



# Enhanced muscle and fat segmentation for CT-based body composition analysis: a comparative study

Benjamin Hou<sup>1</sup> · Tejas Sudharshan Mathai<sup>1</sup> · Jianfei Liu<sup>1</sup> · Christopher Parnell<sup>2</sup> · Ronald M. Summers<sup>1</sup>

Received: 10 January 2024 / Accepted: 2 April 2024  
© The Author(s) 2024

## Abstract

**Purpose** Body composition measurements from routine abdominal CT can yield personalized risk assessments for asymptomatic and diseased patients. In particular, attenuation and volume measures of muscle and fat are associated with important clinical outcomes, such as cardiovascular events, fractures, and death. This study evaluates the reliability of an Internal tool for the segmentation of muscle and fat (subcutaneous and visceral) as compared to the well-established public TotalSegmentator tool.

**Methods** We assessed the tools across 900 CT series from the publicly available SAROS dataset, focusing on muscle, subcutaneous fat, and visceral fat. The Dice score was employed to assess accuracy in subcutaneous fat and muscle segmentation. Due to the lack of ground truth segmentations for visceral fat, Cohen's Kappa was utilized to assess segmentation agreement between the tools.

**Results** Our Internal tool achieved a 3% higher Dice (83.8 vs. 80.8) for subcutaneous fat and a 5% improvement (87.6 vs. 83.2) for muscle segmentation, respectively. A Wilcoxon signed-rank test revealed that our results were statistically different with  $p < 0.01$ . For visceral fat, the Cohen's Kappa score of 0.856 indicated near-perfect agreement between the two tools. Our internal tool also showed very strong correlations for muscle volume ( $R^2=0.99$ ), muscle attenuation ( $R^2=0.93$ ), and subcutaneous fat volume ( $R^2=0.99$ ) with a moderate correlation for subcutaneous fat attenuation ( $R^2=0.45$ ).

**Conclusion** Our findings indicated that our Internal tool outperformed TotalSegmentator in measuring subcutaneous fat and muscle. The high Cohen's Kappa score for visceral fat suggests a reliable level of agreement between the two tools. These results demonstrate the potential of our tool in advancing the accuracy of body composition analysis.

**Keywords** Body composition · CT · Segmentation · Muscle · Fat · Subcutaneous · Visceral

## Introduction

The assessment of body composition, particularly the accurate segmentation of soft tissues such as subcutaneous fat, visceral fat, and muscle, has become a critical component in diagnostic imaging [1, 2]. Advances in computed tomography (CT) imaging have not only facilitated detailed body composition analysis, but also play a pivotal role in a range of medical applications, from disease characterization to surgical planning and radiation therapies [3, 4]. This advancement

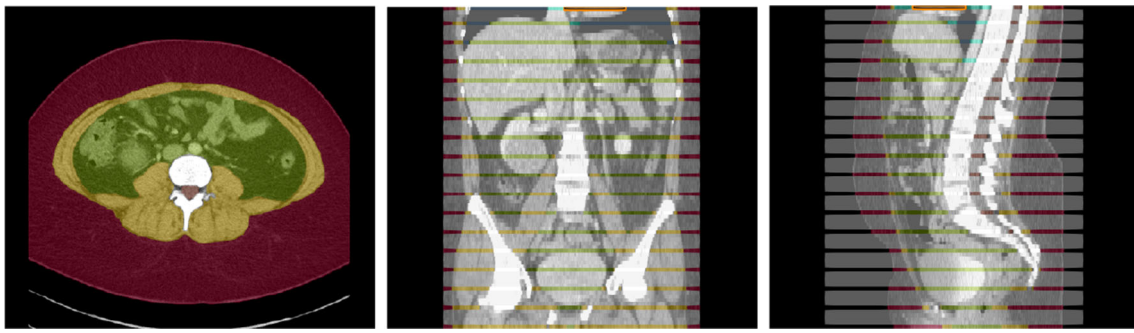
in imaging technology demonstrates potential for enhanced 'incidental' screening and tailored risk evaluation, benefiting both asymptomatic individuals and patients with existing medical conditions. For instance, the distribution and volume of visceral fat are closely linked to metabolic disorders and cardiovascular diseases, making their assessment crucial for early intervention strategies [5]. Similarly, understanding the balance between muscle and fat tissues is essential in evaluating nutritional status, which is particularly relevant in conditions like obesity, sarcopenia, and cachexia [1, 6, 7]. In sports medicine and rehabilitation, analyzing muscle and fat distribution is crucial for creating personalized training and recovery programs [8].

Similarly, in clinical research, such data significantly enhance our understanding of various health conditions and aid in the development of innovative treatments. This knowledge is particularly invaluable in oncology, where it

✉ Benjamin Hou  
benjamin.hou@nih.gov

<sup>1</sup> National Institutes of Health (NIH) Clinical Center, Bethesda, MD, USA

<sup>2</sup> Walter Reed National Military Medical Center, Bethesda, MD, USA



**Fig. 1** Example axial, coronal and sagittal slice of case-042 from SAROS dataset. In the axial slice, the muscle (yellow), subcutaneous fat (red) and abdominal cavity (green) are shown. The gray regions in

the coronal/sagittal views indicate no segmentation masks available in that area, while the streaks in between them contain segmentations

plays a key role in tailoring treatment plans and monitoring the impact of therapies that can significantly alter body composition. Moreover, in surgical planning, especially in reconstructive or plastic surgery, the precise imaging of these tissues is essential for ensuring better outcomes and guiding post-operative care [9–11]. Recent developments in automated segmentation tools, such as TotalSegmentator [12], have shown promising results in enhancing the efficiency and accuracy of these analyses. However, generalized tools in medical imaging, while versatile and broadly applicable, often do not perform with the same level of precision and efficacy as tools that are specifically targeted or tailored to particular tasks or conditions. The effectiveness of such tools compared to specialized solutions remains a subject of ongoing research.

In this study, we compare the effectiveness of the public TotalSegmentator tool against an internally developed tool for the task of muscle and fat (subcutaneous and visceral) segmentation in CT. We hypothesized that the internal tool developed specifically for muscle and fat segmentation would fare better than TotalSegmentator. Through experiments on the public SAROS dataset, we show that the internal tool fares better at the segmentation tasks, with statistical results to corroborate our findings. Our tool has substantial potential to be used for a broad range of clinical applications and offers opportunities for personalized risk assessment for patients.

## Materials and methods

### Patient population

This study utilized deidentified data that are publicly available, thereby obviating the need for IRB approval. The dataset employed, known as the Sparsely Annotated Region and Organ Segmentation (SAROS) [13, 14], comprised of

900 CT series from 882 patients, evenly divided between 450 women and 450 men. These series were randomly selected from various TCIA [14] collections.

The dataset contains CT volumes of 5 mm slice thickness, with annotations provided in NIfTI format. These annotations covers 13 semantic body regions across 6 distinct body parts. The initial generation of annotations was carried out using body composition analysis tools developed by Koitka et al. [15], and subsequently reviewed and corrected by medical residents and students on every fifth axial slice, as illustrated in Fig. 1. Slices that were not reviewed were marked with an ‘ignore’ label of value 255. In this retrospective study, we focused our analysis on three types of soft tissues: subcutaneous fat, visceral fat, and muscle. The SAROS dataset includes annotations for 13 semantic body regions and 6 body parts. However, ground truth segmentation labels within this dataset are only available for subcutaneous fat and muscle. Consequently, our analysis was limited to utilizing only the subcutaneous fat and muscle labels, with all other labels disregarded.

### TotalSegmentator

TotalSegmentator [12] is a publicly accessible tool designed for segmenting over 117 distinct classes in CT images. It is apt for various applications, including organ volumetry, disease characterization, and planning for surgical or radiation therapy. This tool was developed using a training set of 1204 CT examinations, encompassing a diverse array of scanners, institutions, and protocols to ensure its versatility and robustness in different clinical settings. Subcutaneous fat, skeletal muscle, and visceral fat structures fall under a separate task called ‘tissue\_types’, which, while publicly accessible, is subject to a non-commercial license agreement.

## Internal tool

Our internal tool leverages the 3D nnU-Net model [16], which is widely recognized and acclaimed as the *de facto* standard in supervised segmentation. The training data were acquired using a 2D dual-branch network, as described in Liu et al. [17]. This 2D dual-branch network was initially developed to alleviate the extensive and time-consuming annotation burden associated with full CT volumes, enabling the generation of precise segmentations of muscle and fat across all slices of a CT scan.

The dual-branch network features a shared encoder and two duplicate decoders. It was trained using a combination of a few strongly labeled and a large number of weakly labeled datasets; the strongly labeled data included manual annotations of muscle, visceral fat, and subcutaneous fat on each CT slice. The weak labels, generated automatically via a level-set method [18], were prone to segmentation errors. The dual-branch network was trained through a mixed supervision approach utilizing both strong and weak labels. Throughout the training process, the weakly labeled data were periodically refined by the strong decoder in a self-supervised manner. Upon completion of the dual-branch network's training, it was applied to all CT volumes to generate dense annotations across all CT series. These annotations were then utilized as training data for the 3D full-resolution nnU-Net.

## Statistical analysis

As previously mentioned, this retrospective study focuses on three types of soft tissues: subcutaneous fat, visceral fat, and muscle. While both TotalSegmentator (TS) and our Internal tool are capable of segmenting all three tissue types, the SAROS dataset only includes ground truth labels for subcutaneous fat and muscle. After the Internal tool and TotalSegmentator were executed on the CT series in the dataset, the Dice coefficient was utilized to compute the similarity between the predicted segmentations and the ground truth annotations. Since not all slices in the dataset were labeled, Dice score calculation was confined to the "valid" regions of interest, which were delineated by the body mask provided. For all analyses, slices lacking labels, as well as background pixels, were excluded. This approach ensures that our evaluation focused solely on the relevant anatomical areas.

After assessing the normality of the Dice score distribution, a Wilcoxon signed-rank test was employed to determine any statistical differences. Due to the absence of ground truth labels for visceral fat in the dataset, Cohen's Kappa [19] was used to evaluate the agreement between TotalSegmentator and our internal tool in segmenting visceral fat. Cohen's Kappa is a statistical measure that captures the agreement between two raters, taking into account the possibility of

agreement occurring by chance. In addition, graphs correlating the ground truth segmentations contrasted against the predictions were also plotted with overlaid  $R^2$  values. Bland-Altman analysis was also conducted through the calculation of volume differences (biases) and averages for each structure to determine agreement. The Dice and Kappa scores were calculated using the Scikit-learn library (Version 1.3.1) in Python (Version 3.9.10). All statistical tests were performed using RStudio (Version 2023.06.1+524).

## Results

Our study's focus is on comparing the performance of different tools, rather than comparing different scans or patients. Each tool is applied to measure the same scan, with the expectation that the reported volume of tissue types by each tool should be consistent. Should our comparison have been between scans or patients, standardizing the area of measurement would indeed be necessary, such as constraining to the abdomen section (featuring structures L1–L5 and T9–T12) only.

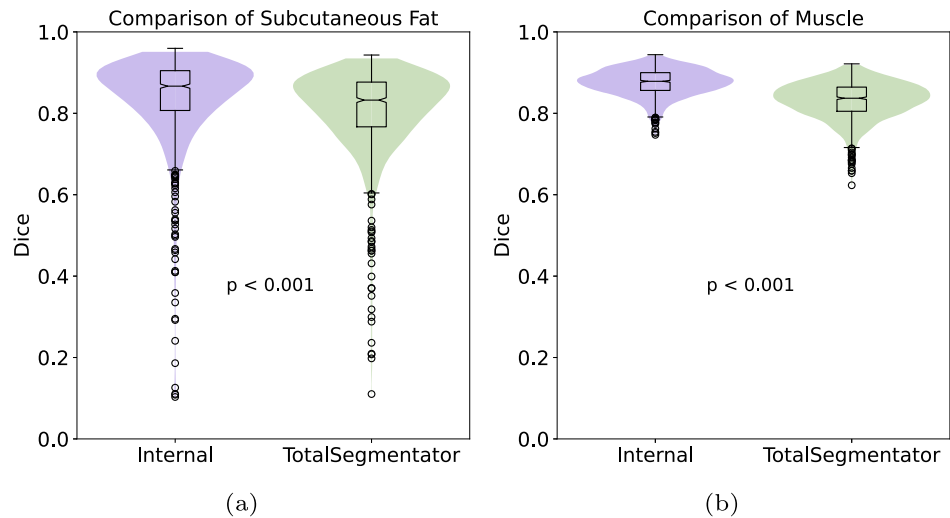
Table 1 presents a direct comparison of the segmentation capabilities of TotalSegmentator and our Internal tool, specifically focusing on subcutaneous fat and muscle segmentation. Figure 2 shows violin plots to visualize the distributions of Dice scores for both TotalSegmentator and Internal tool. Dice scores in Table 1 are presented as means  $\pm$  standard deviation, along with the 25th and 75th percentiles (IQR), for both subcutaneous fat and muscle. For subcutaneous fat, TotalSegmentator achieved an average score of 80.8 ( $\pm$  10.4) with an IQR range of [76.7, 87.7]. In contrast, our Internal tool demonstrated a slightly higher mean Dice score of 83.8 ( $\pm$  10.9) with an IQR range of [80.7, 90.5]. With respect to muscle, TotalSegmentator attained a mean score of 83.2 ( $\pm$  4.6) and [80.5, 86.4] IQR. In contrast, our Internal tool outperformed it by 5% as a mean score of 87.6 ( $\pm$  3.3) and [85.6, 90] IQR was obtained. Notably, as depicted in Fig. 2, the internal tool exhibits fewer outliers compared to TotalSegmentator, particularly in muscle segmentation, indicating a more consistent and reliable performance. These results suggest that while both tools are effective for soft tissue segmentation, the Internal tool was superior in segmenting both subcutaneous fat and muscle with  $p < 0.01$ .

SAROS provides ground truth labels on every fifth axial slice, but these labels are limited to muscle and subcutaneous fat only. Given the absence of labels for visceral fat, the entire CT volume was utilized for comparisons between TotalSegmentator and our Internal tool. It is important to note that subcutaneous fat and visceral fat are considered separate structures and do not overlap. The Kappa scores in Table 2 indicated a high level of concordance between the two tools across all three tissue types. Figure 3 shows  $R^2$  cor-

**Table 1** Table of Dice scores: TotalSegmentator vs. Internal Tool for subcutaneous fat and muscle Segmentation. Scores are shown with mean, standard deviation, and Inter-Quartile Ranges (IQR)

	Subcutaneous Fat	Muscle
TotalSegmentator	80.8 ± 10.4 [76.7, 87.7]	83.2 ± 4.6 [80.5, 86.4]
Internal	83.8 ± 10.9 [80.7, 90.5]	87.6 ± 3.3 [85.6, 90.0]

**Fig. 2** Violin plot of TotalSegmentator (green) vs. our internal tool (blue) for the segmentation of **a** subcutaneous fat and **b** muscle



**Table 2** Cohen's Kappa scores: Agreement of TotalSegmentator and Internal tool for segmentation of subcutaneous fat, visceral fat, and muscle. Scores are shown with mean, standard deviation, and Inter-Quartile Ranges (IQR)

	Cohen's Kappa
Subcutaneous Fat	0.874 ± 0.066 [0.854, 0.913]
Visceral Fat	0.856 ± 0.074 [0.830, 0.906]
Muscle	0.837 ± 0.033 [0.819, 0.861]

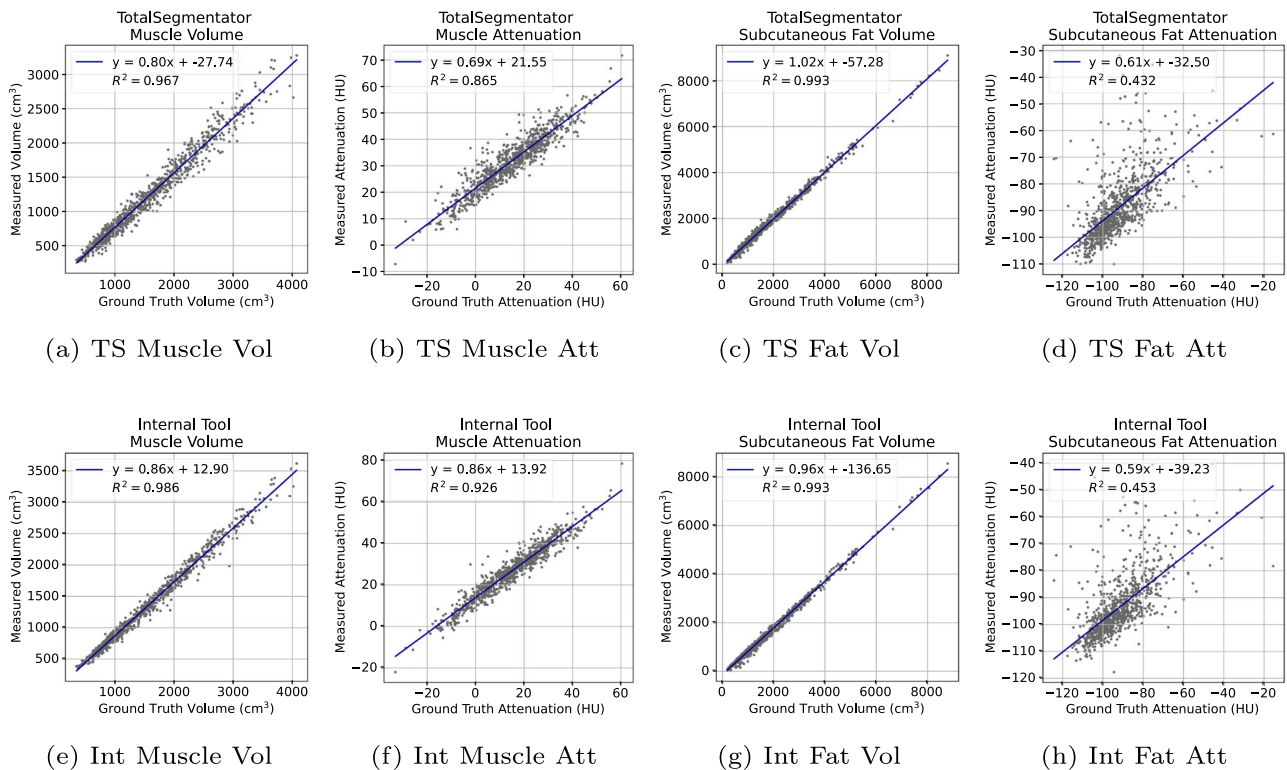
relation plots for the volume and attenuation of the different structures, respectively. The average Hounsfield Unit (HU) of muscle attenuation for both TotalSegmentator and our Internal tool exhibit a strong correlation with  $R^2$  values of 0.87 and 0.93, respectively, with our Internal tool outperforming it by a notable margin. This is supported by the similarly strong correlation observed with muscle volume, yielding  $R^2$  values of 0.97 and 0.99, respectively. For subcutaneous fat, despite a significant uncertainty in the average HU values for both tools, with  $R^2$  values of 0.43 for TotalSegmentator and 0.45 for our Internal tool. Nevertheless, the region was accurately segmented, with fat volume estimation showing a high correlation, evidenced by an  $R^2$  value of 0.99 for both tools.

Figure 4 displays Bland-Altman plots for muscle and subcutaneous fat volume estimation of the tools compared to the manual annotations. For both tools measuring muscle volume, there's a noticeable positive skew in the data. The

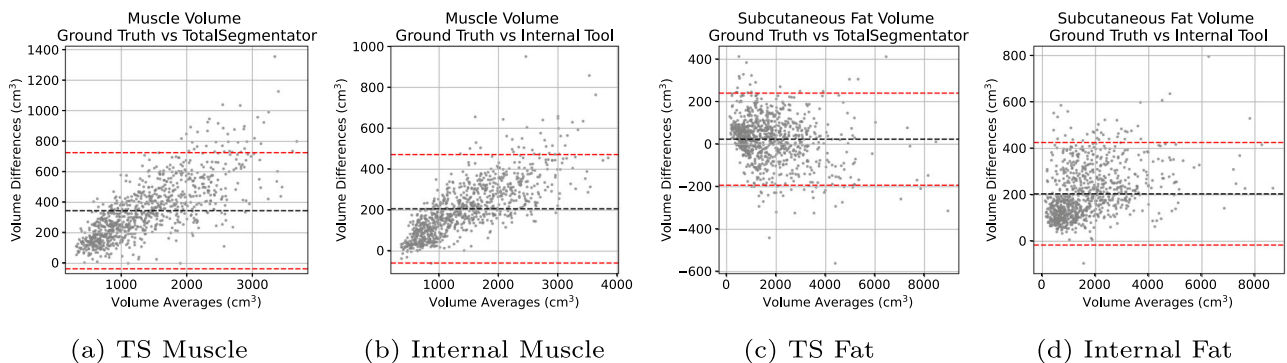
Internal tool demonstrated a significantly lower bias, approximately  $250\text{ cm}^3$ , in comparison to the TotalSegmentator, which exhibited a bias around  $500\text{ cm}^3$ . For the subcutaneous fat volume estimation, there is a distinct concentration of data points on the left-hand side. The Internal tool has a slight positive skew also with a higher bias at around  $+200\text{ cm}^3$  compared to TotalSegmentator that is around  $0\text{ cm}^3$ .

Figure 5 shows an example segmentation of body composition by TotalSegmentator and our Internal tool. In a comparison of segmentation accuracy, our internal tool outperformed TotalSegmentator, achieving Dice scores of 0.947 for Subcutaneous Fat and 0.884 for Muscle, compared to TotalSegmentator's scores of 0.919 and 0.809, respectively. Additionally, our internal tool exhibited a robust Cohen's Kappa score of 0.876, further demonstrating its strong agreement compared to a popular and widely used tool. TotalSegmentator has shown a tendency to over-segment subcutaneous fat, as indicated by the blue arrows in Fig. 5. This is particularly evident in areas such as the muscle between the ribs and within the pelvic cavities.

TotalSegmentator and the Internal Tool demonstrate a high level of segmentation agreement, as evidenced by the Cohen's Kappa scores presented in Table 2. Figure 2 reveals that both tools perform effectively in segmenting muscle tissue, achieving Dice coefficients greater than 0.6, however, this level of performance does not extend to the segmentation of subcutaneous fat. Most instances of segmentation failure (Dice scores < 0.5) occur in patients with a low body



**Fig. 3**  $R^2$  correlation plots of the automatic segmentation results compared against ground truth annotations. Top Row: TotalSegmentator (TS). Bottom Row: Internal (Int) tool. L-to-R: Muscle Volume, Muscle Attenuation, Fat Volume, Fat Attenuation

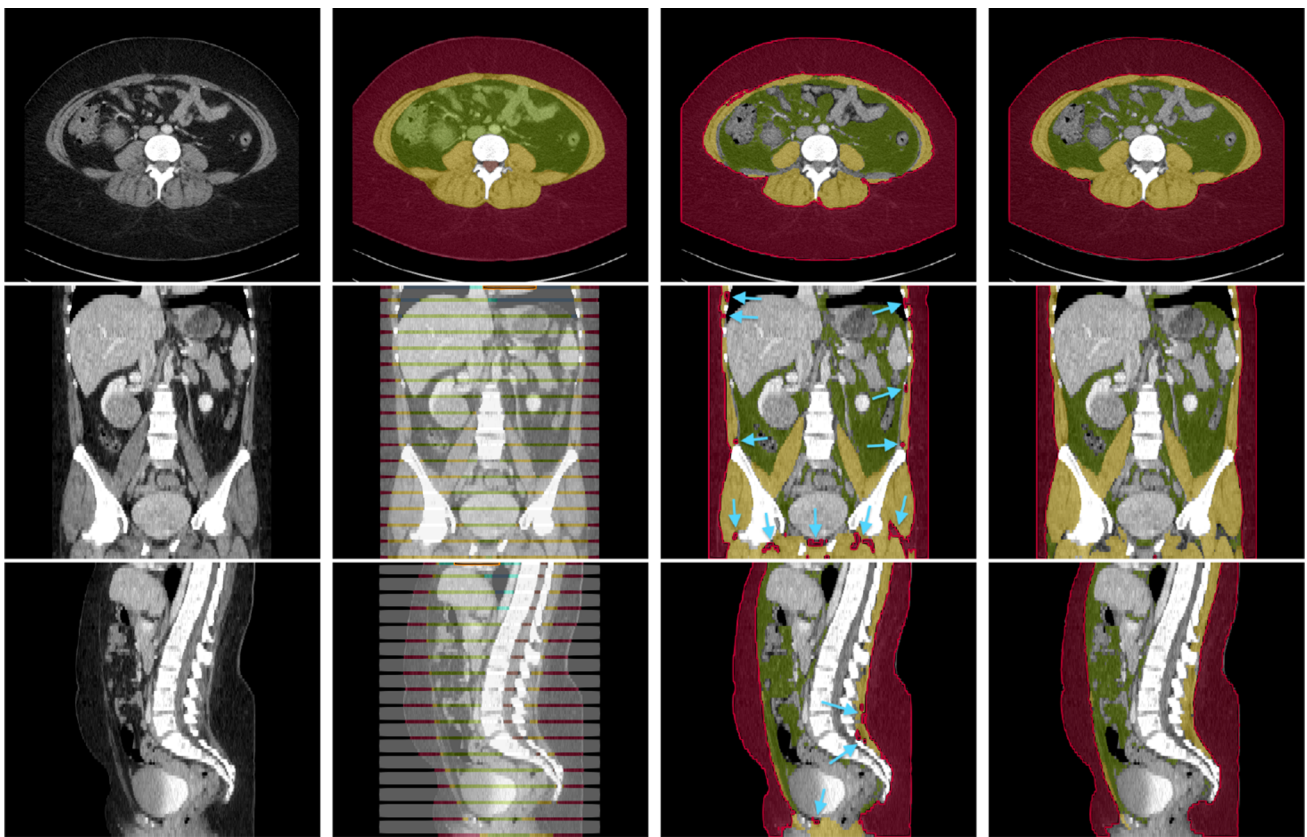


**Fig. 4** Bland-Altman plots of the volume measurements between the automatic segmentations compared against manual annotations. L-to-R: TotalSegmentator Muscle Volume, Internal Muscle Volume, TotalSegmentator Subcutaneous Fat, Internal Subcutaneous Fat

fat percentage. This issue is compounded by the imaging resolution; even at 1 mm, it hinders the clear delineation of subcutaneous fat, which is situated between the skin (dermal layers) and muscle, often covering only a few pixels. The observed low Dice coefficients are attributed to the coarse annotations provided by the annotators, rather than to the segmentation tools themselves as shown in Fig. 6.

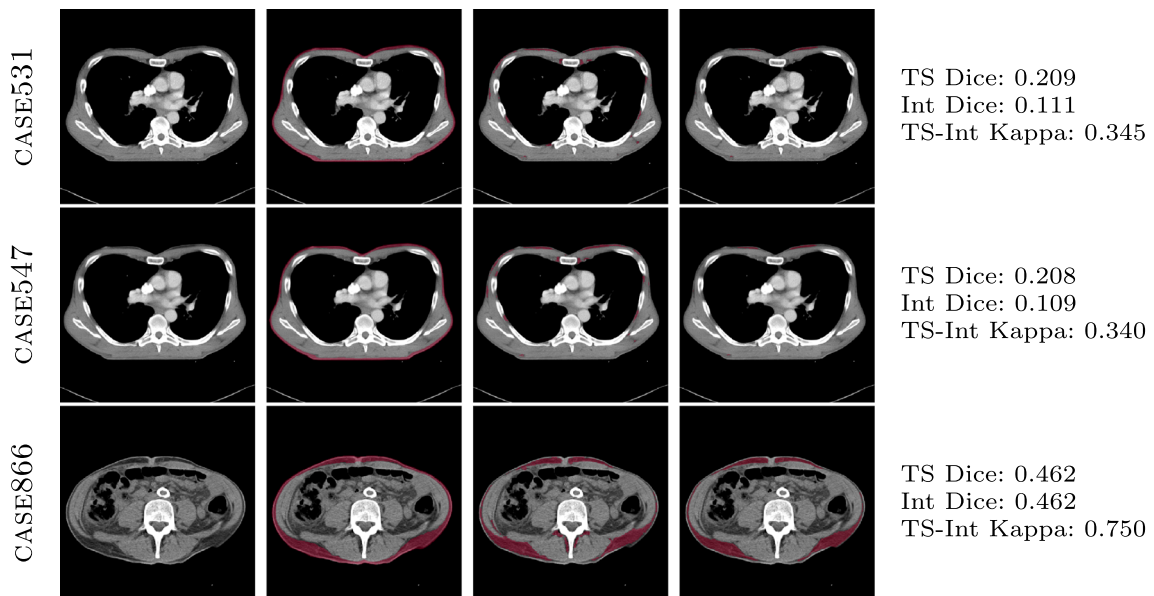
## Discussion and conclusion

Through our experiments, the Internal tool achieved a 3% higher Dice (83.8 vs. 80.8) for subcutaneous fat and a 5% improvement (87.6 vs. 83.2) for muscle segmentation respectively. The results yielded by the internal tool were statistically different  $p < 0.01$ . However, from the  $R^2$  correlation plots in Fig. 3 for subcutaneous fat, a significant uncertainty was seen in the average HU values for both tools: 0.43 for TotalSegmentator and 0.45 for our Internal tool.



**Fig. 5** Example segmentation of case-042. Top-to-Bottom: axial, coronal, sagittal views. L-to-R: CT image, manual annotation (ground truth), TotalSegmentator segmentation, Internal tool segmentation. Red: Subcutaneous Fat, Yellow: Muscle, Green: Internal Abdominal Cavity

(ground truth only) / Visceral Fat, Gray: No ground truth labels. Blue arrows shows over segmentation of subcutaneous fat by TotalSegmentator where it was correctly segmented as muscle by our Internal tool



**Fig. 6** Comparison of subcutaneous fat segmentation failure cases by TotalSegmentator and Internal tool. Top-to-bottom: case-531, case-547, case-886 from SAROS dataset. L-to-R: CT image, manual annotation (ground truth), TotalSegmentator segmentation, Internal tool segmentation

The considerable standard deviation in HU values within the subcutaneous fat layer can be attributed to its diverse composition. This layer, primarily composed of adipocytes, also contains fibroblasts, blood vessels, nerve cells, lymphatic vessels, immune cells, hair follicles, and sweat glands, each with differing densities. These varying densities result in a broad spectrum of HU values, as captured in CT scans. The contrast between the low-density adipocytes and the higher-density components within the layer leads to the observed variability in HU readings.

The variability can also be attributed to several other factors: the quality and noise in CT images affecting segmentation precision, limitations in the segmentation algorithm especially if not tailored for subcutaneous fat, variability in fat composition and density, and the choice of thresholding in segmentation. This complexity not only highlights the multifaceted nature of the subcutaneous layer, but also underscores the challenge in accurately segmenting and analyzing it using CT imaging. Despite the variations in HU, the subcutaneous fat volume demonstrated a high correlation for both tools with an  $R^2$  value of 0.99, indicating accurate segmentation of the subcutaneous fat region.

The skewness in the Bland-Altman plots in Fig. 4 suggests a tendency for the differences between the two methods under comparison to increase as the magnitude of the measurement decreases. Such a distribution pattern indicates a potential systematic bias in the measurements, particularly at lower values. For the concentration of data points on the left-hand side in the subcutaneous fat volume estimation, the pattern indicates that the agreement between the two methods being compared is more consistent at lower measurement values. Such a concentration suggests that for smaller magnitudes of the variable being measured, the two methods yield closer results, implying better concordance in this range. However, this also raises questions about the performance of the methods at higher values, as the relative sparsity of data points on the right-hand side of the plot may indicate a divergence in the methods' readings or a limitation in the range of data sampled.

Furthermore, segmenting muscle tissue is a relatively easier task due to its clearly defined visual boundaries. In contrast, the delineation of fat can be challenging, as its boundaries are not always distinct. This challenge stems from the fact that fat and water-rich tissues (such as specific soft tissues) can exhibit similar Hounsfield Units (HUs), complicating their differentiation. Fat typically has a slightly negative HU value, often in the range of -50 to -100 HU, whereas water has an HU of 0. However, the HU values of soft tissues can range from -10 to +60 HU, depending on the specific tissue type and its water content.

The overlapping HU values between fat and certain soft tissues create a significant challenge for differentiation based solely on attenuation properties. This is particularly true for

visceral fat, where the close proximity and interleaving of blood vessels, bowel, and organs give it a complex shape. Although fat and muscle have distinct HU values, the HU values of the bowel, vessels, and organs may closely resemble those of muscle, especially in non-contrast CT scans, or when the CT scan's resolution is too low to clearly differentiate between these tissue types. Furthermore, fat deposits can be located within muscle tissue, indicating that HU values are not the primary reason for the segmentation difficulty for visceral fat.

In summary, this study has demonstrated that our internal tool significantly outperforms the more generalized TotalSegmentator in accurately segmenting subcutaneous fat, visceral fat, and muscle in CT series. Our findings are supported by high Dice scores and strong correlations ( $R^2$ ) with manual annotations, and is further corroborated by Bland-Altman plots demonstrating consistent agreement and minimal bias. The enhanced accuracy and consistency of our internal tool hold significant promise for a range of clinical applications, such as providing improved personalized risk assessments for patients at risk of adverse cardiovascular events and fractures.

**Funding** Open access funding provided by the National Institutes of Health. This work was supported by the Intramural Research Program of the NIH Clinical Center (project number 1Z01 CL040004).

## Declarations

**Conflict of interest** RMS receives royalties from iCAD, MGB, Philips, PingAn, ScanMed, and Translation Holdings. His lab received research support from PingAn. The authors have no additional Conflict of interest to declare.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. For this study, informed consent was not required.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Lee MH, Liu D, Garrett JW, Perez A, Zea R, Summers RM, Pickhardt PJ (2023) Comparing fully automated AI body composition measures derived from thin and thick slice CT image data. *Abdom Radiol* 49(3):985–996
- Weston AD, Korfiatis P, Kline TL, Philbrick KA, Kostandy P, Sakinis T, Sugimoto M, Takahashi N, Erickson BJ (2019) Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* 290(3):669–679 (PMID: 30526356)
- Makrogiannis S, Okorie A, Di Iorio A, Bandinelli S, Ferrucci L (2022) Multi-atlas segmentation and quantification of muscle, bone and subcutaneous adipose tissue in the lower leg using peripheral quantitative computed tomography. *Front Physiol* 13:951368
- Shah UA, Ballinger TJ, Bhandari R, Dieli-Conwright CM, Guertin KA, Hibler EA, Kalam F, Lohmann AE, Ippolito JE (2023) Imaging modalities for measuring body composition in patients with cancer: opportunities and challenges. *JNCI Monogr* 61:56–67
- Gruzdeva O, Borodkina D, Uchasova E, Dyleva Y, Barbarash O (2018) Localization of fat depots and cardiovascular risk. *Lipids Health Dis* 17(1):1–9
- Holmes CJ, Racette SB (2021) The utility of body composition assessment in nutrition and clinical practice: an overview of current methodology. *Nutrients* 13(8):2493
- Ozen E, Mihaylova R, Weech M, Kinsella S, Lovegrove JA, Jackson KG (2022) Association between dietary saturated fat with cardiovascular disease risk markers and body composition in healthy adults: findings from the cross-sectional bodycon study. *Nutr Metab* 19(1):1–15
- Wackerhage H, Schoenfeld BJ (2021) Personalized, evidence-informed training plans and exercise prescriptions for performance, fitness and health. *Sports Med* 51(9):1805–1813
- Perrin T, Lenfant M, Boisson C, Bert M, Rat P, Facy O (2021) Effects of body composition profiles on oncological outcomes and postoperative intraabdominal infection following colorectal cancer surgery. *Surg Obes Relat Dis* 17(3):575–584
- Wopat H, Harrod T, Brem RF, Kaltman R, Anderson K, Robien K (2023) Body composition and chemotherapy toxicity among women treated for breast cancer: a systematic review. *J Cancer Surviv*. <https://doi.org/10.1007/s11764-023-01380-7>
- Aleixo GFP, Valente SA, Wei W, Moore HC (2023) Association of body composition and surgical outcomes in patients with early-stage breast cancer. *Breast Cancer Res Treat* 202(2):305–311
- Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, Heye T, Boll DT, Cyriac J, Yang S, Bach M, Segeroth M (2023) TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell* 5(5):230024
- Koitka S, Baldini G, Kroll L, Landeghem N van, Haubold J, Kim M Sung, Kleesiek J, Nensa F, Hosch R (2023) SAROS-A large, heterogeneous, and sparsely annotated segmentation dataset on CT imaging data. *Cancer Imaging Arch*
- Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F (2013) The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 26:1045–1057
- Koitka S, Kroll L, Malamutmann E, Oezcelik A, Nensa F (2020) Fully-automated body composition analysis in routine CT imaging using 3d semantic segmentation convolutional neural networks. *CoRR*. [arXiv:2002.10776](https://arxiv.org/abs/2002.10776)
- Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 18(2):203–211
- Liu J, Shafaat O, Summers RM (2023) A dual-branch network with mixed and self-supervision for medical image segmentation: an application to segment edematous adipose tissue. In: *MILLanD@MICCAI*, vol. 14307 of *Lecture Notes in Computer Science*, pp 158–167
- Burns JE, Yao J, Chen JJ, Chalhoub D, Summers RM (2020) A machine learning algorithm to estimate sarcopenia on abdominal CT. *Acad Radiol* 27(3):311–320
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Measur* 20(1):37–46

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.