Published in final edited form as:

Clin Nutr. 2020 October ; 39(10): 3049–3055. doi:10.1016/j.clnu.2020.01.008.

Automated body composition analysis of clinically acquired computed tomography scans using neural networks

Michael T. Paris, MSc¹, Puneeta Tandon, MD², Daren K. Heyland, MSc, MD^{3,4}, Helena Furberg, MSPH, PhD⁵, Tahira Premji, DDS¹, Gavin Low, MD⁶, Marina Mourtzakis, PhD¹

¹Department of Kinesiology, University of Waterloo, Waterloo, ON, Canada.

²Department of Gastroenterology, University of Alberta, Edmonton, AB, Canada

³Department of Critical Care, Kingston General Hospital, Kingston, ON, Canada

⁴Clinical Evaluation Research Unit, Queens University, Kingston, ON, Canada

⁵Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁶Department of Radiology, University of Alberta, Edmonton, AB, Canada

Abstract

Background & Aims—The quantity and quality of skeletal muscle and adipose tissue is an important prognostic factor for clinical outcomes across several illnesses. Clinically acquired computed tomography (CT) scans are commonly used for quantification of body composition, but manual analysis is laborious and costly. The primary aim of this study was to develop an automated body composition analysis framework using CT scans.

Methods—CT scans of the 3rd lumbar vertebrae from critically ill, liver cirrhosis, pancreatic cancer, and clear cell renal cell carcinoma patients, as well as renal and liver donors, were manually analyzed for body composition. Ninety percent of scans were used for developing and validating a neural network for the automated segmentation of skeletal muscle and adipose tissues. Network accuracy was evaluated with the remaining 10 percent of scans using the Dice similarity coefficient (DSC), which quantifies the overlap (0 = no overlap, 1 = perfect overlap) between human and automated segmentations.

Results—Of the 893 patients, 44% were female, with a mean ($\pm$ SD) age and body mass index of 52.7 ($\pm$ 15.8) years old and 28.0 ($\pm$ 6.1) kg/m², respectively. In the testing cohort (n=89), DSC scores indicated excellent agreement between human and network-predicted segmentations for

---

Corresponding author: Marina Mourtzakis, PhD, Department of Kinesiology, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada. mmourtzakis@uwaterloo.ca. Phone: 519-888-4567, Ext 38459.

STATEMENT OF AUTHORSHIP

MTP and MM conceptualized and designed the study, DKH, PT, HF, and DL were responsible for data acquisition, MTP, TP, and MM analyzed and interpreted the data. MTP and MM drafted the initial manuscript. All authors critically revised and gave approval for the final manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

skeletal muscle (0.983 ± 0.013), and intermuscular (0.900 ± 0.034), visceral (0.979 ± 0.019), and subcutaneous (0.986 ± 0.016) adipose tissue. Network segmentation took ~350 milliseconds/scan using modern computing hardware.

Conclusions—Our network displayed excellent ability to analyze diverse body composition phenotypes and clinical cohorts, which will create feasible opportunities to advance our capacity to predict health outcomes in clinical populations.

Keywords

Automated body composition analysis; sarcopenia; computed tomography

INTRODUCTION

Deleterious features of body composition, such as low skeletal muscle mass or elevated visceral adipose tissue, are associated with markedly increased risk of morbidity and mortality in critically ill, cancer, and liver cirrhosis patients [1–9]. Our need to understand body composition beyond the use of body mass index (BMI) is evident in the exponential developments of using clinically accessible modalities such as computed tomography (CT) and ultrasound to better understand skeletal muscle and adipose characteristics [10,11]. Much of the advances in our understanding of body composition in clinical populations has been realized by leveraging CT scans, which have been performed for diagnostic and other clinical purposes, for precise quantification of skeletal muscles, intermuscular adipose tissue (IMAT), visceral adipose tissue (VAT), and subcutaneous adipose tissue (SAT) [10–13]. Typically, CT scans of the 3rd lumbar vertebrae are analyzed for muscle and adipose tissue cross-sectional area (CSA) because these metrics are strongly associated with whole body measures of body composition [13,14]. While these methods have been foundational in our understanding of the multifaceted relationships between body composition and outcomes in clinical populations, current approaches for analysis are impractical for large-scale studies or clinical practice because they are cost- and time-prohibitive.

Development of an automated CT body composition framework would allow a high throughput analysis of body composition and enable larger-scale, in-depth investigations on the prognostic impact of body composition in clinical populations. Several recent publications have developed neural networks for automating the segmentation of skeletal muscle, VAT, and SAT from CT scans [15–22]; however, to best of our knowledge, none of the currently proposed approaches segment IMAT separately from VAT or SAT, which is a significant predictor of mortality in clinical populations [23]. Furthermore, these networks have not been made publicly available, limiting their widespread implementation.

Here, we address the need for developing an accessible approach that reduces the time- and cost-burden of CT body composition analysis. We developed and validated a deep, convolutional neural network for the automated segmentation of skeletal muscle, IMAT, VAT, and SAT CSA of 3rd lumbar vertebrae CT scans. Our network was trained on a large, diverse cohort of intensive care unit (ICU), clear cell renal cell cancer (ccRCC), pancreatic cancer, and liver cirrhosis patients, as well as liver and kidney donors. This multi-morbid cohort was intended to capture the heterogeneity of body composition phenotypes for

greater generalizability. We have termed our approach Automated Muscle and Adipose Tissue Composition Analysis (AutoMATiCA). We have made AutoMATiCA freely available for all to use and have included an easy to use graphical user interface for loading scans, performing analysis, and exporting results: <https://gitlab.com/MichaelParis/AutoMATiCA>

MATERIALS AND METHODS

Data set

Clinically acquired CT scans of the 3rd lumbar vertebrae were retrospectively analyzed for body composition and included in this study. Scans included in this work have been analyzed at the University of Waterloo in collaboration with clinical centres and hospitals from Canada, United States, France, and the Netherlands. The data set includes CT scans from renal and liver donors [24], as well as critically ill [25–27], liver cirrhosis (unpublished), pancreatic cancer [28], and ccRCC patients [29], providing a diverse clinical cohort and range of body composition phenotypes. Scans of several different phases of intravenous (IV) contrast, tube currents, and tube voltages were included in this analysis. Presence of IV contrast was identified visually by a single trained analyst [29]. All research was approved by local and institutional research ethics boards and conducted in accordance with established protocols for human research.

Computed tomography body composition analysis

CT scans of the 3rd lumbar vertebrae were identified from a series of abdominal images and manually segmented by trained analysts at the University of Waterloo using SliceOmatic (TomoVision, Montreal, Canada, version 4.2, 4.3, and 5.0), as previously described [13]. This software utilizes Hounsfield unit (HU) thresholds to aid analysts in segmenting skeletal muscles (–29 to 150 HU), IMAT (–190 to –30 HU), VAT (–150 to –50 HU), and SAT (–190 to –30 HU) using a brush tool. Tissue CSA was calculated by summing pixels of a specific compartment and multiplying by the pixel surface area. Tissue densities were calculated as the mean radiological attenuation, measured in HU, of pixels identified as muscle, IMAT, VAT, and SAT.

Computed tomography scan preparation for neural network training

Greyscale CT scans were processed into a dual-channel image for network training. The dual-channel CT image consists of a standard CT image (windowed between –1024 to +1024 HU and scaled to 0 and 1) and a filtered CT image that removed HU values outside the threshold ranges (–29 to +150 for lean tissues, –190 to –30 for IMAT and SAT, and –150 to –50 for VAT, scaled to 0 and 1) (Figure 1). A dual-channel image was used as input because the filtered image may emphasize structural features within predefined HU ranges, whereas the raw CT image provides the original pixel information [30]. Human-analyzed tissue segmentation maps (known as ground truth labels) were reformatted into acceptable input for neural network training. Full resolution (512 × 512 pixels) dual-channel images and ground truth segmentation maps were used as input. Higher resolution scans (i.e. greater than 512 × 512 pixels) were down sampled to 512 × 512 pixels.

Dataset division for training, validation, and testing

The data set was divided so that 80% of scans were used for training the network (training cohort), 10% for validation during training (validation cohort), and 10% for testing final network accuracy (testing cohort). Previous work has demonstrated that obesity (based on BMI) may negatively impact the accuracy of skeletal muscle segmentation at the 3rd lumbar vertebrae using convolutional neural networks [17]. To ensure the network training and testing is balanced for obesity, patients were stratified by total adipose tissue (IMAT + VAT + SAT CSA) across the training, validation, and testing cohorts. Because the mean total adipose tissue was different across the different groups, cohort specific z-scores were calculated and used for stratification; this ensures each cohort is evenly distributed across the training, validation, and testing data sets.

Network architecture, training, and evaluation metrics

A neural network's architecture refers to the organization of how data is processed through the network. Several architectures have been proposed for different classifications, with certain frameworks having been optimized for specific tasks (e.g. image segmentation). Here, we chose to adapt UNET for our automated approach, as it has demonstrated excellent accuracy for biomedical image segmentation [31]. Details of the network architecture, data augmentation, and training can be found in supplemental materials (Supplemental Methods). The Dice similarity coefficient (DSC) was used for quantifying the overlap between ground truth (human-analyzed) and network predicted segmentations. A DSC score of 1 indicates perfect agreement, whereas 0 indicates no overlap.

Network development and training was performed in Python 3.6.0 (Python Software Foundation, Wilmington, Delaware, United States) using the Keras API for Tensorflow (version 1.12, Google, Mountain View, California, United States). All experiments were run on a single GPU (Nvidia RTX 2070, Santa Clara, California, United States) machine with 16 GB of RAM.

Post processing

The four neural networks (muscle, IMAT, VAT, and SAT) each output a probability segmentation map, which provides a probability score (0 to 1 for each pixel) for whether a pixel belongs to the tissue being analyzed or is part of the background (Figure 1). Since tissue segmentation maps are predicted separately for each network, post processing is necessary to combine segmentation maps. Post processing involves two stages: 1) removing incorrectly segmented pixels using threshold ranges (−29 to +150 for muscle, −190 to −30 for IMAT and SAT, and −150 to −50 for VAT), and 2) combining the probability maps of individual networks into a single segmentation map. Probability maps of adipose tissues (IMAT, VAT, and SAT) were compared for each pixel location, in which the highest probability score is used to label that pixel. Since skeletal muscle HU range does not overlap with adipose tissues, these maps were combined without any comparisons. A threshold greater than 0.1 for all probability maps (0 = background, tissue = 1) was used for final pixel assignment.

Statistical analysis

Student's t-test were used to evaluate differences between training/validation and testing cohorts for demographic, physical, and body composition features. DSC scores were compared across age (young:<40 years of age, middle:40–65 years of age, older:>65 years of age), sex, BMI (low and normal: <25.0 kg/m², overweight: 25–29.9 kg/m², obese: 30.0 kg/m²), total adipose tissue (cohort specific tertiles), presence of IV contrast, and clinical cohort (ICU, cancer, liver cirrhosis, donors) using a one-way ANOVA or Student's t-test. Post-hoc analysis was corrected for multiple comparisons using Tukey's honest significance difference test.

Absolute differences between network and human-analyzed muscle, IMAT, VAT, and SAT CSA and density were evaluated using Bland-Altman plots. Interpretation of the limits of agreement (95% confidence interval) for CSA were made in comparison to limits of agreement derived from inter-rater reliability. Inter-rater reliability was performed between two experienced raters on a subset of scan (n=99) used in the analysis. Proportional bias and homoscedasticity of the differences were evaluated using correlation analysis and visual analysis of the spread of residuals against the average, respectively [32]. Classification of low skeletal muscle mass was made using previously established cutpoints for skeletal muscle CSA normalized to height squared (kg/m²) of <38.9 kg/m² for women and <55.4 kg/m² men [13]. Agreement between network and human identified low skeletal muscle CSA was compared using Cohens Kappa.

All statistical analyses were performed using custom Python scripts (Python Software Foundation, version 3.6.0, Wilmington, Delaware, United States). Statistical significance was set a $p < 0.05$.

RESULTS

In total, 893 patients (44% female) were included in this analysis: 300 critically ill, 110 liver cirrhosis, 181 ccRCC, 49 pancreatic cancer, and 253 donors. There were no differences in demographic or body composition characteristics between the training/validation (n=804) and testing (n=89) cohorts. However, CT tube current was significantly higher in the testing cohort (368.4 ± 125.7 mA) compared with the training/validation group (338.3 ± 122.6 mA, $p = 0.042$) (Table 1).

Within the testing cohort, age, sex, BMI, total adiposity, and presence of contrast agent had small ($DSC < 0.04$), but statistically significant influences on DSC scores (Supplemental Tables 1–5). Similarly, Small ($DSC < 0.04$) differences were also observed in DSC scores across clinical cohorts for skeletal muscle, IMAT, VAT, and SAT (Table 2). Examples of good and poorly automated segmentation are presented to visualize performance compared with human analysis (Figure 2). Segmentation of muscle, IMAT, VAT, and SAT took approximately 350 milliseconds using modern computing hardware.

Across all clinical populations in the testing cohort, Bland-Altman plots comparing network and human-analyzed CSA for skeletal muscle, IMAT, VAT, and SAT displayed non-significant biases [limits of agreement] of $2.5 [-3.3 \text{ and } +8.2] \text{ cm}^2$, $-0.2 [-4.5 \text{ and } 4.0] \text{ cm}^2$,

0.8 [−11.3 and 12.9] cm², and 2.2 [−5.6 and 9.9] cm², respectively (Figure 3). Bland-Altman plots comparing network and human-analyzed density for skeletal muscle, IMAT, VAT, and SAT, displayed non-significant biases [limits of agreement] of 0.01 [−1.61 and 1.63] HU, 0.95 [−2.02 and 3.94] HU, 0.65 [−2.75 and 4.05] HU, and 0.41 [−0.79 and 1.61] HU, respectively (Supplemental figure 2). To compare the absolute error of network predicted CSA with that of two experienced raters, we performed inter-rater reliability analysis using Bland-Altman plots in a subset of patients (n=99). Inter-rater reliability Bland-Altman plots displayed muscle, IMAT, SAT, and VAT biases [limits of agreement] of −0.1 [−8.6 and 8.6] cm², 0.6 [−4.4 and 5.6] cm², 0.0 [−11.6 and 11.6] cm², and −1.7 [−18.5 and 15.1] cm², respectively (Supplemental Figure 3).

There was strong agreement ($\kappa=0.89$) between network and human analyzed identification of low skeletal muscle CSA (Table 3). After correcting for the systematic bias observed in the Bland-Altman plot for skeletal muscle (difference of 2.5 cm²), agreement improved ($\kappa=0.95$) for identification of low skeletal muscle CSA.

DISCUSSION

Prior studies suggest that amongst cancer, liver cirrhosis, and critically ill patients, the presence of deleterious body composition features, such as low skeletal muscle mass or elevated visceral adiposity, negatively impact several clinical endpoints, including length of stay [3,4], rates of post-surgical complications [33–35], rates of infection [36], rates of re-admission [37], response to treatment [38], hospital costs [37,39], and overall survival [2,12,40]. These deleterious body composition features have wide-spread implications, as we and others have shown that these features are highly prevalent amidst clinical cohorts (40–70% of patients display low skeletal muscle mass, despite >50% being classified as overweight or obese based on BMI) [10]. Few studies have leveraged large repositories of CT scans from clinical populations for body composition analysis (e.g. n=5786 breast and colorectal cancer scans [1], estimated analysis time: 1500 – 2000 hours). More often, studies are of limited sample sizes because of the time consuming and expensive body composition analysis of CT scans. There is an urgent need for an open-source framework to automate the analysis of these body composition features, enabling clinicians and researchers to leverage the vast repository of clinically available CT scans (~62 million CT scans in the United States per year) for body composition analysis.

Here, we developed and validated a deep convolutional neural network for the automated segmentation of skeletal muscle, IMAT, VAT, and SAT at the 3rd lumbar vertebrae of CT scans. Across a range of clinical cohorts and body composition phenotypes, our network provides excellent (DSC >0.97) segmentation accuracy for muscle, VAT, and SAT, and good accuracy (DSC >0.88) for IMAT. While the accuracy of segmentation was heterogeneously influenced by age, BMI, sex, total adiposity, presence of contrast, and clinical cohort, these differences were small (DSC <0.04) and likely to be clinically insignificant for CSA analysis.

Several approaches have been proposed for the automated segmentation of muscle, VAT, and SAT within the 3rd lumbar vertebrae region; which have demonstrated segmentation

performance ranging from DSC: 0.85 – 0.99 [15,17,19–21,41–44] (Table 4). AutoMATiCA displays DSC scores on-par with the most accurate automated approaches for muscle, VAT, and SAT. However, our approach maintains accuracy (DSC >0.04) across a diverse cohort of different clinical populations, suggesting that it may be generalizable to other populations for body composition analysis. Furthermore, to the best of our knowledge, AutoMATiCA is the only automated approach to segment IMAT separately from VAT or SAT on 3rd lumbar vertebrae scans, which is an important tissue for understanding impairments in strength, physical function, metabolic control, and clinical outcomes. Lastly, we have made our approach open-source and freely available for all to use, included with an easy to use interface for loading scans, performing segmentation, and saving results. However, because our approach utilizes a combination of 4 separate neural networks (muscle, IMAT, VAT, and SAT), analysis time for a single scan is higher compared to other neural network approaches (350 ms vs 150 ms – hardware dependent); though, we believe this is still an acceptable speed.

While DSC provides an excellent metric to evaluate the accuracy of segmentation compared to human analysis, final interpretation of the results is based on absolute CSA. We compared our network CSA prediction error to the error observed between two well-trained human analysts. Bland-Altman plots comparing the network and human analyzed CSA revealed limits of agreement similar to human inter-rater reliability; indicating that the network-predicted CSA error is similar to what can be expected between two experienced raters. However, it should be noted that while similar, network CSA predictions for skeletal muscle displayed a small bias; which has been observed by others [17].

Several cutpoints have been proposed for identification of low skeletal muscle CSA based on CT imaging. Despite the small bias for network-predicted skeletal muscle CSA, there was strong agreement with human-analyzed identification of low muscle CSA ($\kappa=0.89$). However, given that >95% (85/89) of the network predicted muscle CSA were larger than the human-analyzed, the systematic bias can be easily corrected; which improved agreement for identification of low muscle CSA ($\kappa=0.95$). The accurate identification of low skeletal muscle CSA is a critical aspect, as this metric may be useful in risk classification of patients. For example, it is well established that low muscle CSA negatively influences clinical outcomes following surgical interventions [45]. This metric may be useful in combination with other clinical markers for a risk-score classification, to identify patients that may require pre-habilitation or nutritional support prior to a surgical intervention.

Limitations

Although AutoMATiCA has demonstrated accurate and rapid body composition analysis of CT scans, there are several limitations. During human analysis of CT images, a proportion of scans will be categorized as non-analyzable because of their poor image quality (e.g. inability to distinguish adipose from lean tissues). Since these scans are unable to be included during the training or testing phases, segmentation performance on these scans is unknown. Caution is necessary when applying our approach on these scans, which may require visual confirmation of the segmentation accuracy by trained analysts. However, a significant proportion of scans that are considered challenging to analyze, based on our 15

years of experience, have been included within the training and testing cohorts (Figure 2); these challenging scans may aid the network in developing better segmentation results for scans with increasing magnitude of difficulty. Due to limitations on the availability of training data for body composition segmentation, we were unable to use a large sample of scans for testing final network performance. While we achieved similar results between our training, validation, and testing datasets, an external validation on an independent dataset will be necessary to confirm network accuracy. Although the network automatically segments CT scans of the 3rd lumbar vertebrae, our approach is not fully automated, as the manual identification of the correct scans for analysis by trained individuals is required. Lastly, this network was trained using a diverse geographical cohort of patients; however, the majority of populations from these regions have underlying ethnic diversities that may lead to biases during training. This may be particularly evident amongst Southeast Asian populations, which typically display different body composition phenotypes compared with Caucasians [46].

Conclusions

We have demonstrated that our deep, convolutional neural network, AutoMATiCA, is able to accurately and rapidly segment skeletal muscle, IMAT, VAT, and SAT from CT scans of the 3rd lumbar vertebrae across a diverse range of clinical cohorts. This network has been made freely available for all to use (<https://gitlab.com/MichaelParis/AutoMATiCA>), included with a simple user interface, to enable researchers and clinicians to conduct larger-scale body composition analysis using CT scans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to acknowledge Katie Di Sebastiano and Lesley Moisey for their role in CT body composition analysis. We also wish to acknowledge Stacey Petruzella for her role in CT scan acquisitions of the ccRCC patients.

CONFLICTS OF INTEREST AND SOURCES OF FUNDING

The authors declare no conflicts of interest. Acquisition and analysis of ccRCC CT scans was supported by a grant from the National Cancer Institute.

REFERENCES

1. Caan BJ, Meyerhardt JA, Kroenke CH et al. Explaining the obesity paradox: The association between body composition and colorectal cancer survival (c-scans study). *Cancer Epidemiol. Biomarkers Prev* 2017; 26:1008–1015 [PubMed: 28506965]
2. Vrieling A, Kampman E, Knijnenburg NC et al. Body Composition in Relation to Clinical Outcomes in Renal Cell Cancer: A Systematic Review and Meta-analysis. *Eur. Urol. Focus* 2018; 4:420–434 [PubMed: 28753824]
3. Montano-Loza AJ, Meza-Junco J, Baracos VE et al. Severe Muscle Depletion Predicts Postoperative Length of Stay but Is Not Associated With Survival After Liver Transplantation. *Liver Transplant*. 2014; 20:640–648
4. Moisey LL, Mourtzakis M, Cotton BA et al. Skeletal muscle predicts ventilator-free days, ICU-free days, and mortality in elderly ICU patients. *Crit. Care* 2013; 17:R206 [PubMed: 24050662]

5. Cruz RJ, Dew MA, Myaskovsky L et al. Objective radiologic assessment of body composition in patients with end-stage liver disease: going beyond the BMI. *Transplantation* 2013; 95:617–22 [PubMed: 23348896]
6. Mauricio SF, Xiao J, Prado CM et al. Different nutritional assessment tools as predictors of postoperative complications in patients undergoing colorectal cancer resection. *Clin. Nutr.* 2018; 37:1505–1511 [PubMed: 28918167]
7. Mamane S, Mullie L, Piazza N et al. Psoas Muscle Area and All-Cause Mortality After Transcatheter Aortic Valve Replacement: The Montreal-Munich Study. *Can. J. Cardiol.* 2016; 32:177–182 [PubMed: 26821840]
8. Montano-Loza Aldo J., Mazurak V, Ebadi M et al. VISCERAL ADIPOSITY INCREASES RISK FOR HEPATOCELLULAR CARCINOMA IN MALE PATIENTS WITH CIRRHOSIS AND RECURRENCE AFTER LIVER TRANSPLANT. *Hepatology* 2018; 67:914–923 [PubMed: 29023899]
9. Tamandl D, Paireder M, Asari R et al. Markers of sarcopenia quantified by computed tomography predict adverse long-term outcome in patients with resected oesophageal or gastro-oesophageal junction cancer. *Eur. Radiol* 2016; 26:1358–1367
10. Paris M, Mourtzakis M. Assessment of skeletal muscle mass in critically ill patients: considerations for the utility of computed tomography imaging and ultrasonography. *Curr. Opin. Clin. Nutr. Metab. Care* 2016; 19:125–130 [PubMed: 26808266]
11. Gonzalez MC, Correia MITD, Heymsfield SB. A requiem for BMI in the clinical setting. *Curr. Opin. Clin. Nutr. Metab. Care* 2017; 1
12. Prado C, Purcell SA, Alish C et al. Implications of Low Muscle Mass across the Continuum of Care: A Narrative Review. *Ann. Med* 2018; 0:1–39
13. Mourtzakis M, Prado CMM, Lieffers JR et al. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl. Physiol. Nutr. Metab* 2008; 33:997–1006 [PubMed: 18923576]
14. Shen W, Punyanitya M, Wang Z et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. *J. Appl. Physiol* 2004; 97:2333–8 [PubMed: 15310748]
15. Weston AD, Korfiatis P, Kline TL et al. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. *Radiology* 2018; 181432
16. Obst SJ, Boyd R, Read F et al. Quantitative 3-D Ultrasound of the Medial Gastrocnemius Muscle in Children with Unilateral Spastic Cerebral Palsy. *Ultrasound Med. Biol* 2017;
17. Lee H, Troschel FM, Tajmir S et al. Pixel-Level Deep Segmentation: Artificial Intelligence Quantifies Muscle on Computed Tomography for Body Morphometric Analysis. *J. Digit. Imaging* 2017; 30:487–498 [PubMed: 28653123]
18. Wang Y, Qiu Y, Thai T et al. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. *Comput. Methods Programs Biomed* 2017; 144:97–104 [PubMed: 28495009]
19. Bridge CP, Rosenthal M, Wright B et al. Fully-automated analysis of body composition from CT in cancer patients using convolutional neural networks. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif Intell. Lect. Notes Bioinformatics)* 2018; 11041 LNCS:204–213
20. Dabiri S, Popuri K, Feliciano EMC et al. Muscle Segmentation in axial Computed Tomography (CT) Images at the Lumbar (L3) and Thoracic (T4) levels for Body Composition Analysis. *Comput. Med. Imaging Graph* 2019;
21. Burns JE, Yao J, Chalhoub D et al. A Machine Learning Algorithm to Estimate Sarcopenia on Abdominal CT. 1–10
22. Graffy P, Liu J, Pickhardt P et al. Deep learning-based muscle segmentation and quantification at abdominal CT: application to a longitudinal adult screening cohort for sarcopenia assessment. *Br. J. Radiol* 2019;
23. Looijaard WGPM, Dekker IM, Stapel SN et al. Skeletal muscle quality as assessed by CT-derived skeletal muscle density is associated with 6-month mortality in mechanically ventilated critically ill patients. *Crit. Care* 2016; 20:386 [PubMed: 27903267]

24. Tandon P, Mourtzakis M, Low G et al. Comparing the Variability Between Measurements for Sarcopenia Using Magnetic Resonance Imaging and Computed Tomography Imaging. *Am. J. Transplant* 2016; 16:2766–2767 [PubMed: 27093434]
25. Paris M, Mourtzakis M, Day AG et al. VALIDation of bedside Ultrasound of Muscle layer thickness of the quadriceps in the critically ill patient (VALIDUM study): A prospective multicenter study. *J. Parenter. Enter. Nutr* 2016; In Press:
26. Kuchnia A, Earthman C, Teigen L et al. Evaluation of Bioelectrical Impedance Analysis in Critically Ill Patients: Results of a Multicenter Prospective Study. *J. Parenter. Enter. Nutr* 2017; 41:
27. Moisey LL, Mourtzakis M, Kozar RA et al. Existing equations to estimate lean body mass are not accurate in the critically ill: Results of a multicenter observational study. *Clin. Nutr* 2017; 36:1701–1706 [PubMed: 27697303]
28. Di Sebastiano KM, Yang L, Zbuk K et al. Accelerated muscle and adipose tissue loss may predict survival in pancreatic cancer patients: the relationship with diabetes and anaemia. *Br. J. Nutr.* 2013; 109:302–12 [PubMed: 23021109]
29. Paris MT, Furberg HF, Petruzella S et al. Influence of Contrast Administration on Computed Tomography-Based Analysis of Visceral Adipose and Skeletal Muscle Tissue in Clear Cell Renal Cell Carcinoma. *J. Parenter. Enter. Nutr* 2018; 42:1148–1155
30. Xu X, Zhou F, Liu B. Automatic bladder segmentation from CT images using deep CNN and 3D fully connected CRF-RNN. *Int. J. Comput. Assist. Radiol. Surg.* 2018; 13:967–975 [PubMed: 29556905]
31. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015; 1–8
32. Ludbrook J Confidence in Altman-Bland plots: A critical review of the method of differences. *Clin. Exp. Pharmacol. Physiol* 2010; 37:143–149 [PubMed: 19719745]
33. Simonsen C, de Heer P, Bjerre E et al. Sarcopenia and Postoperative Complication Risk in Gastrointestinal Surgical Oncology: A Meta-analysis. *Ann. Surg* 2018; 268:58–69 [PubMed: 29373365]
34. Friedman J, Lussiez A, Sullivan J et al. Implications of Sarcopenia in Major Surgery Defining Sarcopenia and Its Clinical Using Preoperative Imaging for Risk Assessment. 2015; 1–5
35. Martin L, Hopkins J, Malietzis G et al. Assessment of Computed Tomography (CT) -Defined Muscle and Adipose Tissue Features in Relation to Short-Term Outcomes After Elective Surgery for Colorectal Cancer : A Multicenter Approach. *Ann. Surg. Oncol* 2018; 25:2669–2680 [PubMed: 30006691]
36. Cosquéric G, Sebag A, Ducolombier C et al. Sarcopenia is predictive of nosocomial infection in care of the elderly. *Br. J. Nutr* 2006; 96:895–901 [PubMed: 17092379]
37. Makiura D, Ono R, Inoue J et al. Impact of Sarcopenia on Unplanned Readmission and Survival After Esophagectomy in Patients with Esophageal Cancer. *Ann. Surg. Oncol* 2018; 25:456–464 [PubMed: 29214454]
38. Prado CMM, Baracos VE, Mccargar LJ et al. Cancer Therapy : Clinical Body Composition as an Independent Determinant of 5-Fluorouracil ^ Based Chemotherapy Toxicity. 2007; 13:3264–3269
39. Janssen I, Shepard DS, Katzmarzyk PT et al. The Healthcare Costs of Sarcopenia in the United States. *J. Am. Geriatr. Soc* 2004; 52:80–85 [PubMed: 14687319]
40. Hanai T, Shiraki M, Nishimura K et al. Sarcopenia impairs prognosis of patients with liver cirrhosis. *Nutrition* 2014; 31:193–199 [PubMed: 25441595]
41. Chung H, Cobzas D, Birdsell L et al. Automated segmentation of muscle and adipose tissue on CT images for human body composition analysis. 2009; 72610K
42. Popuri K, Cobzas D, Esfandiari N et al. Body composition assessment in axial CT images using FEM-based automatic segmentation of skeletal muscle. *IEEE Trans. Med. Imaging* 2016; 35:512–520 [PubMed: 26415164]
43. Kullberg J, Hedstrom A, Brandberg J et al. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. *Sci. Rep* 2017; 7:1–11 [PubMed: 28127051]
44. Hu P, Huo Y, Kong D et al. Automated Characterization of Body Composition and Frailty with Clinically Acquired CT. *Comput. Methods Clin. Appl. Musculoskelet. Imaging* 2017; 171

45. Bokshan S, DePasse M, Daniels A. Sarcopenia in Orthopedic Surgery. *Orthopedics* 2016; 39:295–300 [PubMed: 27248339]
46. Wulan SN, Westerterp KR, Plasqui G. Ethnic differences in body composition and the associated metabolic profile : A comparative study between Asians and Caucasians. *Maturitas* 2010; 65:315–319 [PubMed: 20079586]

Figure 1.

Schematic overview of network. Raw CT scans are processed into dual channel images: 1) lean tissue enhanced images consist of a CT scan filtered to contain HU pixels in the -29 to $+150$ range concatenated with a raw CT scan, and 2) adipose tissue enhanced images consist of a CT scan filtered to contain HU pixels in -190 to -30 (IMAT and SAT) or -150 to -50 (VAT) range concatenated with a raw CT scan. Individual neural networks, based on the UNET architecture [31], output probability segmentations maps for muscle, IMAT, VAT, and SAT. Post processing combines tissue specific segmentation maps into a single output.

	Raw scan	Manual analysis	Network analysis
Good performance			
Kidney caner			
Muscle – 0.99			
IMAT – 0.89			
SAT – 0.99			
VAT – 0.99			
Good performance			
Liver donor			
Muscle – 0.99			
IMAT – 0.98			
SAT – 0.99			
VAT – 0.99			
Poor performance			
Liver cirrhosis			
Muscle – 0.93			
IMAT – 0.69			
SAT – 0.90			
VAT – 0.97			
Poor performance			
Critically ill			
Muscle – 0.92			
IMAT – 0.55			
SAT – 0.95			
VAT – 0.96			

Figure 2.

Example of good and poor performance scans compared with human analysis. Muscle – red; intermuscular adipose tissue (IMAT) – green; visceral adipose tissue (VAT) – yellow; subcutaneous adipose tissue (SAT) – turquoise.

Figure 3.

Bland-Altman plots of A) muscle, B) IMAT, C) VAT, and D) SAT CSA between network-predicted and manual segmentations. Solid lines indicate average bias, dashed lines indicate upper and lower limits of agreement (95% confidence intervals).

Table 1.

Demographic, physical, and body composition features of patients

	Training and validation cohort (n=804)	Testing cohort (n=89)	p-value
Age, years	52.6 (15.8)	53.9 (15.6)	0.426
Sex, % female	0.43	0.48	0.425
BMI, kg/m ²	27.9 (6.0)	28.9 (6.5)	0.158
Underweight, %	1.3	1.2	
Normal, %	33.7	25.6	
Overweight, %	36.0	40.7	
Obese, %	29.0	32.5	
Voltage, kV	119.7 (6.3)	120.9 (5.1)	0.077
Current, mA	338.3 (122.6)	368.4 (125.7)	0.042
IV contrast, %	55	56	0.863
Muscle CSA, cm ²	145.6 (39.3)	147.0 (39.7)	0.736
Muscle density, HU	36.4 (12.2)	36.5 (11.2)	0.947
IMAT CSA, cm ²	14.0 (12.3)	14.3 (13.0)	0.814
IMAT density, HU	-61.2 (6.5)	-61.7 (5.5)	0.507
VAT CSA, cm ²	140.8 (107.2)	140.0 (111.9)	0.822
VAT density, HU	-89.2 (11.3)	-89.5 (11.3)	0.788
SAT CSA, cm ²	203.6 (115.5)	208.3 (116.3)	0.715
SAT density, HU	-95.0 (12.7)	-94.6 (12.5)	0.774

BMI: body mass index, CSA: cross-sectional area, HU: Hounsfield units, IMAT: intermuscular adipose tissue, IV: intravenous, SAT: subcutaneous adipose tissue, VAT: visceral adipose tissue. Data is presented as mean (standard deviation).

Table 2.

Dice similarity coefficient comparison across population cohorts

	All (n=89)	ICU (n=30)	Liver cirrhosis (n=11)	Cancer (n=23)	Donors (n=25)
Muscle	0.983 (0.013)	0.976 (0.017) ^d	0.972 (0.015) ^{c,d}	0.989 (0.008) ^b	0.991 (0.003) ^{a,b}
IMAT	0.900 (0.034)	0.882 (0.047) ^d	0.892 (0.030)	0.904 (0.020)	0.921 (0.024) ^a
VAT	0.979 (0.019)	0.978 (0.018)	0.973 (0.018) ^c	0.992 (0.009) ^{b,d}	0.972 (0.022) ^c
SAT	0.986 (0.016)	0.980 (0.014) ^{b,c,d}	0.965 (0.029) ^{a,c,d}	0.995 (0.003) ^{a,b}	0.996 (0.002) ^{a,b}

^a different from ICU^b different from cirrhosis^c different from cancer^d different from donors.

ICU: intensive care unit, IMAT: intermuscular adipose tissue, SAT: subcutaneous adipose tissue, SD: standard deviation, VAT: visceral adipose tissue. Data is presented as mean (SD)

Table 3.

Identification of low skeletal muscle mass between human and network analyzed scans

	Testing cohort (n=89)	Kappa agreement with human-analyzed identification
Human-analyzed low skeletal muscle mass, %	34.9	-
Network-predicted low skeletal muscle mass, %	30.2	0.89
Bias corrected network-predicted low skeletal muscle mass, %	32.6	0.95

Table 4.

Dice similarity coefficient scores for automated segmentation of the 3rd lumbar vertebrae of computed tomography scans

Article	Methodology	Cohort (n)	Muscle Dice	IMAT Dice	VAT Dice	SAT Dice
Chung et al. 2009 [41]	Deformable shape modeling	Cancer (20)	0.975	-	0.977	0.986
Popuri et al. 2016 [42]	Finite element modeling	Head and neck cancer (1069)	0.950	-	0.951 *	
Lee et al. 2017 [17]	Neural network	Lung cancer (400)	0.93	-	-	-
Wang et al. 2017 [18]	Neural network	Clinical cohorts (50)	-	-	0.915	0.980
Kullberg et al. 2017 [43]	Shape modeling	Healthy (50)	-	-	0.971	0.986
Hu et al. 2017 [44]	Atlas model	Clinical (60)	0.854	-	0.887	0.933
Weston et al. 2018 [15]	Neural network	Cancers (2700)	0.96	-	0.94	0.98
Bridge et al. 2018 [19]	Neural network	Pancreatic cancer (595)	0.968	-	0.954	0.984
Dabiri et al. 2019 [20]	Neural Network	Cancers (9 179)	0.982	-	-	-
Burns et al. 2019 [21]	Neural Network	Older adults (102)	0.938	-	-	-
Present study	Neural network	Clinical cohorts (893)	0.983	0.900	0.979	0.986

Note – direct comparisons cannot be made due to the use of different datasets in development and accuracy testing.

* VAT and SAT measures were combined.