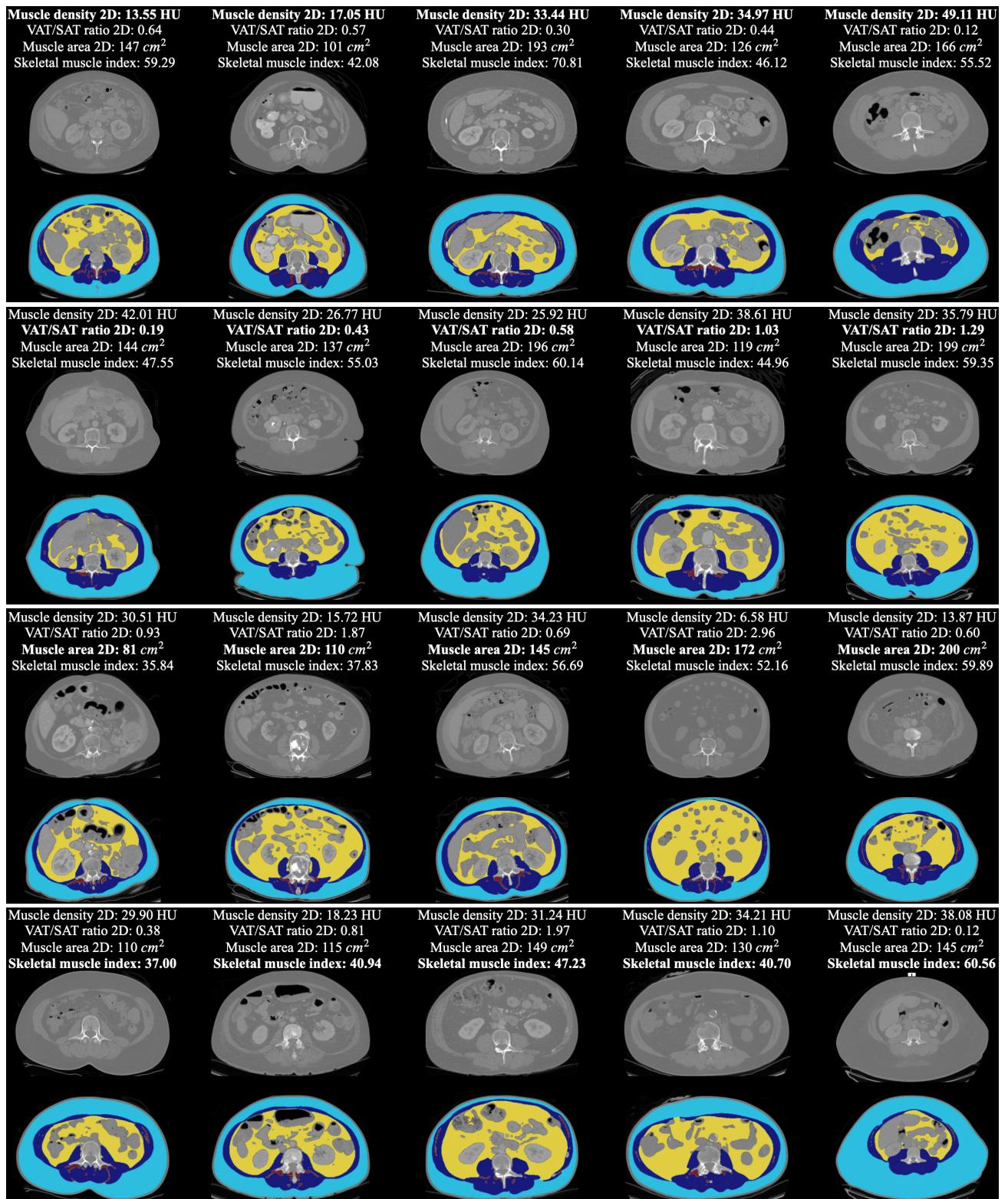


Figure 1: Qualitative evaluation of our segmentation model: Figure shows segmentation results of the abdominal L3 slice. Each row represents a specific body composition metric (in bold), with five patients arranged from left to right in categories: Low, Moderately Low, Moderate, Moderately High, and High. For example, in the first row (muscle density), the first patient has low muscle density, and the fifth has high muscle density. The second, third, and fourth rows show the VAT/SAT ratio, muscle area/volume, and SMI, respectively, following the same left-to-right order. In the segmentation, *dark blue* shows skeletal muscle, *light blue* SAT, *yellow* VAT, and *maroon* muscular fat.

In figure 2, three volumes are randomly selected from the demographic analysis dataset to demonstrate the algorithm’s performance for 3D body composition measurement. For each volume, four slices corresponding to T12, L1, L2, and L4 are displayed on the left, along with their automatically generated segmentation masks. L3 segmentation is not included in this figure, as it is fully demonstrated in Figure 1. Additionally, the stacked slices from T12 to L4 are visualized in the sagittal view, shown on the right side of the figure.

4.2 Quantitative evaluation

Quantitative segmentation model evaluation is demonstrated with three experiments. Subsection 4.2.1 demonstrate the model segmentation performance on our internal dataset and public available SAROS dataset based on the Dice coefficient between model segmentation and manual annotation for skeletal muscle, SAT, VAT, and muscular fat. The performance on multiple body location has been analyzed, specifically we evaluate the segmentation result on both L3, T12-L4, and every slice among the selected volume (chest, abdomen, and pelvis). 4.2.2 compares our segmentation performance with two of external segmentation model (Wasserthal et al., 2023; Hou et al., 2024) on the public available SAROS dataset (Koitka et al., 2023; Clark et al., 2013). The evaluation focus on skeletal muscle and SAT performance on the location from to align with the evaluation proposed by (Clark et al., 2013). Section 4.2.3 evaluates the segmentation accuracy by analyzing the mean error between automatically measured body composition metrics and those derived from manual annotations. Specifically, we present the mean error for metrics such as muscle density, VAT/SAT ratio, muscle area/volume, and SMI in both 2D and 3D settings. For more flexible use of muscular fat (including intra-muscular and inter-muscular fat), we also provide the Dice coefficient for muscular fat alone in Section 4.2.4, along with the Dice coefficients for different applications. Specifically, we calculate the Dice coefficients for muscular fat + VAT and muscular fat + SAT, compared with the manual annotations in our internal dataset.

4.2.1 Internal segmentation performance

For our internal evaluation, without comparisons to other methods, we utilize the Dice coefficient to compare our auto-segmented labels with manual annotations for skeletal muscle, subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT). Table 2 summarizes the segmentation performance for both internal and external datasets, using the Dice coefficient (Section 3.2.2) and MRAE (Section 3.2.3). The model consistently performs best at the L3 slice, with higher average Dice coefficients (93.19%) and

lower MRAE (5.31%) compared to the T12-L4 sub-volume and all slices. Among the tissues, the model achieves the highest segmentation accuracy for SAT, which consistently shows superior Dice scores and lower MRAEs compared to skeletal muscle and VAT. The external dataset follows a similar trend, with the best performance observed at L3 (average Dice of 91.91% and MRAE of 6.15%) and the highest accuracy for SAT.

Notably, there is a label inconsistency between the annotations in our internal dataset and those in the SAROS dataset. Specifically, the SAROS annotation includes skin as part of the SAT label. To address this discrepancy, we applied a simple post-processing step to our model by dilating our SAT segmentation to include the skin. The detailed process for this post-processing is described in Appendix B. However, the post-processing step only mimics the inclusion of skin in the segmentation, which still leaves a gap between the two segmentation approaches.

4.2.2 Comparison with benchmark models

In this experiment we choose the algorithm proposed by Hou et al. (Hou et al., 2024) and TotalSegmentator model (Wasserthal et al., 2023) as the benchmark. Performance is evaluated by the Dice coefficient compared with segmented mask and the public available skeletal muscle and SAT on SAROS dataset (Koitka et al., 2023; Clark et al., 2013). To compare our method with the chosen benchmarks, we follow the instructions provided in (Hou et al., 2024), constraining the analysis to the abdomen section, specifically L1–L5 and T9–T12 following the instructions provided (Hou et al., 2024). The performance results are illustrated in Table 3. As a result, our model outperform the enhanced segmentation model (Hou et al., 2024) by 2.40% for skeletal muscle and 10.26% for SAT. Additionally, it surpasses the TotalSegmentator (Wasserthal et al., 2023) by 7.81% for skeletal muscle and 14.36% for SAT. Notably, due to license restrictions, our evaluation dataset is a large subset of theirs, with 650 commercially licensed volumes used in our study compared to 900 volumes in theirs. However, due to the considerable amount of data and data overlap, it is still representative of the original dataset, ensuring the confidence of our advancements.

4.2.3 Analysis metric evaluation

For the metric evaluation, we utilize our auto-segmentation to measure four selected body composition metrics: muscle density, VAT/SAT ratio, muscle volume, and SMI, in both 2D and 3D settings on our internal test dataset. These results are compared with the body composition metrics derived from manual annotations. The MRAE result is presented in Table 4.

The model demonstrates the best performance in mea-

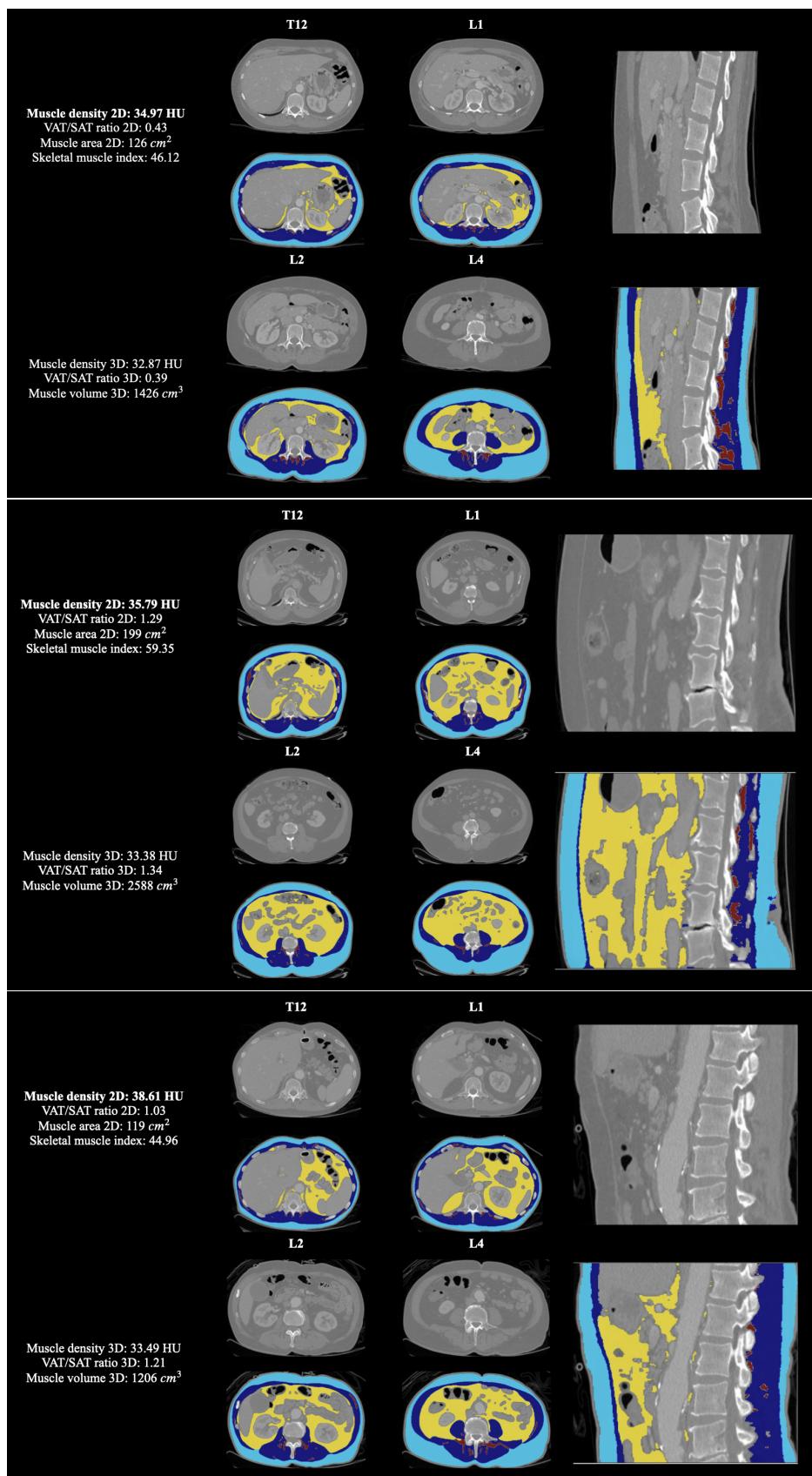


Figure 2: Qualitative evaluation of our segmentation model for 3D setting: Four slices corresponding to T12, L1, L2, and L4 are displayed on the left, while the stacked slices from T12 to L4 are visualized in the sagittal view on the right side of the figure. The visualization volumes are ordered by increasing of 2D muscle density (in bold). In the segmentation, *dark blue* shows skeletal muscle, *light blue* SAT, *yellow* VAT, and *maroon* muscular fat.

Internal Dataset									
	Skeletal Muscle		SAT		VAT		Average		
	Dice ↑ (%)	MRAE ↓ (%)	Dice ↑ (%)	MRAE ↓ (%)	Dice ↑ (%)	MRAE ↓ (%)	Dice ↑ (%)	MRAE ↓ (%)	
L3	92.33 ± 4.42	7.02 ± 8.16	94.55 ± 5.21	3.29 ± 3.77	92.70 ± 5.00	5.61 ± 6.19	93.19 ± 4.88	5.31 ± 6.04	
T12-L4	91.95 ± 5.76	6.47 ± 7.46	93.38 ± 8.57	5.44 ± 8.68	91.29 ± 8.93	7.02 ± 10.49	92.21 ± 7.09	6.31 ± 8.88	
All Slices	91.85 ± 3.37	4.13 ± 4.10	94.06 ± 4.25	3.06 ± 3.97	89.45 ± 7.09	5.91 ± 5.51	91.79 ± 4.90	4.37 ± 4.53	
External Dataset									
	Skeletal Muscle		SAT		VAT		Average		
	Dice ↑ (%)	MRAE ↓ (%)	Dice ↑ (%)	MRAE ↓ (%)	Dice ↑ (%)	MRAE ↓ (%)	Dice ↑ (%)	MRAE ↓ (%)	
L3	91.70 ± 3.45	8.49 ± 5.26	92.12 ± 4.54	3.80 ± 3.37	-	-	91.91 ± 3.99	6.15 ± 4.32	
T12-L4	91.03 ± 3.77	6.61 ± 4.24	92.59 ± 4.35	4.39 ± 3.32	-	-	91.81 ± 4.06	5.50 ± 3.78	
All Slices	89.68 ± 2.75	7.35 ± 3.80	90.27 ± 4.83	3.60 ± 3.39	-	-	89.98 ± 3.79	5.48 ± 3.60	

Table 2: **Internal segmentation Performance:** Segmentation performance for skeletal muscle, subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT) across the internal and external datasets, reported using Dice scores (↑) and MRAE (↓). Results are provided for L3, T12-L4, and all slices, highlighting the model’s superior performance at the L3 slice and on SAT compared to skeletal muscle and VAT. The “Average” column provides the mean Dice score and MRAE across the reported tissues. VAT performance is unavailable due to the absence of VAT annotation in the SAROS dataset.

	Skeletal Muscle (%)	SAT (%)
TotalSegmentator (Wasserthal et al., 2023)	83.2 ± 4.6 [80.5, 86.4]	80.8 ± 10.4 [76.7, 87.7]
Enhanced Segmentation (Hou et al., 2024)	87.6 ± 3.3 [85.6, 90.0]	83.8 ± 10.9 [80.7, 90.5]
Ours	89.7 ± 3.2 [88.3, 91.7]	92.4 ± 3.7 [91.0, 94.7]

Table 3: **Comparison with benchmark models:** Segmentation performance for skeletal muscle and subcutaneous adipose tissue (SAT) across the external dataset is reported using Dice scores. The scores are presented as mean, standard deviation, and interquartile range (IQR). The bolded result highlights the best-performing model among the evaluated models.

suring muscle density among four body compositions, with an MRAE lower than 5% on both internal and external datasets for both 2D and 3D measurements. Across all body composition metrics for both datasets, the model achieves an MRAE lower than 10%, showcasing its robustness.

4.2.4 Muscular fat segmentation

During the literature review, we observed inconsistencies in how muscular fat (both intra-muscular and inter-muscular fat) is classified in research. While some studies include muscular fat as part of skeletal muscle measurements (Hou et al., 2024; Van der Werf et al., 2018), others classify it under VAT (Camus et al., 2014; Wirtz et al., 2021; Connelly et al., 2013), and a smaller subset considers it part of SAT (Ozturk et al., 2020; Magudia et al., 2021). Consequently, we attempted to segment muscular fat as a separate label in our segmentation model.

In this section, we present the Dice coefficient perfor-

mance for muscular fat segmented independently, muscular fat included as part of SAT, muscular fat included as part of VAT, and muscular fat included as part of muscle. Notably, for all skeletal muscle evaluations discussed in the previous sections, we follow the methodology adopted in prior studies, where muscular fat is evaluated as part of muscle segmentation (Section 4.2.14.2.2). Although the segmentation of muscular fat itself demonstrates a relatively low Dice coefficient ($56.27 \pm 10.33\%$) compared to manual annotations on our internal dataset, incorporating muscular fat into other labels—specifically muscle, SAT, and VAT, as is common in body composition measurements—results in high Dice coefficients across all slices ($91.85 \pm 3.37\%$, $92.35 \pm 4.6\%$, and $85.19 \pm 6.73\%$, respectively).

5. Body composition vs. demographic analysis

In the following sections, we analyze the body composition measurements generated by our algorithm, highlight-

Internal Dataset				
	Muscle density (% among range -29 to +150 HU)	VAT/SAT ratio (%)	Muscle area/volume (%)	SMI (%)
2D	2.43 ± 1.67	5.43 ± 5.26	9.81 ± 10.27	9.81 ± 10.27
3D	4.11 ± 2.94	5.09 ± 4.82	8.44 ± 8.03	-
External Dataset				
	Muscle density (% among range -29 to +150 HU)	VAT/SAT ratio (%)	Muscle area/volume (%)	SMI (%)
2D	4.47 ± 2.44	-	9.20 ± 5.31	-
3D	4.71 ± 2.22	-	6.61 ± 4.24	-

Table 4: **Analysis metric evaluation performance:** The performance of our segmentation model on both internal and external datasets is evaluated by comparing the four body composition metrics automatically calculated by our model with the ground truth measured from manual annotations. The evaluation is based on MRAE (\downarrow).

ing their relationships with patients' age, sex, and racial groups. The results produced by our algorithm are compared with previous body composition findings reported in leading medical journals. The aim of this section is to demonstrate the accuracy of our body composition metrics calculation. While the calculated metric values may differ slightly from those reported in previous studies due to variations in population distribution, the trends shown in this analysis strongly resemble those established before.

5.1 Body composition metrics vs. age

To ensure a sufficient sample size for analysis, we divided the age range into six distinct groups, each containing at least 20 instances. The observed trends in muscle area, SAT, VAT area, and SMI with increasing age closely align with findings from previous studies (Magudia et al., 2021), with both the trends and absolute measurement values showing strong consistency across age groups. Specifically, muscle density decreases with age, while VAT and the VAT/SAT ratio increase.

5.2 Body composition metrics vs. sex

The left column sub-figures in Figure 4 illustrate muscle density, SAT/VAT ratio, muscle area, and SMI versus gender, respectively, as measured at the L3 level. Similarly, the corresponding measurements for muscle density, SAT/VAT ratio, muscle volume as measured from T12 to L4, are shown in the right column sub-figures. The mean and standard deviation of these body composition metrics are consistent with those reported in previous studies (Van der Werf et al., 2018; Graffy et al., 2019). Our measurement also aligns with the previous findings that compared to female, male typically have a higher muscle density, VAT/SAT ratio, muscle area, SMI (Kammerlander et al., 2021; Van der Werf et al., 2018; Graffy et al., 2019).

5.3 Body composition metrics vs. race group

For the race groups, to ensure sufficient data size for analysis, we conducted the analysis based only on two race groups: Caucasian/White and Black or African American. The relationship between race groups and multiple body composition metrics is demonstrated in Figure 5. Few previous studies have exclusively analyzed body composition across different races, limiting our ability for direct comparisons. However, several studies have examined the combined impact of both sex and race. For example, (Magudia et al., 2021) demonstrates that Black or African American individuals, on average, have larger muscle areas and higher SMI for both males and females. Similarly, (Beasley et al., 2009) reports an average abdominal visceral fat area of 152.0 for White individuals and 129.9 for Black individuals, as well as an average abdominal subcutaneous fat area of 266.0 for White individuals and 312.1 for Black individuals. Although our study and (Beasley et al., 2009) have different population distributions, with the latter being limited to healthy elderly adults, the measurement differences between the two studies for all metrics are within 10 %.

6. Discussion and future work

To mitigate the gap that there are few publicly available deep learning-based CT segmentation and body composition measurement models for abdominal muscle and fat, we built this model based on nnU-Net ResEnc XL. This model is able to segment skeletal muscle, subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT) across the chest, abdomen, and pelvis in axial CT images. It additionally automatically measures muscle density, visceral-to-subcutaneous fat (VAT/SAT) ratio, muscle area/volume, and SMI. All the code will be made publicly available at <https://github.com/mazurowski-lab/CT-Muscle-and-Fat-Segmentation.git>.

This study highlights the strong capability of our model

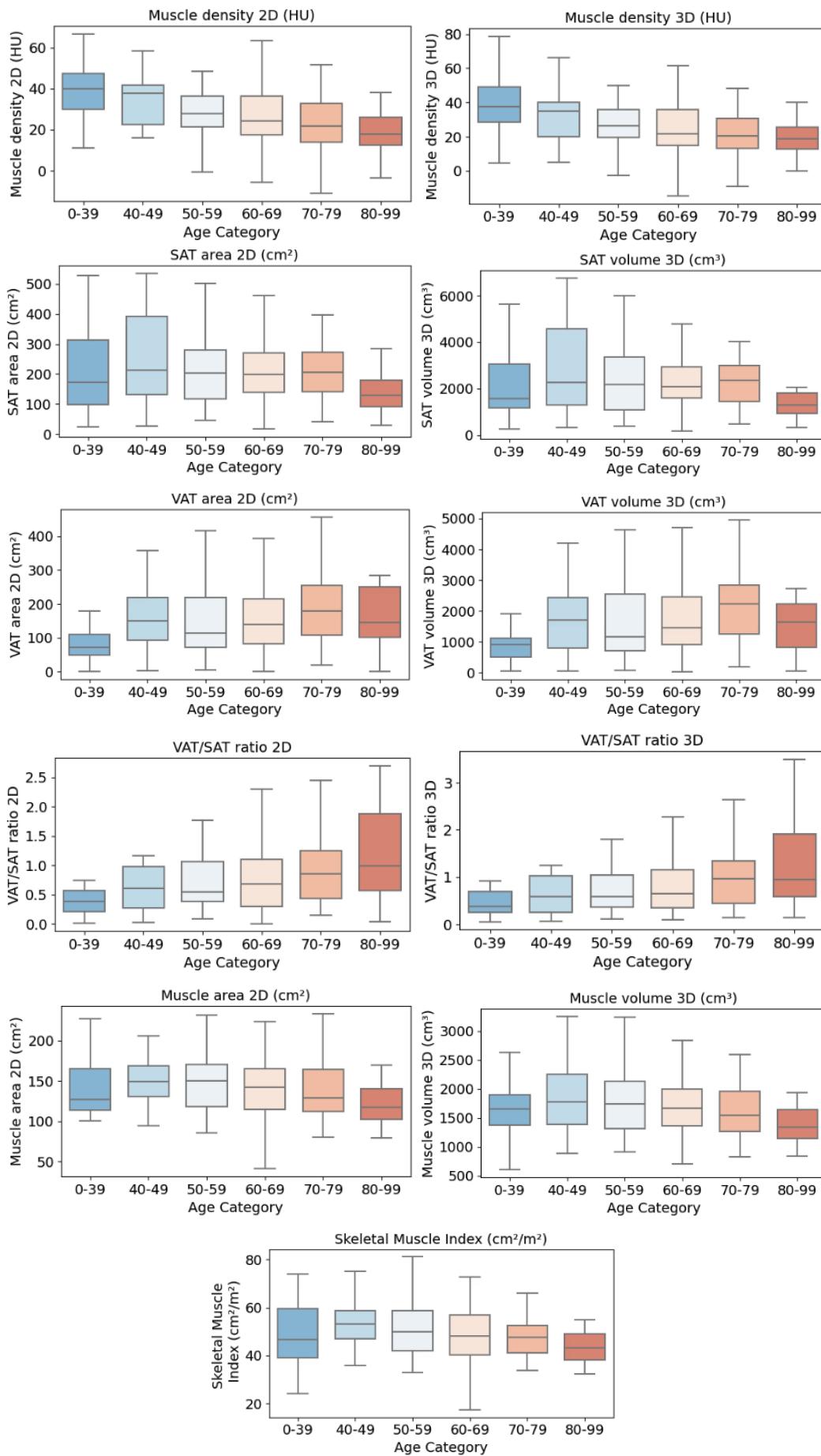


Figure 3: Body composition metrics vs. age categories.

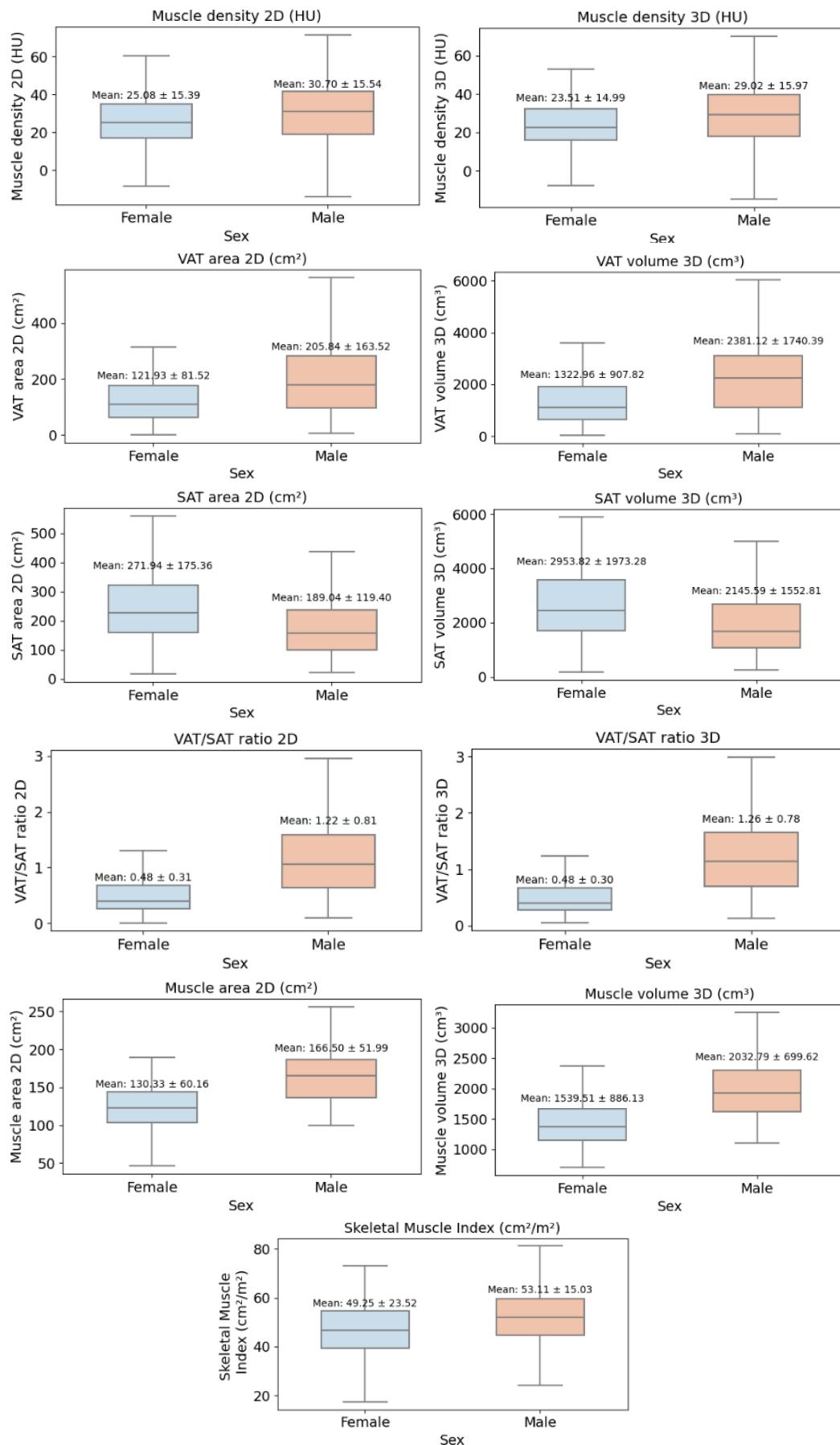


Figure 4: Body composition metrics vs. sex.

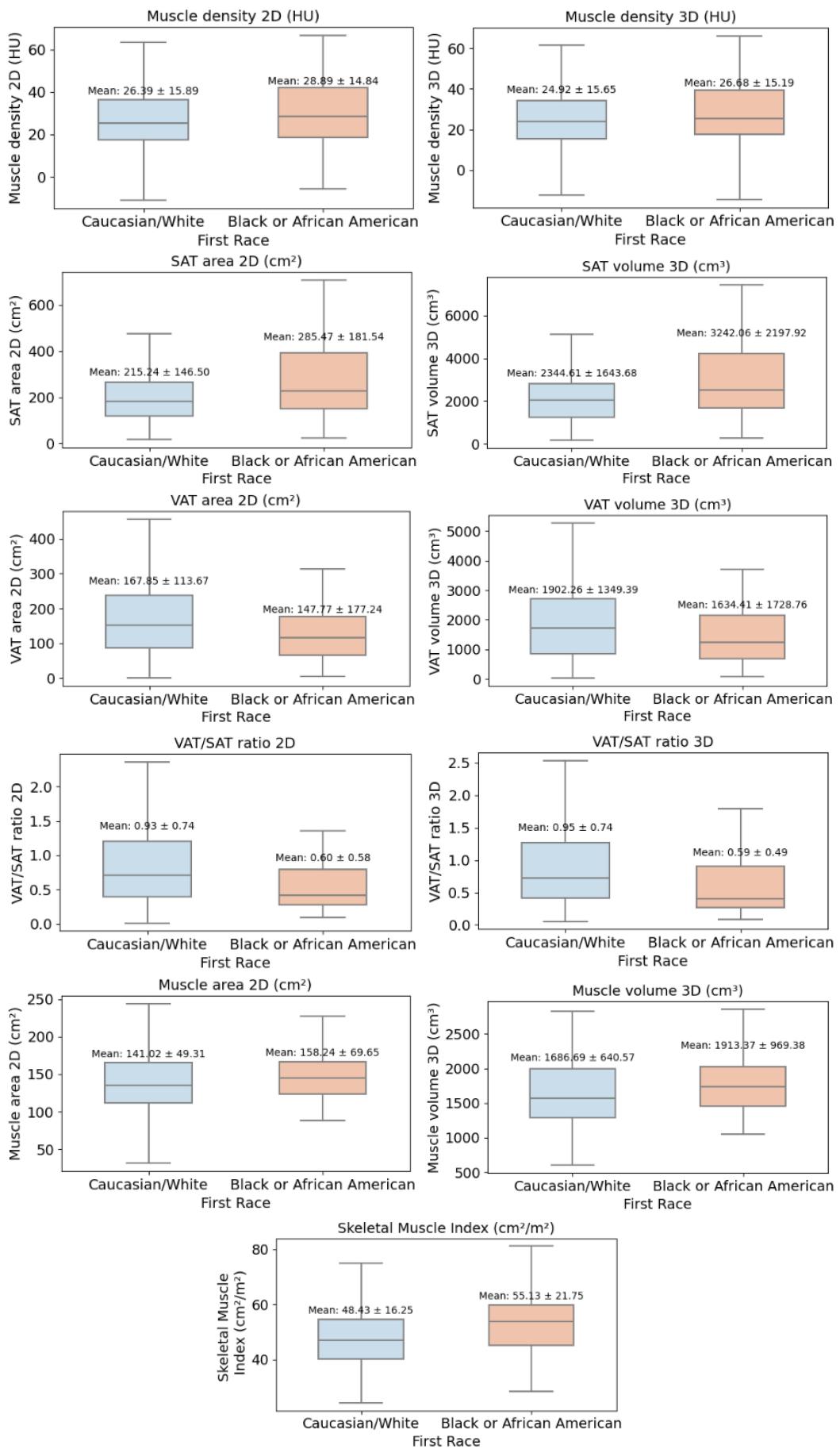


Figure 5: Body composition metrics vs. race.

in segmenting skeletal muscle, SAT, and VAT across the chest, abdomen, and pelvis in axial CT images. As detailed in Section 4.2.1, the model achieves an average Dice score of $91.79 \pm 4.90\%$ across all slices and all four labels in the internal dataset. In the external dataset, the average Dice score is $89.98 \pm 3.79\%$.

The model demonstrates even better performance when segmenting the L3 slice and the T12–L4 region. For the L3 slice, it achieves an average Dice score of $93.19 \pm 4.88\%$ in the internal dataset and $91.91 \pm 3.99\%$ in the external dataset. For the T12–L4 region, the average Dice score is $92.21 \pm 7.09\%$ on internal dataset and $91.81 \pm 4.06\%$ on external dataset.

When compared to previous methods, our model shows significant improvements, outperforming the recently published in-house segmentation model (Hou et al., 2024) by 2.40% for skeletal muscle and 10.26% for SAT. Additionally, it surpasses the TotalSegmentator (Wasserthal et al., 2023) by 7.81% for skeletal muscle and 14.36% for SAT. These evaluations are based on manual annotations from the publicly available SAROS dataset. A detailed comparison with benchmark models is provided in Section 4.2.2.

Apart from segmentation performance, our model also demonstrates high accuracy in measuring commonly used body composition metrics, including muscle density, visceral-to-subcutaneous fat ratio, muscle area/volume, and SMI in both 2D and 3D settings. The average MRAE for all metrics is below 10%. As detailed in Section 4.2.3, the model achieves its best performance in measuring muscle density, with an MRAE of less than 5% compared to manual annotations across internal and external datasets in both 2D and 3D settings.

Furthermore, utilizing our model, we performed body composition metrics analysis across different age, sex, and race groups on 371 randomly selected patients from Duke Hospital. The results demonstrate clear differences in muscle density, adipose tissue distribution, and SMI among patients of different ages, sexes, and races. With increasing age, there was a noticeable decline in muscle density and SMI, coupled with an increase in visceral adipose tissue (VAT) and the VAT/SAT ratio, indicating age-related muscle loss and fat redistribution. Sex-based comparisons revealed that males generally had higher muscle density, muscle volume, and SMI, while females exhibited higher subcutaneous adipose tissue (SAT) levels. Additionally, race-based analysis showed that Black or African American individuals had higher muscle mass and SAT but lower VAT levels and VAT/SAT ratios compared to Caucasian/White individuals. All the findings also align with results from previous studies, highlighting the robustness of our model for both segmentation and body composition measurement.

Despite the promising results, this study has several limitations that present opportunities for further development and improvement.

First, the scope of the study is restricted to three body regions and relies solely on axial views. To enhance the generalizability and clinical utility of our approach, future work will focus on expanding the analysis to additional body regions, such as the hip, leg, and shoulder, which are also commonly assessed in body composition studies.

While our model demonstrates reasonable performance on sagittal and coronal views by stacking segmented axial slices and extracting intersections across different planes, as illustrated in Figure 2, this approach has inherent limitations. For example, since axial slices are segmented independently, which may lead to inconsistencies between adjacent slices. We recognize the potential benefits of directly incorporating sagittal and coronal views into the training and evaluation pipeline, which may improve segmentation accuracy and consistency across all anatomical planes.

Lastly, although the segmentation model includes a muscular fat label, its performance is comparatively lower than that of the other three labels. This discrepancy is primarily due to variability in annotation granularity among different annotators. To enhance annotation consistency in future versions, we will establish clear annotation standards. Specifically, we will define muscular fat regions by applying a Hounsfield Unit (HU) threshold between -220 and -50 for fat tissue, as suggested by Chougule et al. (Chougule et al., 2018), and retain only contiguous fat regions comprising more than six pixels.

Acknowledgments

This work was supported through a partnership between the Duke Departments of Surgery and Radiology and the Duke Spark Initiative for AI in Medical Imaging.

Ethical Standards

The research protocol was approved by the Duke Health System Institutional Review Board (IRB) with ethical standards for research and manuscript preparation, adhering to all relevant laws and regulations concerning the treatment of human subjects and animals.

Conflicts of Interest

We declare we don't have conflicts of interest.

Data availability

The external dataset utilized for model evaluation and anal-

ysis in this study is publicly accessible (Koitka et al., 2023; Clark et al., 2013). However, the internal dataset is currently unavailable, as de-identifying the data requires an extensive institutional review process. Readers interested in evaluating the accuracy of the method can access and utilize the publicly available datasets, which are readily accessible and user-friendly. The code for this study is also publicly available at <https://github.com/mazurowski-lab/CT-Muscle-and-Fat-Segmentation.git>.

References

- Othman Al-Sawaf, Jakob Weiss, Marcin Skrzypski, Jie Min Lam, Takahiro Karasaki, Francisco Zambrana, Andrew C Kidd, Alexander M Frankell, Thomas BK Watkins, Carlos Martínez-Ruiz, et al. Body composition and lung cancer-associated cachexia in tracerx. *Nature medicine*, 29(4):846–858, 2023.
- B. Albertina, M. Watson, C. Holback, R. Jarosz, S. Kirk, Y. Lee, K. Rieger-Christ, and J. Lemmerman. The cancer genome atlas lung adenocarcinoma collection (tcga-luad) (version 4) [data set], 2016. URL <https://doi.org/10.7937/K9/TCIA.2016.JGNIHEP5>.
- Sumbul Ali and Jose M Garcia. Sarcopenia, cachexia and aging: diagnosis, mechanisms and therapeutic options—a mini-review. *Gerontology*, 60(4):294–305, 2014.
- Behrang Amini, Sean P Boyle, Robert D Boutin, and Leon Lenchik. Approaches to assessment of muscle mass and myosteatosis on computed tomography: a systematic review. *The Journals of Gerontology: Series A*, 74(10):1671–1678, 2019.
- George A Antoniou, Djamila Rojo, Stavros A Antoniou, Aws Alfahad, Francesco Torella, and Maciej T Juszczak. Effect of low skeletal muscle mass on post-operative survival of patients with abdominal aortic aneurysm: a prognostic factor review and meta-analysis of time-to-event data. *European Journal of Vascular and Endovascular Surgery*, 58(2):190–198, 2019.
- Aisha A Arayne, Richard Gartrell, Jing Qiao, Paul N Baird, and Justin MC Yeung. Comparison of ct derived body composition at the thoracic t4 and t12 with lumbar l3 vertebral levels and their utility in patients with rectal cancer. *BMC cancer*, 23(1):56, 2023.
- S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazemiraj, H. MacMahon, E. J. R. Van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P. Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. V. Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke. Data from lidc-idri [data set], 2015. URL <https://doi.org/10.7937/K9/TCIA.2015.L09QL9SX>.
- Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011.
- S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, J. Shrager, A. Quon, D. Rubin, S. Plevritis, and S. Napel. Data for nsclc radiogenomics (version 4) [data set], 2017. URL <https://doi.org/10.7937/K9/TCIA.2017.7hs46erv>.
- Shaimaa Bakr, Olivier Gevaert, Sebastian Echegaray, Kelsey Ayers, Mu Zhou, Majid Shafiq, Hong Zheng, Jalen Anthony Benson, Weiruo Zhang, Ann NC Leung, et al. A radiogenomic dataset of non-small cell lung cancer. *Scientific data*, 5(1):1–9, 2018.
- Savita Bansal, Meenakshi Vachher, Taruna Arora, Bhupender Kumar, and Archana Burman. Visceral fat: A key mediator of nafld development and progression. *Human Nutrition & Metabolism*, page 200210, 2023.
- Vickie Baracos and Seyyed Mohammad Reza Kazemi-Bajestani. Clinical outcomes related to muscle mass in humans with cancer and catabolic illnesses. *The international journal of biochemistry & cell biology*, 45(10):2302–2308, 2013.
- Vickie E Baracos, Tony Reiman, Marina Mourtzakis, Ioannis Gioulbasanis, and Sami Antoun. Body composition in patients with non- small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. *The American journal of clinical nutrition*, 91(4):1133S–1137S, 2010.
- David DB Bates and Perry J Pickhardt. Ct-derived body composition assessment as a prognostic tool in oncologic patients: from opportunistic research to artificial intelligence-based clinical implementation. *American Journal of Roentgenology*, 219(4):671–680, 2022.

- Lydia E Beasley, Annemarie Koster, Anne B Newman, M Kassim Javaid, Luigi Ferrucci, Stephen B Kritchevsky, Lewis H Kuller, Marco Pahor, Laura A Schaap, Marjolein Visser, et al. Body composition measures from ct and inflammation. *Obesity (Silver Spring, Md.)*, 17(5):1062, 2009.
- Paul Blanc-Durand, J-B Schiratti, Kathryn Schutte, Paul Jehanno, Paul Herent, Frédéric Pigneur, Olivier Lucidarme, Yassine Benaceur, Alexandre Sadate, Alain Luciani, et al. Abdominal musculature segmentation and surface prediction from ct using deep learning for sarcopenia assessment. *Diagnostic and Interventional Imaging*, 101(12):789–794, 2020.
- RB Boer, KI Jones, S Ash, GI Boxel, RS Gillies, T O'Donnell, JP Ruurda, B Sgromo, MA Silva, and ND Maynard. Impact on postoperative complications of changes in skeletal muscle mass during neoadjuvant chemotherapy for gastro-oesophageal cancer. *BJS open*, 4(5):847–854, 2020.
- NA Bradley, CSD Roxburgh, DC McMillan, and GJK Guthrie. The relationship between pre-operative psoas and skeletal muscle parameters and survival following endovascular aneurysm repair: a systematic review and meta-analysis. *Scientific Reports*, 12(1):16663, 2022.
- Vincent Camus, Helene Lanic, Jérôme Kraut, Romain Modzelewski, Florian Clatot, Jean M Picquenot, Nathalie Contentin, Pascal Lenain, Luminata Groza, Emilie Lemasle, et al. Prognostic impact of fat tissue loss and cachexia assessed by computed tomography scan in elderly patients with diffuse large b-cell lymphoma treated with immunochemotherapy. *European Journal of Haematology*, 93(1):9–18, 2014.
- Peggy M Cawthon. Assessment of lean mass and physical performance in sarcopenia. *Journal of Clinical Densitometry*, 18(4):467–471, 2015.
- Elizabeth M Cespedes Feliciano, Karteek Popuri, Dana Cobzas, Vickie E Baracos, Mirza Faisal Beg, Arafat Dad Khan, Cydney Ma, Vincent Chow, Carla M Prado, Jingjie Xiao, et al. Evaluation of automated computed tomography segmentation to assess body composition and mortality associations in cancer patients. *Journal of cachexia, sarcopenia and muscle*, 11(5):1258–1269, 2020.
- Alan Chait and Laura J Den Hartigh. Adipose tissue distribution, inflammation and its metabolic consequences, including diabetes and cardiovascular disease. *Frontiers in cardiovascular medicine*, 7:522637, 2020.
- Yu-Yao Chang and Bill Cheng. Prognostic impact of myosteatosis in patients with colorectal cancer undergoing curative surgery: an updated systematic review and meta-analysis. *Frontiers in Oncology*, 14:1388001, 2024.
- Chiao-Nan Chen, Kuo-Jen Hsu, Shu-Chen Chen, and Kuei-Yu Chien. What really matters mobility of middle-aged and older adults: low muscle mass or obesity? *Physiology*, 38(S1):5730639, 2023.
- VN Chougule, Arati Mulay, and BB Ahuja. Clinical case study: spine modeling for minimum invasive spine surgeries (miss) using rapid prototyping. *Bone (CT)*, 226:3071, 2018.
- MC Jobin Christ and RMS Parvathi. Fuzzy c-means algorithm for medical image segmentation. In *2011 3rd International conference on electronics computer technology*, volume 4, pages 33–36. IEEE, 2011.
- Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26:1045–1057, 2013.
- Laura C Cleary, Leslie J Crofford, Douglas Long, Richard Charnigo, Jody Clasey, Francesca Beaman, Kirk A Jenkins, Natasha Fraser, Archana Srinivas, Nicole Dhaon, et al. Does computed tomography-based muscle density predict muscle function and health-related quality of life in patients with idiopathic inflammatory myopathies? *Arthritis Care & Research*, 67(7):1031–1040, 2015.
- TM Connelly, R Sehgal, RF Tappouni, FE Luke, S Deiling, WA Koltun, and E Messaris. Volumetric fat ratio and not body mass index is predictive of ileocolectomy outcomes in crohn's disease patients. *Journal of Surgical Research*, 2(179):190, 2013.
- National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The clinical proteomic tumor analysis consortium pancreatic ductal adenocarcinoma collection (cptac-pda), 2018. URL <https://doi.org/10.7937/K9/TCIA.2018.SC20F018>. Data set.
- Alfonso J Cruz-Jentoft, Gülistan Bahat, Jürgen Bauer, Yves Boirie, Olivier Bruyère, Tommy Cederholm, Cyrus Cooper, Francesco Landi, Yves Rolland, Avan Aihie Sayer, et al. Sarcopenia: revised european consensus on definition and diagnosis. *Age and ageing*, 48(1):16–31, 2019.

- Elise Deluche, Sophie Leobon, Jean Claude Desport, Laurence Venat-Bouvet, Julie Usseglio, and Nicole Tubiana-Mathieu. Impact of body composition on outcome in patients with early breast cancer. *Supportive Care in Cancer*, 26:861–868, 2018.
- Ellen W Demerath, Wei Shen, Miryoung Lee, Audrey C Choh, Stefan A Czerwinski, Roger M Siervogel, and Bradford Towne. Approximation of total visceral adipose tissue with a single magnetic resonance image. *The American journal of clinical nutrition*, 85(2):362–368, 2007.
- Manoj Diwakar, Pardeep Kumar, and Amit Kumar Singh. Ct image denoising using nlm and its method noise thresholding. *Multimedia Tools and Applications*, 79:14449–14464, 2020.
- Richard Matthew Dodds, Helen Clare Roberts, Cyrus Cooper, and Avan Aihie Sayer. The epidemiology of sarcopenia. *Journal of Clinical Densitometry*, 18(4):461–466, 2015.
- Haoyu Dong, Hanxue Gu, Yaqian Chen, Jichen Yang, Yuwen Chen, and Maciej A Mazurowski. Segment anything model 2: an application to 2d and 3d medical images. *arXiv preprint arXiv:2408.00756*, 2024.
- Klaus Engelke, Oleg Museyko, Ling Wang, and Jean-Denis Laredo. Quantitative analysis of skeletal muscle by computed tomography imaging—state of the art. *Journal of orthopaedic translation*, 15:91–103, 2018.
- B. J. Erickson, S. Kirk, Y. Lee, O. Bathe, M. Kearns, C. Gerdes, K. Rieger-Christ, and J. Lemmerman. The cancer genome atlas liver hepatocellular carcinoma collection (tcga-lihc) (version 5) [data set], 2016a. URL <https://doi.org/10.7937/K9/TCIA.2016.IMMQW8UQ>.
- B. J. Erickson, D. Mutch, L. Lippmann, and R. Jarosz. The cancer genome atlas uterine corpus endometrial carcinoma collection (tcga-ucec) (version 4) [data set], 2016b. URL <https://doi.org/10.7937/K9/TCIA.2016.GKJ0ZWAC>.
- Gláucia Regina Falsarella, Lívia Pimenta Renó Gasparotto, Caroline Coutinho Barcelos, Ibsen Bellini Coimbra, Maria Clara Moretto, Mauro Alexandre Pascoa, Talita CB Rezende Ferreira, and Arlete Maria Valente Coimbra. Body composition as a frailty marker for the elderly community. *Clinical interventions in aging*, pages 1661–1667, 2015.
- KCH Fearon and T Preston. Body composition in cancer cachexia. *Transfusion Medicine and Hemotherapy*, 17 (Suppl. 3):63–66, 1990.
- Rocío Fernández-Jiménez, Alicia Sanmartín-Sánchez, Eva Cabrera-César, Francisco Espíndola-Hernández, Isabel Vegas-Aguilar, María del Mar Amaya-Campos, Fiorella Ximena Palmas-Candia, María Claro-Brandner, Josefina Olivares-Alcolea, Víctor José Simón-Frapolli, et al. Ia-body composition ct at t12 in idiopathic pulmonary fibrosis: Diagnosing sarcopenia and correlating with other morphofunctional assessment techniques. *Nutrients*, 16(17):2885, 2024.
- Aaron P Frank, Roberta de Souza Santos, Biff F Palmer, and Deborah J Clegg. Determinants of body fat distribution in humans may provide insight about obesity-related health risks. *Journal of lipid research*, 60(10):1710–1719, 2019.
- Yabo Fu, Joseph E Ippolito, Daniel R Ludwig, Rehan Nizamuddin, Harold H Li, and Deshan Yang. Automatic segmentation of ct images for ventral body composition analysis. *Medical physics*, 47(11):5723–5730, 2020.
- Olivier Gevaert, Jiajing Xu, Chuong D Hoang, Ann N Leung, Yue Xu, Andrew Quon, Daniel L Rubin, Sandy Napel, and Sylvia K Plevritis. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology*, 264(2):387–396, 2012.
- Bret H Goodpaster, F Leland Thaete, and David E Kelley. Composition of skeletal muscle evaluated with computed tomography. *Annals of the New York Academy of Sciences*, 904(1):18–24, 2000.
- Gijs H Goossens. The metabolic phenotype in obesity: fat mass, body fat distribution, and adipose tissue function. *Obesity facts*, 10(3):207–215, 2017.
- Peter M Graffy, Jiamin Liu, Perry J Pickhardt, Joseph E Burns, Jianhua Yao, and Ronald M Summers. Deep learning-based muscle segmentation and quantification at abdominal ct: application to a longitudinal adult screening cohort for sarcopenia assessment. *The British journal of radiology*, 92(1100):20190327, 2019.
- Hanxue Gu, Roy Colglazier, Haoyu Dong, Jikai Zhang, Yaqian Chen, Zafer Yildiz, Yuwen Chen, Lin Li, Jichen Yang, Jay Willhite, et al. Segmentanybone: A universal model that segments any bone at any location on mri. *arXiv preprint arXiv:2401.12974*, 2024.
- Peggi Guenter, Ruba Abdelhadi, Pat Anthony, Allison Blackmer, Ainsley Malone, Jay M Mirtallo, Wendy Phillips, and Helaine E Resnick. Malnutrition diagnoses and associated outcomes in hospitalized patients: United

- states, 2018. *Nutrition in Clinical Practice*, 36(5):957–969, 2021.
- Jin Soo Han, Hyoseon Ryu, In Ja Park, Kyung Won Kim, Yongbin Shin, Sun Ok Kim, Seok-Byung Lim, Chan Wook Kim, Yong Sik Yoon, Jong Lyul Lee, et al. Association of body composition with long-term survival in non-metastatic rectal cancer patients. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, 52(2):563–572, 2020.
- Olga T Hardy, Michael P Czech, and Silvia Corvera. What causes the insulin resistance underlying obesity? *Current Opinion in Endocrinology, Diabetes and Obesity*, 19(2):81–87, 2012.
- N. Heller, N. Sathianathan, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikopoulos, and C. Weight. Data from c4kc-kits [data set], 2019. URL <https://doi.org/10.7937/TCIA.2019.IX49E8NX>.
- Nicholas Heller, Fabian Isensee, Klaus H Maier-Hein, Xiaoshuai Hou, Chunmei Xie, Fengyi Li, Yang Nan, Guangrui Mu, Zhiyong Lin, Miofei Han, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical image analysis*, 67:101821, 2021.
- Robert Hemke, Colleen G Buckless, Andrew Tsao, Benjamin Wang, and Martin Torriani. Deep learning for automated segmentation of pelvic muscles, fat, and bone from ct studies for body composition assessment. *Skeletal radiology*, 49:387–395, 2020.
- Jordan Hernandez-Martinez, Joaquín Perez-Carcamo, Bayron Coñapi-Union, Sebastian Canales-Canales, Mario Negron-Molina, Sergio Avila-Valencia, Izham Cid-Calfucura, Tomas Herrera-Valenzuela, Diego Cisterna, Braulio Henrique Magnani Branco, et al. Relationship between body composition and physical performance by sex in professional basketball players. *Applied Sciences*, 14(20):9165, 2024.
- Benjamin Hou, Tejas Sudharshan Mathai, Jianfei Liu, Christopher Parnell, and Ronald M Summers. Enhanced muscle and fat segmentation for ct-based body composition analysis: a comparative study. *International journal of computer assisted radiology and surgery*, pages 1–8, 2024.
- Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- Fabian Isensee, Tassilo Wald, Constantin Ulrich, Michael Baumgartner, Saikat Roy, Klaus Maier-Hein, and Paul F Jaeger. nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 488–498. Springer, 2024.
- Toshiaki Iwase, Takafumi Sangai, Takeshi Nagashima, Masahiro Sakakibara, Junta Sakakibara, Shouko Hayama, Emi Ishigami, Takahito Masuda, and Masaru Miyazaki. Impact of body fat distribution on neoadjuvant chemotherapy outcomes in advanced breast cancer patients. *Cancer medicine*, 5(1):41–48, 2016.
- Wei Ji, XiangLiang Liu, KaiWen Zheng, PengFei Liu, YiXin Zhao, Jin Lu, LingLing Zhao, TingTing Liang, JiuWei Cui, and Wei Li. Thresholds of visceral fat area and percent of body fat to define sarcopenic obesity and its clinical consequences in chinese cancer patients. *Clinical Nutrition*, 41(3):737–745, 2022.
- Nina Kaegi-Braun, Marlena Mueller, Philipp Schuetz, Beat Mueller, and Alexander Kutz. Evaluation of nutritional support and in-hospital mortality in patients with malnutrition. *JAMA Network Open*, 4(1):e2033433–e2033433, 2021.
- BM Kaess, A Pedley, JM Massaro, J Murabito, U Hoffmann, and CS Fox. The ratio of visceral to subcutaneous fat, a metric of body fat distribution, is a unique correlate of cardiometabolic risk. *Diabetologia*, 55:2622–2630, 2012.
- Lisa Kakinami, Sabine Plummer, Tamara R Cohen, Sylvia Santosa, and Jessica Murphy. Body-composition phenotypes and their associations with cardiometabolic risks and health behaviours in a representative general us sample. *Preventive Medicine*, 164:107282, 2022.
- N Kamiya, X Zhou, H Chen, T Hara, H Hoshi, R Yokoyama, M Kanematsu, and H Fujita. Automated recognition of the psoas major muscles on x-ray ct images. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3557–3560. IEEE, 2009.
- Naoki Kamiya, Xiangrong Zhou, H Chen, Chisako Muramatsu, Takeshi Hara, Ryujiro Yokoyama, Masayuki Kanematsu, Hiroaki Hoshi, and Hiroshi Fujita. Automated segmentation of rectus abdominis muscle using shape model in x-ray ct images. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 7993–7996. IEEE, 2011.

- Andreas A Kammerlander, Asya Lyass, Taylor F Mahoney, Joseph M Massaro, Michelle T Long, Ramachandran S Vasan, and Udo Hoffmann. Sex differences in the associations of visceral adipose tissue and cardiometabolic and cardiovascular disease risk: the framingham heart study. *Journal of the American Heart Association*, 10(11):e019968, 2021.
- P. Kinahan, M. Muzi, B. Bialecki, and L. Coombs. Data from acrin-flt-breast (version 2) [data set], 2017. URL <https://doi.org/10.7937/K9/TCIA.2017.o120zmxg>.
- P. Kinahan, M. Muzi, B. Bialecki, B. Herman, and L. Coombs. Data from the acrin 6668 trial nsclc-fdg-pet (version 2) [data set], 2019. URL <https://doi.org/10.7937/tcia.2019.30ilqfc1>.
- S. Kirk, Y. Lee, P. Kumar, J. Filippini, B. Albertina, M. Watson, K. Rieger-Christ, and J. Lemmerman. The cancer genome atlas lung squamous cell carcinoma collection (tcga-lusc) (version 4) [data set], 2016. URL <https://doi.org/10.7937/K9/TCIA.2016.TYGKKFMQ>.
- Sven Koitka, Lennard Kroll, Eugen Malamutmann, Arzu Oezcelik, and Felix Nensa. Fully automated body composition analysis in routine ct imaging using 3d semantic segmentation convolutional neural networks. *European radiology*, 31:1795–1804, 2021.
- Sven Koitka, Giulia Baldini, Lennard Kroll, Natalie van Landeghem, Johannes Haubold, et al. Saros-a large, heterogeneous, and sparsely annotated segmentation dataset on ct imaging data (saros). *The Cancer Imaging Archive*, 2023.
- Lale Kostakoglu, Fenghai Duan, Michael O Idowu, Paul R Jolles, Harry D Bear, Mark Muzi, Jean Cormack, John P Muzi, Daniel A Pryma, Jennifer M Specht, et al. A phase ii study of 3'-deoxy-3'-18f-fluorothymidine pet in the assessment of early response of breast cancer to neoadjuvant chemotherapy: results from acrin 6688. *Journal of Nuclear Medicine*, 56(11):1681–1689, 2015.
- Amanika Kumar, Michael R Moynagh, Francesco Multinu, William A Cliby, Michaela E McGree, Amy L Weaver, Phillip M Young, Jamie N Bakkum-Gamez, Carrie L Langstraat, Sean C Dowdy, et al. Muscle composition measured by ct scan is a measurable predictor of overall survival in advanced ovarian cancer. *Gynecologic oncology*, 142(2):311–316, 2016.
- Ricardo Ladeiras-Lopes, Francisco Sampaio, Nuno Bettencourt, Ricardo Fontes-Carvalho, Nuno Ferreira, Adelino Leite-Moreira, and Vasco Gama. The ratio between visceral and subcutaneous abdominal fat assessed by computed tomography is an independent predictor of mortality and cardiac events. *Revista Española de Cardiología (English Edition)*, 70(5):331–337, 2017.
- Hyunkwang Lee, Fabian M Troschel, Shahein Tajmir, Georg Fuchs, Julia Mario, Florian J Fintelmann, and Synho Do. Pixel-level deep segmentation: artificial intelligence quantifies muscle on computed tomography for body morphometric analysis. *Journal of digital imaging*, 30:487–498, 2017.
- Yoon Seong Lee, Namki Hong, Joseph Nathanael Witanto, Ye Ra Choi, Junghoan Park, Pierre Decazes, Florian Eude, Chang Oh Kim, Hyeon Chang Kim, Jin Mo Goo, et al. Deep neural network for automatic volumetric segmentation of whole-body ct images for body composition assessment. *Clinical Nutrition*, 40(8):5038–5046, 2021.
- P. Li, S. Wang, T. Li, J. Lu, Y. HuangFu, and D. Wang. A large-scale ct and pet/ct dataset for lung cancer diagnosis (lung-pet-ct-dx) [data set], 2020. URL <https://doi.org/10.7937/TCIA.2020.NNC2-0461>.
- Daniel Liu, John W Garrett, Matt H Lee, Ryan Zea, Ronald M Summers, and Perry J Pickhardt. Fully automated ct-based adiposity assessment: comparison of the l1 and l3 vertebral levels for opportunistic prediction. *Abdominal Radiology*, 48(2):787–795, 2023.
- Wilhelmus GPM Looijaard, Ingeborg M Dekker, Sandra N Stapel, Armand RJ Girbes, Jos WR Twisk, Heleen M Oudemans-van Straaten, and Peter JM Weijs. Skeletal muscle quality as assessed by ct-derived skeletal muscle density is associated with 6-month mortality in mechanically ventilated critically ill patients. *Critical care*, 20:1–10, 2016.
- F. R. Lucchesi and N. D. Aredes. The cancer genome atlas stomach adenocarcinoma collection (tcga-stad) (version 3) [data set], 2016. URL <https://doi.org/10.7937/K9/TCIA.2016.GDHL9KIM>.
- Mitchell Machtay, Fenghai Duan, Barry A Siegel, Bradley S Snyder, Jeremy J Gorelick, Janet S Reddin, Reginald Munden, Douglas W Johnson, Larry H Wilf, Albert DeNittis, et al. Prediction of survival by [18f] fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: results of the acrin 6668/rtoi 0235 trial. *Journal of clinical oncology*, 31(30):3823–3830, 2013.
- P. Madhavi, S. Patel, and A. S. Tsao. Data from anti-pd-1 immunotherapy lung [data set], 2019. URL <https://doi.org/10.7937/tcia.2019.zjjwb9ip>.

- Kirti Magudia, Christopher P Bridge, Camden P Bay, Ana Babic, Florian J Fintelmann, Fabian M Troschel, Nityanand Miskin, William C Wrobel, Lauren K Brais, Katherine P Andriole, et al. Population-scale ct-based body composition analysis of a large outpatient population using deep learning to derive age-, sex-, and race-specific reference curves. *Radiology*, 298(2):319–329, 2021.
- Dinh Van Chi Mai, Ioanna Drami, Edward T Pring, Laura E Gould, Phillip Lung, Karteek Popuri, Vincent Chow, Mirza F Beg, Thanos Athanasiou, John T Jenkins, et al. A systematic review of automated segmentation of 3d computed-tomography scans for volumetric body composition analysis. *Journal of Cachexia, Sarcopenia and Muscle*, 14(5):1973–1986, 2023.
- George Malietzis, Omer Aziz, Nigel M Bagnall, N Johns, KC Fearon, and JT Jenkins. The role of body composition evaluation by computerized tomography in determining colorectal cancer treatment outcomes: a systematic review. *European Journal of Surgical Oncology (EJSO)*, 41(2):186–196, 2015.
- Lisa Martin, Laura Birdsall, Neil MacDonald, Tony Reiman, M Thomas Clandinin, Linda J McCargar, Rachel Murphy, Sunita Ghosh, Michael B Sawyer, and Vickie E Baracos. Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. *Journal of clinical oncology*, 31(12):1539–1547, 2013.
- Maciej A Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment anything model for medical image analysis: an experimental study. *Medical Image Analysis*, 89:102918, 2023.
- Takako Momose, Yutaka Inaba, Hyonmin Choe, Naomi Kobayashi, Taro Tezuka, and Tomoyuki Saito. Ct-based analysis of muscle volume and degeneration of gluteus medius in patients with unilateral hip osteoarthritis. *BMC musculoskeletal disorders*, 18:1–7, 2017.
- Sandy Napel and Sylvia K. Plevritis. Nsclc radiogenomics: Initial stanford study of 26 cases (nsclc radiogenomics-stanford) [data set], 2014. URL <https://doi.org/10.7937/K9/TCIA.2014.X7ONY6B1>.
- National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The clinical proteomic tumor analysis consortium cutaneous melanoma collection (cptac-cm) (version 11) [data set], 2018a. URL <https://doi.org/10.7937/K9/TCIA.2018.0DU24GZE>.
- National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The clinical proteomic tumor analysis consortium lung squamous cell carcinoma collection (cptac-lscc) (version 15) [data set], 2018b. URL <https://doi.org/10.7937/K9/TCIA.2018.6EMUB5L2>.
- National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The clinical proteomic tumor analysis consortium lung adenocarcinoma collection (cptac-luad) (version 12) [data set], 2018c. URL <https://doi.org/10.7937/K9/TCIA.2018.PAT12TBS>.
- National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The clinical proteomic tumor analysis consortium uterine corpus endometrial carcinoma collection (cptac-ucec) (version 12) [data set], 2019. URL <https://doi.org/10.7937/K9/TCIA.2018.3R3JUISW>.
- Sebastian Nowak, Anton Faron, Julian A Luetkens, Helena L Geißler, Michael Praktiknjo, Wolfgang Block, Daniel Thomas, and Alois M Sprinkart. Fully automated segmentation of connective tissue compartments for ct-based body composition analysis: a deep learning approach. *Investigative radiology*, 55(6):357–366, 2020.
- Yun Hwan Oh, Ji Hyun Moon, Hyeon Ju Kim, and Mi Hee Kong. Visceral-to-subcutaneous fat ratio as a predictor of the multiple metabolic risk factors for subjects with normal waist circumference in korea. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, pages 505–511, 2017.
- Cansu Ozturk, Ozlem Gungor, Veysel Kaplanoglu, MERVE KAŞIKCI, and Selma Ramadan. Relationship between abdominal aortic calcification, abdominal adiposity, and liver density. *JCPSP-Journal of the College of Physicians and Surgeons Pakistan*, 30(12), 2020.
- Michael T Paris. Body composition analysis of computed tomography scans in clinical populations: the role of deep learning. *Lifestyle genomics*, 13(1):28–31, 2020.
- Marie-Eve Piché, Paul Poirier, Isabelle Lemieux, and Jean-Pierre Després. Overview of epidemiology and contribution of obesity and body fat distribution to cardiovascular disease: an update. *Progress in cardiovascular diseases*, 61(2):103–113, 2018.
- Perry J Pickhardt, Peter M Graffy, Ryan Zea, Scott J Lee, Jiamin Liu, Veit Sandfort, and Ronald M Summers. Automated ct biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study. *The Lancet Digital Health*, 2(4):e192–e200, 2020a.

- Perry J Pickhardt, Peter M Graffy, Ryan Zea, Scott J Lee, Jiamin Liu, Veit Sandfort, and Ronald M Summers. Automated abdominal ct imaging biomarkers for opportunistic prediction of future major osteoporotic fractures in asymptomatic adults. *Radiology*, 297(1):64–72, 2020b.
- Karteek Popuri, Dana Cobzas, Nina Esfandiari, Vickie Baracos, and Martin Jägersand. Body composition assessment in axial ct images using fem-based automatic segmentation of skeletal muscle. *IEEE transactions on medical imaging*, 35(2):512–520, 2015.
- Vincenzo Positano, Amalia Gastaldelli, Anna maria Sironi, Maria Filomena Santarelli, Massimo Lombardi, and Luigi Landini. An accurate and robust method for unsupervised assessment of abdominal fat by mri. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 20(4):684–689, 2004.
- Vincenzo Positano, Tore Christiansen, Maria Filomena Santarelli, Steffen Ringgaard, Luigi Landini, and Amalia Gastaldelli. Accurate segmentation of subcutaneous and intermuscular adipose tissue from mr images of the thigh. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 29(3):677–684, 2009.
- Carla MM Prado and Steven B Heymsfield. Lean tissue imaging: a new era for nutritional assessment and intervention. *Journal of Parenteral and Enteral Nutrition*, 38(8):940–953, 2014.
- Ilse Reinders, Marjolein Visser, and Laura Schaap. Body weight and body composition in old age and their relationship with frailty. *Current Opinion in Clinical Nutrition & Metabolic Care*, 20(1):11–15, 2017.
- Ladina Risch, Florian Hotzy, Stefan Vetter, Sascha Hiller, Kathrin Wallimann, Erich Seifritz, and Sonja Mötteli. Assessment of nutritional status and risk of malnutrition using adapted standard tools in patients with mental illness and in need of intensive psychiatric treatment. *International Journal of Environmental Research and Public Health*, 20(1):109, 2022.
- H. Roth, A. Farag, E. B. Turkbey, L. Lu, J. Liu, and R. M. Summers. Data from pancreas-ct (version 2) [data set], 2016. URL <https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU>.
- Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I* 18, pages 556–564. Springer, 2015.
- Iris JG Rutten, David PJ van Dijk, Roy FPM Kruitwagen, Regina GH Beets-Tan, Steven WM Olde Damink, and Toon Van Gorp. Loss of skeletal muscle during neoadjuvant chemotherapy is related to decreased survival in ovarian cancer patients. *Journal of cachexia, sarcopenia and muscle*, 7(4):458–466, 2016.
- Babak Salam, Muntadher Al Zaidi, Alois M Sprinkart, Sebastian Nowak, Maike Theis, Daniel Kuettling, Adem Aksoy, Georg Nickenig, Ulrike Attenberger, Sebastian Zimmer, et al. Opportunistic ct-derived analysis of fat and muscle tissue composition predicts mortality in patients with cardiogenic shock. *Scientific Reports*, 13(1):22293, 2023.
- J. Saltz, M. Saltz, P. Prasanna, R. Moffitt, J. Hajagos, E. Bremer, J. Balsamo, and T. Kurc. Stony brook university covid-19 positive cases [data set], 2021. URL <https://doi.org/10.7937/TCIA.BBAG-2923>.
- Philipp Schuetz, David Seres, Dileep N Lobo, Filomena Gomes, Nina Kaegi-Braun, and Zeno Stanga. Management of disease-related malnutrition for patients being treated in hospital. *The Lancet*, 398(10314):1927–1938, 2021.
- Rashmita Sehgal and Vandana Dixit Kaushik. Ct image denoising using bilateral filter and method noise thresholding in shearlet domain. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 1*, pages 99–106. Springer, 2022.
- Giuseppe Sergi, Caterina Trevisan, Nicola Veronese, Paola Lucato, and Enzo Manzato. Imaging of sarcopenia. *European journal of radiology*, 85(8):1519–1524, 2016.
- Azadeh Sharifi, Andrew P Klein, and Kevin M Koch. Quantitative mri assessment of post-surgical spinal cord injury through radiomic analysis. *Journal of Imaging*, 10(12):312, 2024.
- Andrey Sinelnikov, Chuanxing Qu, David T Fetzer, Jean-Sébastien Pelletier, Michael A Dunn, Allan Tsung, and Alessandro Furlan. Measurement of skeletal muscle area: comparison of ct and mr imaging. *European Journal of Radiology*, 85(10):1716–1721, 2016.
- Sarah Soh, Young Joo Suh, Suji Lee, Yun Ho Roh, Young-Lan Kwak, and Young Jin Kim. Prognostic value of ct body composition analysis for 1-year mortality after transcatheter aortic valve replacement. *European Radiology*, pages 1–11, 2024.

- Stuart-Allison Moffat Staley, Katherine Tucker, Jorge Oldan, Dominic T Moore, Meredith Newton, Michelle Ertel, Lindsay West, and Victoria Bae-Jump. Visceral adiposity as a predictor of survival in patients with epithelial ovarian cancer receiving platinum and taxane-based chemotherapy., 2019.
- Alberto Stefano Tagliafico, Bianca Bignotti, Lorenzo Torri, and Federica Rossi. Sarcopenia: how to measure, when and why. *La radiologia medica*, 127(3):228–237, 2022.
- Qi Tan, Bingliang Gao, Pengde Gao, Chao Fan, and Haijun Li. Quantitative analysis of complex aluminum electrolytes by x-ray fluorescence. *Luminescence*, 39(11):e70013, 2024.
- Muhei Tanaka, Hiroshi Okada, Yoshitaka Hashimoto, Muneaki Kumagai, Hiromi Nishimura, and Michiaki Fukui. Distinct associations of intraperitoneal and retroperitoneal visceral adipose tissues with metabolic syndrome and its components. *Clinical Nutrition*, 40(5):3479–3484, 2021.
- Antti Tolonen, Tomppa Pakarinen, Antti Sassi, Jere Kyttä, William Cancino, Irina Rinta-Kiikka, Said Pertuz, and Otso Arponen. Methodology, clinical applications, and future directions of body composition analysis using computed tomography (ct) images: a review. *European journal of radiology*, 145:109943, 2021.
- Yubing Tong, Jayaram K Udupa, and Drew A Torigian. Optimization of abdominal fat quantification on ct imaging through use of standardized anatomic space: a novel approach. *Medical physics*, 41(6Part1):063501, 2014.
- Michelle L Torres, Lynn C Hartmann, William A Cliby, Kimberly R Kalli, Phillip M Young, Amy L Weaver, Carrie L Langstraat, Aminah Jatoi, Sanjeev Kumar, and Andrea Mariani. Nutritional status, ct body composition measures and survival in ovarian cancer. *Gynecologic oncology*, 129(3):548–553, 2013.
- Inez Trouwborst, Kelly M Jardon, Anouk Gijbels, Gabby Hul, Edith JM Feskens, Lydia A Afman, Jennifer Linge, Gijs H Goossens, and Ellen E Blaak. Body composition and body fat distribution in tissue-specific insulin resistance and in response to a 12-week isocaloric dietary macronutrient intervention. *Nutrition & Metabolism*, 21(1):20, 2024.
- Martin Vallières, Carolyn R Freeman, Sonia R Skamene, and Issam El Naqa. A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology*, 60(14):5471, 2015.
- Martin Vallières, Carolyn R. Freeman, Sonia R. Skamene, and Issam El Naqa. A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities (soft-tissue-sarcoma) [dataset], 2015. URL <https://doi.org/10.7937/K9/TCIA.2015.7G02GSKS>.
- A Van der Werf, JAE Langius, MAE De Van Der Schueren, SA Nurmohamed, KAMI Van Der Pant, S Blauwhoff-Buskermolen, and NJ I Wierdsma. Percentiles for skeletal muscle index, area and radiation attenuation based on computed tomography imaging in a healthy caucasian population. *European journal of clinical nutrition*, 72(2):288–296, 2018.
- Sathish Babu Vasamsetti, Niranjana Natarajan, Samreen Sadaf, Jonathan Florentin, and Partha Dutta. Regulation of cardiovascular health and disease by visceral adipose tissue-derived metabolic hormones. *The Journal of Physiology*, 601(11):2099–2120, 2023.
- Chantal A Vella, Megan C Nelson, Jonathan T Unkart, Iva Miljkovic, and Matthew A Allison. Skeletal muscle area and density are associated with lipid and lipoprotein cholesterol levels: The multi-ethnic study of atherosclerosis. *Journal of clinical lipidology*, 14(1):143–153, 2020.
- Chiara Villa and Niels Lynnerup. Hounsfield units ranges in ct-scans of bog bodies and mummies. *Anthropologischer Anzeiger*, 69(2):127–145, 2012.
- Dennis T Villareal, Marian Banks, Catherine Siener, David R Sinacore, and Samuel Klein. Physical frailty and body composition in obese elderly men and women. *Obesity research*, 12(6):913–920, 2004.
- Benjamin Wang and Martin Torriani. Artificial intelligence in the evaluation of body composition. In *Seminars in Musculoskeletal Radiology*, volume 24, pages 030–037. Thieme Medical Publishers, 2020.
- Ling Wang, Lu Yin, Yue Zhao, Yongbin Su, Wei Sun, Shuo Chen, Yandong Liu, Minghui Yang, Aihong Yu, Giuseppe Guglielmi, et al. Muscle density, but not size, correlates well with muscle strength and physical performance. *Journal of the American Medical Directors Association*, 22(4):751–759, 2021.
- Yunzhi Wang, Yuchen Qiu, Theresa Thai, Kathleen Moore, Hong Liu, and Bin Zheng. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on ct images. *Computer methods and programs in biomedicine*, 144:97–104, 2017.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5), 2023.

Connor A Wathen, Nathan Foje, Tony van Avermaete, Bernadette Miramontes, Sarah E Chapman, Todd A Sasser, Raghuraman Kannan, Steven Gerstler, and W Matthew Leevy. In vivo x-ray computed tomographic imaging of soft tissue with native, intravenous, or oral contrast. *Sensors*, 13(6):6957–6980, 2013.

Alexander D Weston, Panagiotis Korfiatis, Timothy L Kline, Kenneth A Philbrick, Petro Kostandy, Tomas Sakinis, Motokazu Sugimoto, Naoki Takahashi, and Bradley J Erickson. Automated abdominal segmentation of ct scans for body composition analysis using deep learning. *Radiology*, 290(3):669–679, 2019.

Theresa H Wirtz, Sven H Loosen, Maximilian Schulze-Hagen, Ralf Weiskirchen, Lukas Buendgens, Samira Abu Jhaisha, Jonathan F Brozat, Tobias Puengel, Michael Vucur, Pia Paffenholz, et al. Ct-based determination of excessive visceral adipose tissue is associated with an impaired survival in critically ill patients. *PLoS One*, 16(4):e0250321, 2021.

Y Xu, XY Zhu, H Feng, XP Yu, Y Wang, X Rong, and TY Qi. The value of quantitative contrast-enhanced ultrasonography analysis in evaluating central retinal artery microcirculation in patients with diabetes mellitus: comparison with colour doppler imaging. *Clinical Radiology*, 79(4):e560–e566, 2024.

Nobuhisa Yoshikawa, Akira Shirakawa, Kosuke Yoshida, Satoshi Tamauchi, Shiro Suzuki, Fumitaka Kikkawa, and Hiroaki Kajiyama. Sarcopenia as a predictor of survival among patients with organ metastatic cervical cancer. *Nutrition in Clinical Practice*, 35(6):1041–1046, 2020.

Lin Zhang, Huijuan Zhao, Zhongxing Zhou, Mengyu Jia, Limin Zhang, Jingying Jiang, and Feng Gao. Improving spatial resolution with an edge-enhancement model for low-dose propagation-based x-ray phase-contrast computed tomography. *Optics Express*, 29(23):37399–37417, 2021.

Appendix A. Data collections from SAROS

While only a subset of the collections within SAROS is provided with a commercially permitted license (specifically CC BY 3.0 and CC BY 4.0), we exclusively utilized this subset in external evaluation to ensure maximum flexibility for our model. In this section, we provide a detailed list of the dataset collections used in this study, shown in Table 5, including the collection name, scan region (Abdomen, Thorax, Whole-body) assigned by SAROS (Koitka et al., 2023; Clark et al., 2013), and their license type.

Appendix B. Post-processing for label inconsistencies

Label post-processing is performed in two steps. First, a 5×5 structuring element is applied to morphologically dilate the SAT label. Second, the expanded region is constrained to ensure that (1) it does not overlap with any previously labeled areas and (2) it remains within the abdominal region. The abdominal boundary is determined by thresholding at -800 Hounsfield Units (HU), as skin typically exhibits an HU value around this level (Chougule et al., 2018; Villa and Lynnerup, 2012). Figure 6 demonstrates the dilation result with on two randomly selected slices. Notably, the post-processing step merely approximates the inclusion of skin in the segmentation, leaving a remaining discrepancy between the two approaches.

Appendix C. Body composition metrics relationship

Pearson correlation coefficient (r) is the most common method of measuring a linear correlation between two variables, with its definition shown in Equation (3), where $\text{Cov}(X, Y)$ represents the covariance of X and Y , and σ_X , σ_Y are standard deviations of X and Y respectively. This section demonstrates the correlation between four selected body composition metrics: muscle density, VAT/SAT ratio, muscle area, and SMI based both on Pearson correlation coefficient and scatter plots. The 2D and 3D measurements of the same metrics are highly correlated, as shown in Figure 7, with all three Pearson correlation coefficients exceeding 0.96. In this experiment, we utilize only the body composition metrics measured in 2D settings (at the L3 level). The results are shown in Figure 8, as we can observe, except for the relationship between muscle area and SMI (with r equals to 0.94), all other pairs of metrics show insignificant or no linear correlation, with r having an absolute value smaller than 0.2. Scatter plots also do not show a clear monotonic relationship.

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (3)$$

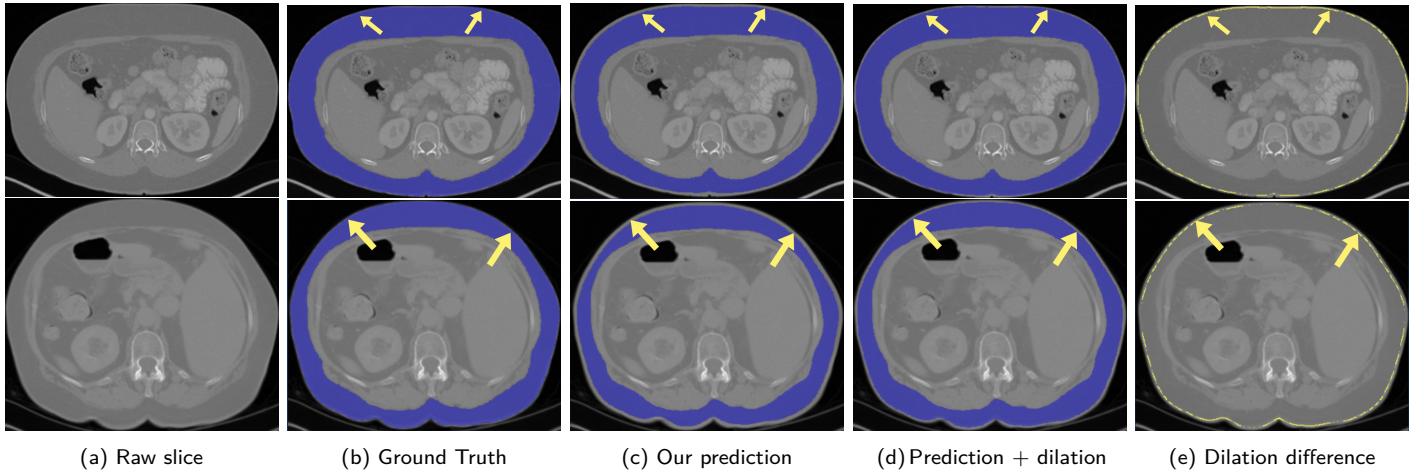


Figure 6: **Dilation examples for two randomly selected slices:** The first column shows raw data without a mask. The second column displays the ground truth from the SAROS dataset. The third column is our original prediction, which excludes skin for SAT. The fourth column demonstrates our prediction after dilation, and the last column illustrates the area added by dilation (in yellow). The yellow arrow highlights the difference introduced by the dilation and blue mask shows SAT.

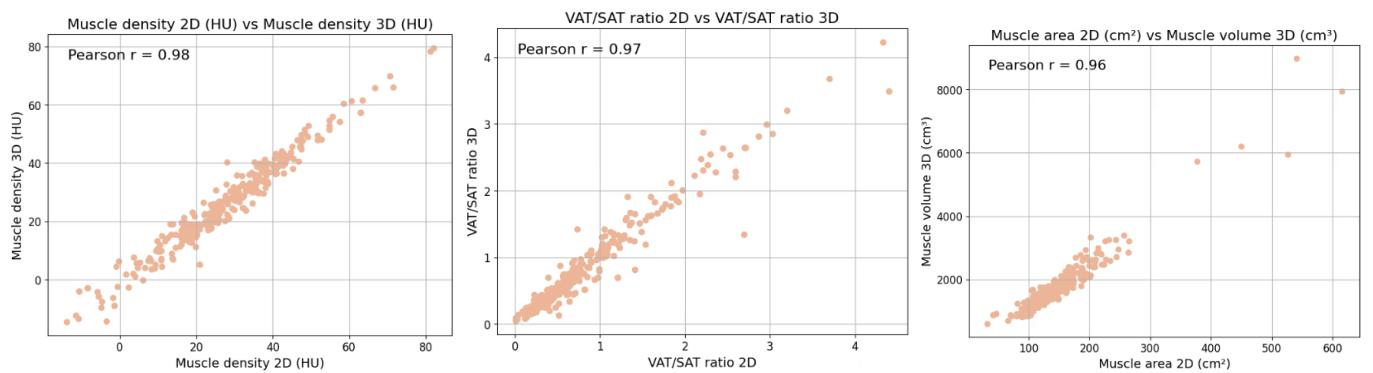


Figure 7: Scatter plots illustrating the relationships between 2D and 3D settings of body composition metrics: Muscle density 2D (HU), VAT/SAT ratio 2D, and Muscle area 2D (cm^2). Each subplot represents a specific metric pair, displaying the distribution of data points alongside the calculated Pearson correlation coefficient (Pearson r) to quantify the strength and direction of their linear relationship.

Collection	Number of studies	Abdomen	Thorax	Whole-body	License
ACRIN-FLT-Breast ^{1,2}	32	0	0	32	CC BY 3.0
ACRIN-NSCLC-FDG-PET ^{3,4}	129	0	78	51	CC BY 3.0
Anti-PD-1_Lung ⁵	12	0	0	12	CC BY 3.0
C4KC-KiTS ^{6,7}	175	175	0	0	CC BY 3.0
CPTAC-CM ⁸	1	0	0	1	CC BY 3.0
CPTAC-LSCC ⁹	3	0	0	3	CC BY 3.0
CPTAC-LUAD ¹⁰	1	0	0	1	CC BY 3.0
CPTAC-PDA ¹¹	8	0	0	8	CC BY 3.0
CPTAC-UCEC ¹²	26	25	0	1	CC BY 3.0
LIDC-IDRI ^{13,14}	133	0	133	0	CC BY 3.0
NSCLC Radiogenomics ^{15,16,17,18}	7	0	0	7	CC BY 3.0
Pancreas CT ^{19,20}	58	58	0	0	CC BY 3.0
Soft-tissue-Sarcoma ^{21,22}	6	0	0	6	CC BY 3.0
TCGA-LIHC ²³	33	32	0	1	CC BY 3.0
TCGA-LUAD ²⁴	2	0	0	2	CC BY 3.0
TCGA-LUSC ²⁵	3	0	0	3	CC BY 3.0
TCGA-STAD ²⁶	2	2	0	0	CC BY 3.0
TCGA-UCEC ²⁷	1	0	0	1	CC BY 3.0
COVID-19-NY-SBU ²⁸	1	0	0	1	CC BY 4.0
Lung-PET-CT-Dx ²⁹	17	0	15	2	CC BY 4.0
In total	650	292	226	132	-

Table 5: Dataset collections from SAROS. References: ¹(Kostakoglu et al., 2015), ²(Kinahan et al., 2017), ³(Machta et al., 2013), ⁴(Kinahan et al., 2019), ⁵(Madhavi et al., 2019), ⁶(Heller et al., 2021), ⁷(Heller et al., 2019), ⁸(National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), 2018a), ⁹(National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), 2018b), ¹⁰(National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), 2018c), ¹¹(CPTAC), ¹²(National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), 2019), ¹³(Armato III et al., 2011), ¹⁴(Armato III et al., 2015), ¹⁵(Napel and Plevritis, 2014), ¹⁶(Bakr et al., 2017), ¹⁷(Bakr et al., 2018), ¹⁸(Gevaert et al., 2012), ¹⁹(Roth et al., 2016), ²⁰(Roth et al., 2015), ²¹(Vallières et al., 2015), ²²(Vallières et al., 2015), ²³(Erickson et al., 2016a), ²⁴(Albertina et al., 2016), ²⁵(Kirk et al., 2016), ²⁶(Lucchesi and Arede, 2016), ²⁷(Erickson et al., 2016b), ²⁸(Saltz et al., 2021), ²⁹(Li et al., 2020).

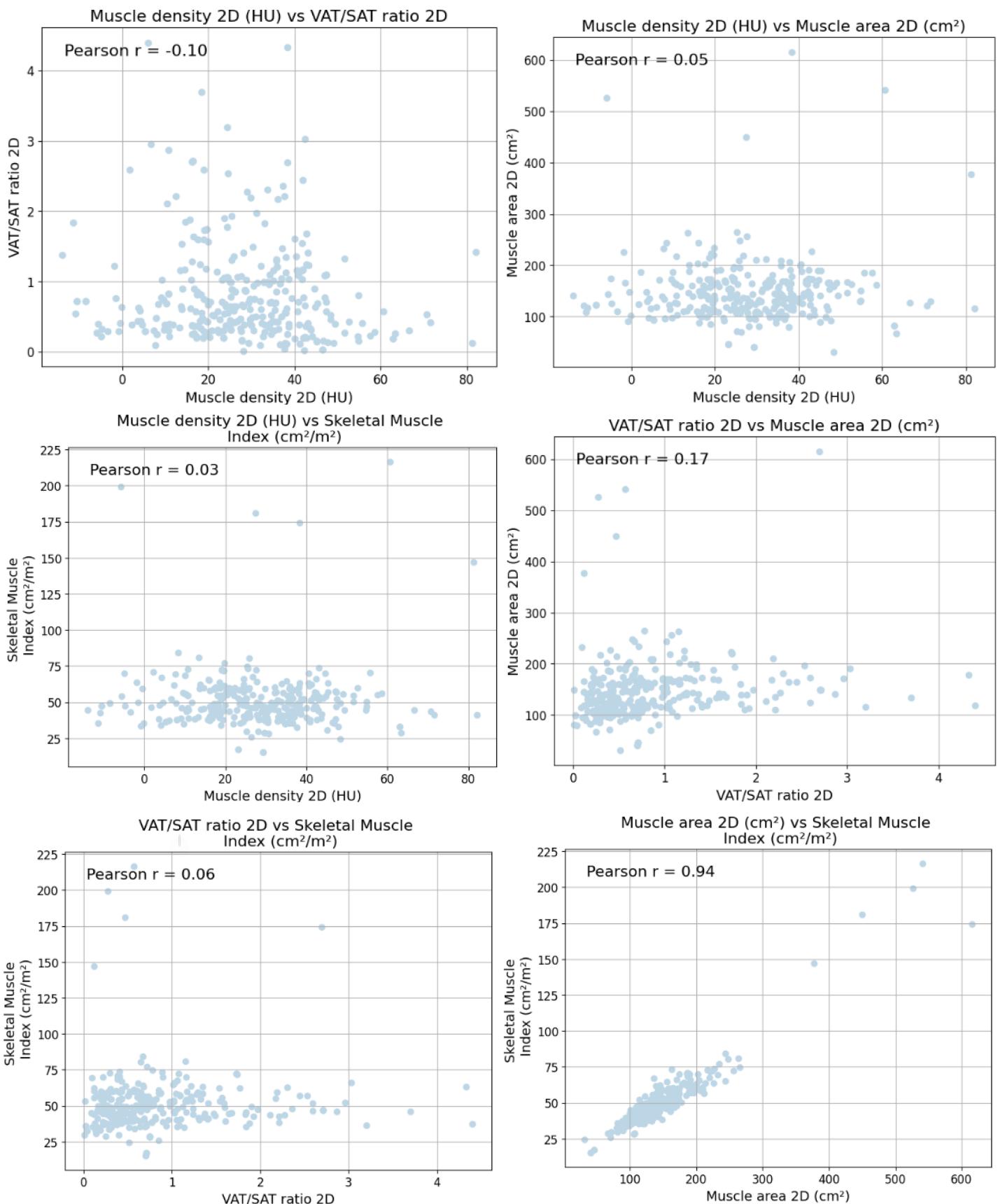


Figure 8: Scatter plots illustrating the relationships between pairs of body composition metrics: Muscle density 2D (HU), VAT/SAT ratio 2D, Muscle area 2D (cm^2), and Skeletal Muscle Index (cm^2/m^2). Each subplot represents a specific metric pair, displaying the distribution of data points alongside the calculated Pearson correlation coefficient (Pearson r) to quantify the strength and direction of their linear relationship.