Heliyon 8 (2023) e12536

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Validation of skeletal muscle and adipose tissue measurements using a fully automated body composition analysis neural network versus a semi-automatic reference program with human correction in patients with lung cancer

Cecily A. Byrne [a], Yanyu Zhang [b], Giamila Fantuzzi [a], Thomas Geesey [c], Palmi Shah [c], Sandra L. Gomez [d,*]

[a] University of Illinois at Chicago, College of Applied Health Sciences, Department of Kinesiology and Nutrition, 1919 W. Taylor Street Chicago IL 60612, USA
[b] Rush University Medical Center, Rush Bioinformatics and Biostatistics Core, 1653 West Congress Parkway Chicago IL 60612, USA
[c] Rush University Medical Center, Department of Radiology, 1653 West Congress Parkway, Chicago, IL 60612, USA
[d] Rush University, Department of Clinical Nutrition, 600 S Paulina St, AAC 737D, Chicago, IL 60612, USA

ARTICLE INFO

Keywords:
Body composition
Validation
Computed tomography
Automated segmentation
Lumbar 1 and 2
Skeletal muscle
Adipose tissue

ABSTRACT

Rationale and objectives: To validate skeletal muscle and adipose tissues cross sectional area (CSA) and densities between a fully automated neural network (test program) and a semi-automated program requiring human correction (reference program) for lumbar 1 (L1) and lumbar 2 (L2) CT scans in patients with lung cancer.
Materials and methods: Agreement between the reference and test programs was measured using Dice-similarity coefficient (DSC) and Bland-Altman plots with limits of agreement within 1.96 standard deviation.
Results: A total of 49 L1 and 47 L2 images were analyzed from patients with lung cancer (mean age = 70.51 ± 9.48 years; mean BMI = 27.45 ± 6.06 kg/m^2; 71% female, 55% self-identified as Black and 96% as non-Hispanic ethnicity). The DSC indicates excellent overlap (>0.944) or agreement between the two measurement methods for muscle, visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) CSA and all tissue densities at L1 and L2. The DSC was lowest for intermuscular adipose tissue (IMAT) CSA at L1 (0.889) and L2 (0.919).
Conclusion: The use of a fully automated neural network to analyze body composition at L1 and L2 in patients with lung cancer is valid for measuring skeletal muscle and adipose tissue CSA and densities when compared to a reference program. Further validation in a more diverse sample and in different disease and health states is warranted to increase the generalizability of the test program at L1 and L2.

1. Introduction

Computed tomography (CT) scans are routinely used in patients with cancer for the diagnosis, staging and progression of disease to image tumors and their response to treatment. These scans can also be used to detect changes in body composition [1, 2, 3]. Altered body composition in cancer, specifically low muscle mass, is associated with poor functional status [4], reduced survival [4, 5, 6, 7], chemotherapy toxicity [8, 9, 10], discontinuation of chemotherapy and dose reduction [11]. Skeletal muscle density or the infiltration of fatty tissue into the muscle, a measure of muscle quality, has been associated with physical function impairments and poor prognosis in cancer [12, 13, 14].

Lumbar 3 (L3) cross sectional area (CSA) from an abdominal CT scan has been validated as the gold standard for body composition analysis, as skeletal muscle and adipose tissue areas in this region are highly correlated with total body skeletal muscle and adipose tissue [1, 15]. While the L3 landmark is the ideal location for estimating total body skeletal muscle, results indicate that other lumbar and thoracic landmarks (L2, L4, L5, L1, T12, T11, and T10, in preferential order) can be used as alternatives when L3 is unavailable [13], and regional skeletal muscle at L1 and L2 are also highly correlated with total body skeletal muscle [15]. Using CT scans at the L3 landmark is not always feasible for patients with lung cancer, as most individuals with lung cancer receive chest CT scans initially, which do not always extend to L3 [16]. In addition, the availability of L3 from a chest CT scan can vary from 65-84%, whereas L1 is

* Corresponding author.
E-mail address: Sandra_L_Gomez@rush.edu (S.L. Gomez).

https://doi.org/10.1016/j.heliyon.2022.e12536
Received 27 July 2022; Received in revised form 10 November 2022; Accepted 14 December 2022

available in 94.5% of chest CT scans [17]. Therefore, there is a need to utilize additional vertebral landmarks for measuring skeletal muscle in individuals with lung cancer; based on previous studies in healthy adults [13, 15] and individuals with lung cancer [16], L1 or L2 may be appropriate.

Body composition analysis via CT scans is currently expensive as well as time- and labor-intensive [18], requiring trained human research analysts to obtain images at specific vertebral landmarks and complete tagging of skeletal muscle and adipose tissue either wholly manually or after a semi-automatic analysis by specialized software, which requires the purchase of a software license plus a specialized module for semi-automation. In recent years, the use of supervised machine learning, a type of artificial intelligence, has increased in the field of medical imaging to help radiologists and researchers with automated segmentation of skeletal muscle and adipose tissues [19]. Programs that perform fully automated segmentation of skeletal muscle and adipose tissues have been developed but the cost can be a barrier for use. Supervised machine learning involves providing an algorithm with a large number of CT scans with marked tissues, allowing the machine to learn from the labeled tissues, and then allowing the algorithm to perform automated segmentation on a new data set. Neural networks are the most popular supervised machine learning used in medical imaging [19]. Automated Muscle and Adipose Tissue Composition Analysis (AutoMATiCA; https://gitlab.com/Michael_Paris/AutoMATiCA) was developed as a free, open-source neural network to analyze skeletal muscle, intermuscular adipose tissue (IMAT), visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) at L3 in a fraction of the time compared to current manual reference methods and has been validated in diverse clinical cohorts [18, 20]. As L3 may not always be available, particularly for patients with lung cancer, there is a need to validate other vertebral landmarks, specifically L1 and L2, using the fully automated body composition analysis neural network.

1.1. Specific aim

Because the use of AutoMATiCA has not yet been validated for L1 and L2, which are the landmarks most likely to be available from CT scans of lung cancer patients, the primary aim of this study was to validate skeletal muscle and adipose tissue CSA and densities between a fully automated body composition analysis neural network (test program, AutoMATiCA) and a semi-automated program requiring human correction (reference program, SliceOmatic) for CT scans of the L1 and L2 vertebral landmarks in patients with lung cancer. In addition, this study will provide further guidance on which landmark, L1 or L2, might be most appropriate to use according to the performance of the test program since including both landmarks in future body composition analysis studies may not be warranted.

Table 1. Demographics of subjects included in the study.

Variable	Mean ± SD or N (%) (N = 49)
Age (years)	70.51 ± 9.48
BMI (kg/m^2)	27.45 ± 6.06
Age Group	
Middle (40–65 years)	15 (30.61)
Older (>65 years)	34 (69.39)
BMI Group	
Low/Normal: <25 kg/m^2	18 (36.73)
Overweight: 25–29.9 kg/m^2	16 (32.65)
Obese: >30 kg/m^2	15 (30.61)
Sex	
Male	14 (28.57)
Female	35 (71.43)
Race	
White	21 (42.86)
Black	27 (55.10)
Other	1 (2.04)
Ethnicity	
Non-Hispanic	47 (95.92)
Hispanic	2 (4.08)

BMI- body mass index; kg-kilograms; m^2-meters squared; SD -standard deviation; N- number.

2. Materials and Methods

2.1. Data set

Clinically available CT scans of the L1 and L2 vertebral landmarks for patients with lung cancer diagnosed between 2014 and 2016 were retrospectively analyzed for body composition and included in this study. Sequential patients who were 19 years or older, diagnosed with lung cancer, and had a diagnostic CT scan within 45 days before or after cancer diagnosis and before the initiation of treatment were included in this pilot from a parent study of 1251 lung cancer cases from an urban, tertiary medical center. CT scans included were analyzed at Rush University Medical Center in collaboration with University of Illinois Chicago.

2.2. Computed tomography body composition analysis

Conventional CT scans of the L1 and L2 vertebral landmarks were identified by a trained radiologist from either a chest or whole-body CT scan relative to the cancer diagnosis date. CT images were anonymized prior to body composition analysis. Greyscale CT scans in DICOM® (Digital Imaging and Communications in Medicine) format were reviewed, and images with muscle cut-off were excluded from this validation study. Two methods were used for body composition analysis. Method one (test program) consisted of using a fully automated neural network (AutoMATiCA) as previously developed and validated for L3 [18, 20]. As described in the manual, greyscale CT scans in DICOM® format were uploaded into AutoMATiCA's neural network and analyzed, yielding segmentation maps and an excel spreadsheet with body composition parameters including skeletal muscle and adipose tissue CSA and tissue densities in Hounsfield Units (HU). This analysis took approximately 350 ms per CT image [18]. Segmentation of CSA for muscle and adipose tissue was visually inspected after the analysis to determine if any of the images were incompletely segmented (defined as absence of or erroneous color tagging). AutoMATiCA uses HU thresholds to segment skeletal muscles (−29 to 150 HU), VAT (−150 to -50 HU), and IMAT/SAT (−190 to -30 HU). Method two (reference program) consisted of using a semi-automatic reference program with human correction using standard protocols (SliceOmatic, Tomovision, Montreal, Canada, version 5.0) as previously described [1]. Greyscale CT images in DICOM format were initially analyzed by the ABACS module (Voronoi Health Analytics), followed by the manual correction of muscle and adipose tissue tagging by a trained researcher to capture tissues missed or erroneously tagged in the ABACS analysis. The manual correction of images was reviewed with the senior researcher (trained expert in CT body composition analysis) and corrected as needed. Each image took approximately 20 min to correct. HU thresholds are set in SliceOmatic to aid research analysts in segmenting skeletal muscles (−29 to 150 HU), VAT (−150 to -50 HU), and IMAT/SAT (−190 to -30 HU) using a brush tool to tag tissues. Tissue CSA and density were calculated by the reference program.

The demographic and clinical variables collected for each patient with lung cancer were age (years), sex (male, female), race (White, Black, Other), ethnicity (Hispanic, Non-Hispanic), height (cm), weight (kg), and body mass index (BMI, kg/m^2). This study was approved by the University of Illinois Chicago (Protocol #2020-0677) and the Rush University Medical Center (ORA #18013002-IRB01) Institutional Review Boards.

Table 2. Summary agreement statistics for Lumbar 1 measured by Dice-similarity coefficient and Bland-Altman summary statistics including assessment of proportional bias using Pearson correlation coefficients.

Comparisons		N	DSC		Bland-Altman					
Test Method	Reference method				Bland-Altman plot (difference between Test and Reference)				Proportional bias	
			Mean	SD	mean	SD	lower	upper	CC	P-value
Muscle CSA	SM	49	0.955	0.058	3.94	14.55	-24.57	32.46	0.32	**0.02**
VAT CSA	VAT	49	0.954	0.051	2.55	9.78	-16.62	21.73	0.00	0.98
SAT CSA	SAT	49	0.951	0.048	10.25	10.26	-9.87	30.36	0.39	**0.006**
IMAT CSA	IMAT	49	0.890	0.097	-3.37	3.91	-11.04	4.29	-0.52	**0.0001**
Muscle HU	SMHU	49	0.956	0.072	1.49	3.77	-5.90	8.89	-0.07	0.66
VAT HU	VATHU	49	0.996	0.007	-0.05	1.28	-2.55	2.46	-0.34	**0.02**
SAT HU	SATHU	49	0.992	0.009	1.12	2.16	-3.10	5.34	0.00	1.00
IMAT HU	IMATHU	49	0.985	0.015	1.38	2.42	-3.37	6.13	-0.39	**0.005**

Test Method: AutoMATiCA (automated segmentation).
Reference Method: SliceOmatic plus ABACS and manual correction.
DSC: Dice-similarity coefficient.
CC: Pearson Correlation Coefficient.
Muscle CSA: Cross-sectional area for muscle.
VAT CSA: Cross-sectional area for visceral adipose tissue (VAT).
SAT CSA: Cross-sectional area for subcutaneous adipose tissue (SAT).
IMAT CSA: Cross-sectional area for intermuscular adipose tissue (IMAT).
Muscle HU: Mean Hounsfield Units (or density) for muscle.
VAT HU: Mean Hounsfield Units (or density) for visceral adipose tissue (VAT).
SAT HU: Mean Hounsfield Units (or density) for subcutaneous adipose tissue (SAT).
IMAT HU: Mean Hounsfield Units (or density) for intermuscular adipose tissue (IMAT). Bold values shows $p<0.05$.

Table 3. Summary agreement statistics for Lumbar 2 measured by Dice-similarity coefficient and Bland-Altman summary statistics including assessment of proportional bias using Pearson correlation coefficients.

Comparisons		N	DSC		Bland-Altman					
Test Method	Reference method				Bland-Altman plot (difference between Test and Reference)				Proportional bias	
			Mean	SD	mean	SD	lower	upper	CC	P-value
Muscle CSA	SM	47	0.973	0.057	0.24	11.61	-22.52	22.99	0.08	0.58
VAT CSA	VAT	47	0.944	0.101	7.83	13.28	-18.19	33.86	0.51	**0.0002**
SAT CSA	SAT	47	0.961	0.035	11.13	11.32	-11.05	33.31	0.35	**0.02**
IMAT CSA	IMAT	47	0.919	0.092	-2.60	4.50	-11.42	6.21	-0.48	**0.0007**
Muscle HU	SMHU	47	0.972	0.065	-0.68	8.43	-17.21	15.84	-0.51	**0.0002**
VAT HU	VATHU	47	0.994	0.022	-2.95	22.44	-46.92	41.03	-0.78	**<.0001**
SAT HU	SATHU	47	0.990	0.021	-3.23	30.28	-62.57	56.12	-0.83	**<.0001**
IMAT HU	IMATHU	47	0.969	0.103	0.05	10.47	-20.48	20.57	-0.65	**<.0001**

Test Method: AutoMATiCA (automated segmentation).
Reference Method: SliceOmatic plus ABACS and manual correction.
DSC: Dice-similarity coefficient.
CC: Pearson Correlation Coefficient.
Muscle CSA: Cross-sectional area for muscle.
VAT CSA: Cross-sectional area for visceral adipose tissue (VAT).
SAT CSA: Cross-sectional area for subcutaneous adipose tissue (SAT).
IMAT CSA: Cross-sectional area for intermuscular adipose tissue (IMAT).
Muscle HU: Mean Hounsfield Units (or density) for muscle.
VAT HU: Mean Hounsfield Units (or density) for visceral adipose tissue (VAT).
SAT HU: Mean Hounsfield Units (or density) for subcutaneous adipose tissue (SAT).
IMAT HU: Mean Hounsfield Units (or density) for intermuscular adipose tissue (IMAT). Bold values shows $p<0.05$.

2.3. Statistical analysis

Continuous variables are presented as mean ± standard deviation, whereas categorical variables are presented as number and frequency. Dice similarity coefficients (DSC) were used to evaluate agreement between measurements from fully automated analysis (test method) and the semi-automated plus human correction (reference method) for muscle and adipose tissues. The DSC quantifies the overlap between the two analysis programs (0 = no overlap; 1 = perfect overlap or agreement) [18]. Excellent agreement was defined as a DSC >0.9. Mann Whitney U and Kruskal-Wallis tests were used to determine differences in DSC by sex and race and by BMI group, respectively. Bland Altman Plots were used to graph the difference between the test and reference method against the average of the two measurements and are used to determine limits of agreement (LOA) within 1.96 standard deviations [21]. Statistical significance was defined as $p < 0.05$. All statistical analyses were conducted using SAS® Studio, SAS® OnDemand for Academics (Copyright© 2021, SAS Institute Inc., Cary, NC, USA).

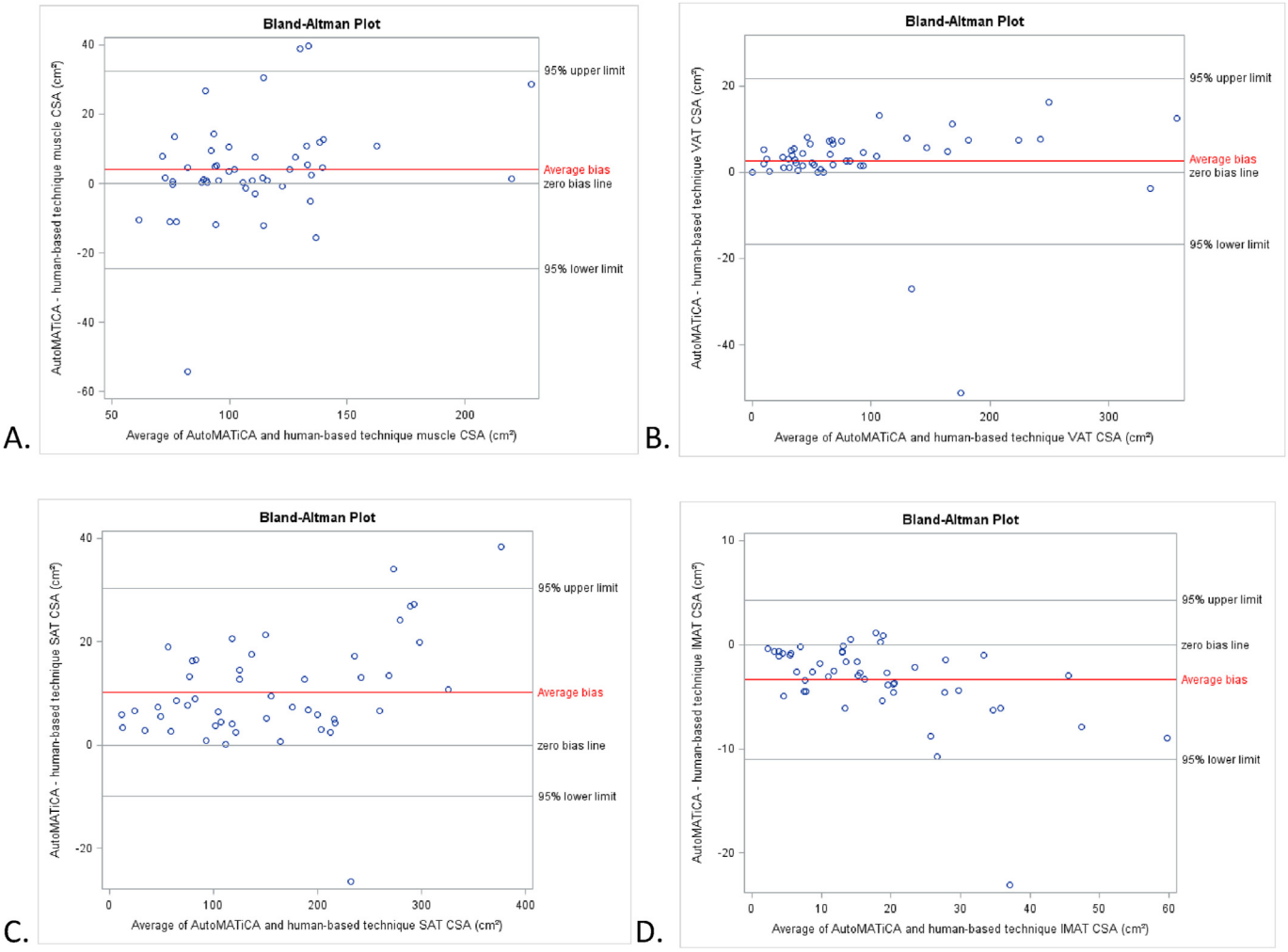

Figure 1. Bland-Altman plots of body composition parameters between test (AutoMATiCA) and reference (SliceOmatic plus ABACS and manual correction) method for Lumbar 1. Plots show cross-sectional area (CSA) for muscle (A), visceral adipose tissue (VAT, B), subcutaneous adipose tissue (SAT, C), and intermuscular adipose tissue (IMAT, D). Limits of agreement within 1.96 standard deviations are shown with average bias (red line) for each plot. The average bias line for muscle (A) and SAT (C) suggests that on average, the test method values are higher than the reference method values. The average bias line for IMAT CSA (D) suggests that on average the test method values are lower than the reference method values.

3. Results

Forty-nine subjects were included in this study for a total of 49 L1 and 47 L2 CT scans analyzed. One L2 CT scan was excluded as muscle was cut off, and the radiologist was not able to obtain one L2 image from a chest CT scan. Seventy-one percent of subjects were female, 55% self-identified as Black, 96% self-identified as non-Hispanic ethnicity, 69% were 65 years of age or older. The mean BMI was 27.45 ± 6.06 kg/m^2, indicating overweight status. See Table 1 for demographics of subjects included in the study.

The mean CSA for muscle (112.28 ± 36.09 versus 108.34 ± 31.51 cm^2), VAT (90.98 ± 81.37 versus 88.43 ± 81.34 cm^2) and IMAT (16.32 ± 11.67 versus 19.69 ± 13.70 cm^2) were similar at L1 between the test and reference programs. For L1, SAT CSA had the largest difference in mean between the two measurement methods (160.46 ± 93.14 cm^2 in the test program versus 150.21 ± 89.15 cm^2 in the reference program).

The two measurement methods produced similar mean values for muscle (115.89 ± 36.94 versus 115.65 ± 35.99 cm^2) and IMAT (19.32 ± 11.93 versus 21.93 ± 14.05 cm^2) CSA at L2, with the biggest differences in mean for VAT and SAT CSA at L2 (105.54 ± 93.21 versus 97.71 ± 86.40 cm^2 for VAT CSA; 196.41 ± 104.56 versus 185.28 ± 100.65 cm^2 for SAT CSA).

Tissue density means were similar between the two measurement methods at L1 and L2. Supplementary table 1 demonstrates that skeletal muscle CSA was significantly higher in males compared to females at both L1 (139.70 ± 36.37 versus 95.79 ± 18.13 cm^2; p = 0.0006) and L2 (152.70 ± 39.78 versus 101.50 ± 21.97 cm^2; p = 0.0005), which is consistent with other studies that have found higher muscle mass in males versus females [4, 16]. VAT was also significantly higher in males at L1 (137.70 ± 111.80 versus 68.71 ± 56.24 cm^2; p = 0.04) and approaching statistical difference at L2 (146.00 ± 122.00 versus 79.26 ± 61.03 cm^2; p = 0.08).

3.1. Concordance statistics

The DSC (Tables 2 and 3) indicates excellent overlap (>0.944) or agreement between the two measurement methods for muscle, VAT and SAT CSA and all tissue densities at both vertebral landmarks. The DSC was lowest for IMAT CSA at L1 (0.889) and L2 (0.919). At L1 and L2, the DSC for muscle and adipose tissues did not differ by sex or race (see supplementary tables 2, 3, 5 and 6). The DSC differed significantly for IMAT CSA and density and VAT density at L1 by BMI group, with a significantly higher DSC for IMAT CSA and density and VAT density in the obese group, with all DSC values indicating good (IMAT CSA in low/normal and overweight groups) to excellent agreement between the test and reference program (supplementary table 4). The DSC for muscle and adipose tissues did not differ significantly by BMI group at L2 (supplementary table 7).

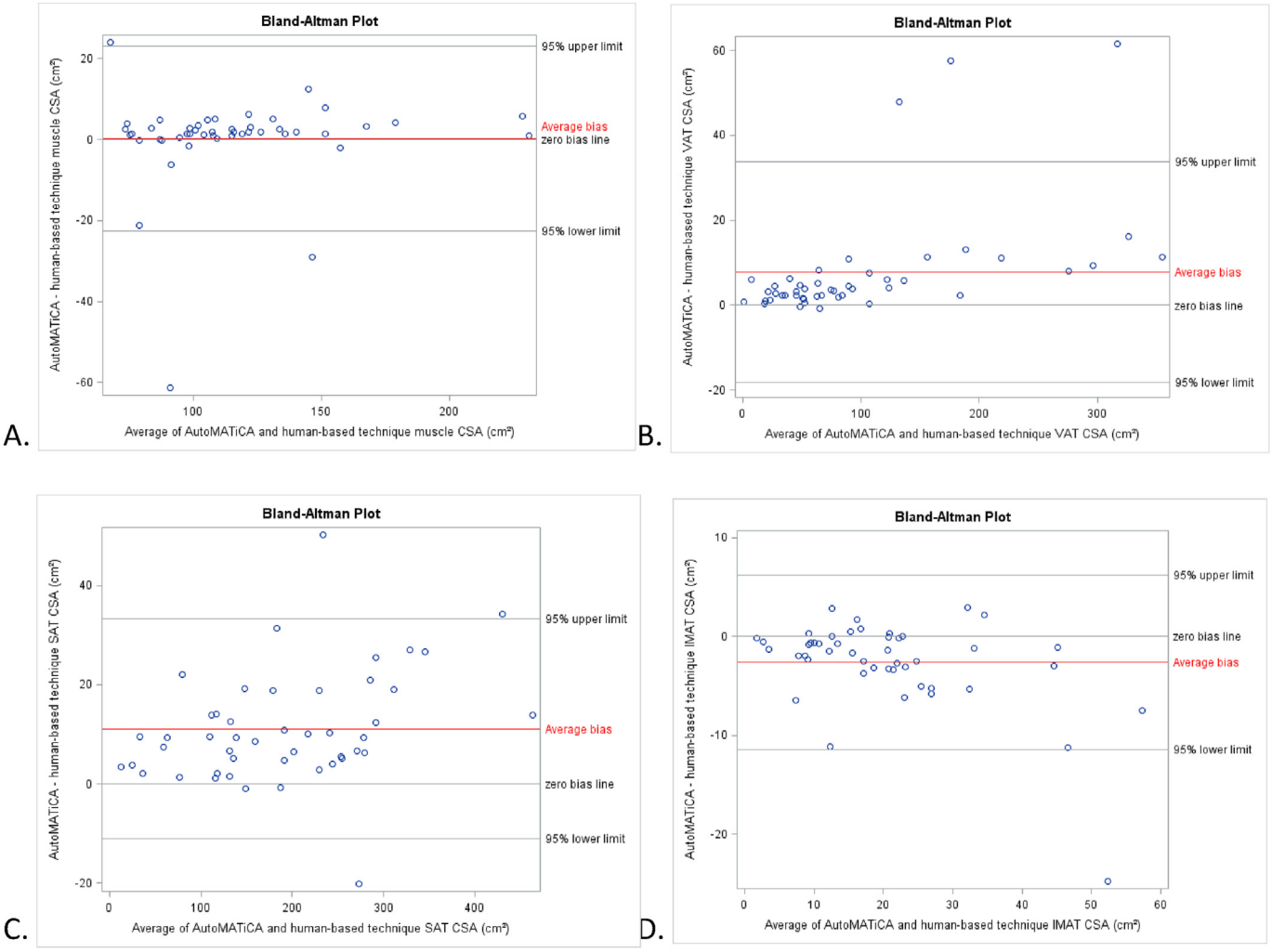

Figure 2. Bland-Altman plots of body composition parameters between test (AutoMATiCA) and reference (SliceOmatic plus ABACS and manual correction) method for Lumbar 2. Plots show cross-sectional area (CSA) for muscle (A), visceral adipose tissue (VAT, B), subcutaneous adipose tissue (SAT, C), and intermuscular adipose tissue (IMAT, D). Limits of agreement within 1.96 standard deviations are shown with average bias (red line) for each plot. The average bias line for VAT (B) and SAT (C) suggests that on average, the test method values are higher than the reference method values. The average bias line for IMAT CSA (D) suggests that on average, the test method values are lower than the reference method values.

Bland Altman Plots showed LOA of 3.94 [−24.57 and 32.46] cm^2, 2.55 [−16.62 and 21.73] cm^2, 10.25 [−9.87, 30.36] cm^2, -3.37 [−11.04, 4.29] cm^2 for muscle, VAT, SAT, IMAT CSA at L1, respectively (Table 2). Analysis for proportional bias indicated significant proportional bias for muscle, SAT and IMAT CSA with the fully automated neural network producing higher values compared to the semi-automated plus human correction reference program at L1 (Table 2 and Figure 1 A, C) and lower values for IMAT (Figure 1, D). There was no significant proportional bias for VAT CSA at L1 (Figure 1, B). For tissue densities at L1, there was significant proportional bias for VAT HU and IMAT HU with the fully automated neural network producing higher values than the reference program for both VAT and IMAT HU (Table 2).

For L2, Bland Altman Plots showed LOA of 0.24 [−22.52, 22.99] cm^2, 7.83 [−18.19, 33.86] cm^2, 11.13 [−11.05, 33.31] cm^2, -2.60 [−11.42, 6.21] cm^2 for muscle, VAT, SAT, IMAT CSA (Table 3). There was significant proportional bias for VAT, SAT and IMAT CSA with higher values for VAT and SAT Figure 2, B and C) and lower values for IMAT (Figure 2, D) by the fully automated neural network (Table 3 and Figure 2). There was no proportional bias for muscle CSA at L2 (Figure 2, A). For tissue densities at L2, there was significant proportional bias for muscle HU, VAT HU, SAT HU, and IMAT HU with the fully automated neural network producing values slightly lower than the reference program (Table 3).

4. Discussion

To the best of our knowledge, the validation of the test program (AutoMATiCA) has only been tested in two previous studies utilizing L3 in various clinical populations not including lung cancer [18, 20]; this is the first study to validate the use of the test program at L1 and L2 in patients with lung cancer.

The test program AutoMATiCA, a fully automated neural network for the analysis of skeletal muscle and adipose tissues, is valid for L1 and L2 when compared to measurements derived from a semi-automated plus human correction reference program (SliceOmatic). The test program produced measurements that are in excellent agreement when compared to the reference program for muscle, VAT, and SAT CSA per DSC values at L1 and L2. The test program also produced measurements for IMAT CSA that were in good to excellent agreement (DSC 0.889 and 0.919 for IMAT CSA at L1 and L2, respectively) with the reference program. DSC values produced by this study for muscle CSA (0.955 and 0.973), VAT CSA (0.954 and 0.944), SAT CSA (0.951 and 0.961) and IMAT CSA (0.890 and 0.919) at L1 and L2 are similar to the DSC values produced by Paris et al. (muscle 0.983; VAT 0.979; SAT 0.979; IMAT 0.900) [18], and Gomez-Perez et al. [20] (muscle 0.97; VAT 0.92; SAT 0.93; IMAT 0.83) in the validation of AutoMATiCA for L3 in diverse clinical cohorts. In

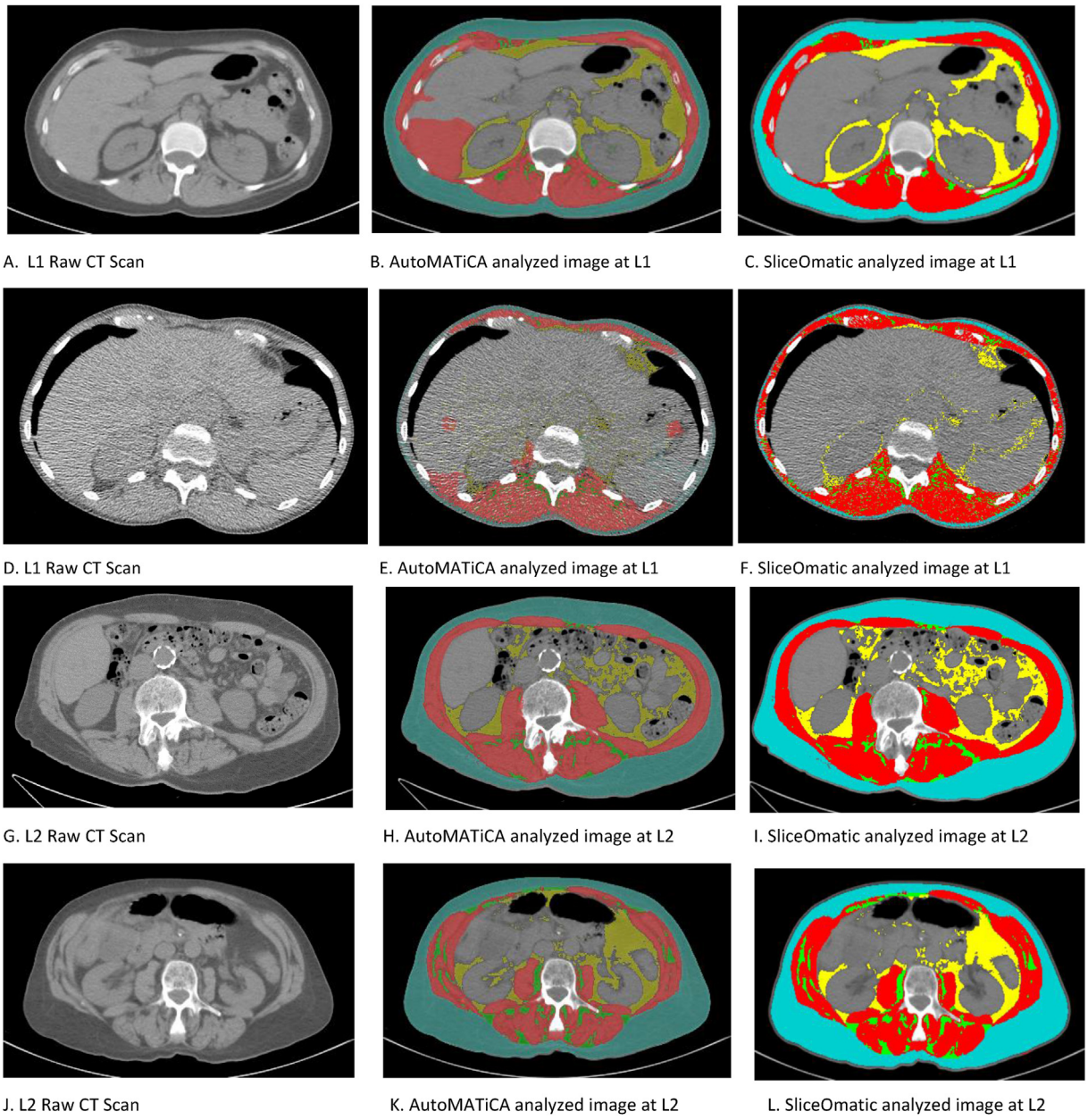

Figure 3. Examples of Raw CT scans and Analyzed Images by the Test program (AutoMATiCA) versus the Reference program (SliceOmatic plus ABACS and manual correction) at Lumbar 1 and Lumbar 2. Example of erroneous tagging of the liver as skeletal muscle (image B); poor-quality image resulting in tagging of organs as skeletal muscle and unsuccessful tagging of obliques (image E) and segmentation by AutoMATiCA at L2 (images H and K).

addition, the DSC values produced for muscle CSA at L1 and L2 in this study are higher compared to ones produced by Burns et al. in a study that compared manual segmentation of lateral wall, paraspinous, psoas, quadratus, and rectus muscles CSA to a computer system with fully automated segmentation of truncal musculature at L1-L5 [22]. In the study by Burns et al., the average DSC for all muscle groups at L1 was 0.879 $\pm$ 0.099, and at L2, the average DSC was 0.917 $\pm$ 0.055 [22] compared to 0.955 $\pm$ 0.058 and 0.973 $\pm$ 0.057 for L1 and L2, respectively, in our study. The higher DSC values at L1 and L2 in our study indicate stronger agreement between the test and reference methods for skeletal muscle CSA in our study compared to the study by Burns et al. In addition, the agreement for tissue densities at L1 (DSC 0.956–0.996) and L2 (DSC 0.969–0.994) indicate excellent agreement for muscle and adipose tissue densities between the two methods. Based on a recent review of the use of artificial intelligence in the imaging of sarcopenia, DSC is the most frequently used metric to evaluate the effectiveness of artificial intelligence, including neural networks, against a reference method in measuring muscle [19]. Therefore, the DSC for muscle and adipose tissue CSA and tissue densities from our study can be used to compare effectiveness of the test method (AutoMATiCA) against other fully automated segmentation methods. In addition, the most frequently investigated regional vertebral landmark in artificial intelligence studies is L3 [19] with only one study investigating L1 and L2 [22]. Our results thus contribute to extending the use of this neural network (AutoMATiCA) to other vertebral landmarks for body composition analysis, particularly in lung cancer patients where L3 may not always be available.

Finally, Bland Altman Plots indicated significant proportional bias for muscle, SAT and IMAT CSA at L1 and for VAT, SAT and IMAT CSA at L2, with the test program on average producing higher results for muscle, SAT, and VAT and lower results for IMAT compared to the reference program. These results are different than those produced by Paris et al. in which AutoMATiCA produced no significant proportional bias compared to the reference program at L3 [18]. In addition, Gomez-Perez et al. indicated no significant proportional bias for VAT and SAT CSA at L3 [20]. This

difference in results in our study may be due to the complexity of muscle at the chest wall and ribs at L1, where Burns et al. also found the lowest level of performance of their algorithm compared to the reference program [22]. In addition, because the fully automated test program was trained to tag skeletal muscle and adipose tissues at L3, proportional bias may occur at other vertebral levels. In manually reviewing the images at L1 with the greatest difference in skeletal muscle measurements, the proportional bias stems from erroneous tagging of liver as skeletal muscle (Figure 3, A-C), resulting in higher muscle mass measurements, and from pixilated or low-quality images resulting in lower muscle measurements from AutoMATiCA (Figure 3, D-F). At L2, there were two images with the liver erroneously tagged as skeletal muscle, but to a lower extent than at L1. Figure 3, G-I and J-L, are examples of segmentation by AutoMATiCA at L2, which closely resemble the segmentation by SliceOmatic. There was no proportional bias for skeletal muscle at L2, which may be due to closer anatomical similarities between images at L2 and L3. Poor quality images and artifacts contributed to proportional bias for VAT and SAT at L2. Further investigation with a variety of CT images at L1 and L2 are needed to determine the types of images that result in inaccuracies when analyzed in AutoMATiCA and may be contributing to proportional bias.

4.2. Limitations

We experienced delays in the ability to analyze the images via the fully automated neural network due to corruption of DICOM images upon initial extraction, making them unreadable with AutoMATiCA. Once the images were re-extracted using a different imaging software, the automated analysis via the fully automated neural network was quick and accurate. While we used conventional CT scans for body composition analysis, the use of low-dose chest CT scans before cancer diagnosis cannot be ruled out completely and may have contributed to lower performance by the automated neural network. In addition, the images were manually inspected following the fully automated analysis, however, the images cannot be manually corrected using the test program as is done with the reference program. While this study helps to validate the use of a fully automated neural network to analyze L1 and L2 landmarks in lung cancer patients, a larger and more diverse sample, including different races and ethnicities and a higher proportion of males in different disease and health states is needed to confirm validation of the fully automated neural network at L1 and L2 to increase its generalizability to other populations. Finally, L3 images from patients with lung cancer were not included in this study since L3 images from various cancer and non-cancer populations have already been evaluated using the test program [18, 20]. However, including L3 images in this study may have helped to further validate the utility and capabilities of this fully automated neural network at L3.

5. Conclusion

The use of a fully automated neural network to analyze body composition at L1 and L2 in patients with lung cancer is valid for measuring skeletal muscle and adipose tissue CSA and densities when compared to a reference program. In subsequent body composition analysis studies for patients with lung cancer, either the L1 or L2 vertebral landmark can be analyzed using the fully automated neural network and does not require the analysis of both vertebral landmarks. Further validation in a more diverse sample and in different disease and health states is warranted to increase the generalizability of the test program to analyze body composition at L1 and L2.

Declarations

Author contribution statement

Cecily A. Byrne; Sandra L. Gomez: Conceived and designed the experiments; Performed the experiments; Wrote the paper.

Yanyu Zhang: Analyzed and interpreted the data.

Giamila Fantuzzi; Thomas Geesey; Palmi Shah: Contributed reagents, materials, analysis tools or data.

Funding statement

Cecily A. Byrne was supported by the University of Illinois Chicago Graduate College Access to Excellence Fellowship.

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no competing interests.

Additional information

Supplementary content related to this article has been published online at 10.1016/j.heliyon.2022.e12536.

Acknowledgements

Cecily A. Byrne was supported by the University of Illinois Chicago Graduate College Access to Excellence Fellowship.

References

[1] M. Mourtzakis, C.M. Prado, J.R. Lieffers, T. Reiman, L.J. McCargar, V.E. Baracos, 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. 33 (5) (2008) 997–1006.

[2] A.J. Cruz-Jentoft, G. Bahat, J. Bauer, et al., Sarcopenia: revised European consensus on definition and diagnosis, Age Ageing (2019), 05.

[3] E.J. Roeland, J.D. Ma, S.H. Nelson, et al., Weight loss versus muscle loss: re-evaluating inclusion criteria for future cancer cachexia interventional trials, Support. Care Cancer 25 (2) (2017) 365–369, 02.

[4] C.M. Prado, J.R. Lieffers, L.J. McCargar, et al., Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study, Lancet. Oncol. 9 (7) (Jul 2008) 629–635.

[5] S.I. Go, M.J. Park, H.N. Song, et al., Sarcopenia and inflammation are independent predictors of survival in male patients newly diagnosed with small cell lung cancer, Support. Care Cancer 24 (5) (May 2016) 2075–2084.

[6] T. Tsukioka, N. Izumi, S. Mizuguchi, et al., Positive correlation between sarcopenia and elevation of neutrophil/lymphocyte ratio in pathological stage IIIA (N2-positive) non-small cell lung cancer patients, Gen. Thorac. Cardiovasc. Surg. 66 (12) (2018) 716–722, 12.

[7] E.M.C. Feliciano, C.H. Kroenke, J.A. Meyerhardt, et al., Association of systemic inflammation and sarcopenia with survival in nonmetastatic colorectal cancer: results from the C SCANS study, JAMA Oncol. 3 (12) (2017), e172319, 12.

[8] C.M. Prado, V.E. Baracos, L.J. McCargar, et al., Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity, *Clin Cancer Res.* 13 (11) (Jun 2007) 3264–3268.

[9] C.M. Prado, V.E. Baracos, L.J. McCargar, et al., Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment, *Clin Cancer Res.* 15 (8) (Apr 2009) 2920–2926.

[10] O. Mir, R. Coriat, B. Blanchet, et al., Sarcopenia predicts early dose-limiting toxicities and pharmacokinetics of sorafenib in patients with hepatocellular carcinoma, PLoS One 7 (5) (2012), e37563.

[11] M.C. Vega, A. Laviano, G.D. Pimentel, Sarcopenia and chemotherapy-mediated toxicity, Einstein (Sao Paulo). 14 (4) (2016 Oct-Dec 2016) 580–584.

[12] G.R. Williams, A.M. Deal, H.B. Muss, et al., Skeletal muscle measures and physical function in older adults with cancer: sarcopenia or myopenia? Oncotarget 8 (20) (May 16 2017) 33658–33665.

[13] B.A. Derstine, S.A. Holcombe, B.E. Ross, N.C. Wang, G.L. Su, S.C. Wang, Skeletal muscle cutoff values for sarcopenia diagnosis using T10 to L5 measurements in a healthy US population, *Sci Rep.* 8 (1) (2018), 11369, 07.

[14] L. Martin, L. Birdsell, N. Macdonald, et al., Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index, *J Clin Oncol.* 31 (12) (Apr 2013) 1539–1547.

[15] W. Shen, M. Punyanitya, Z. Wang, et al., Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image, J. Appl. Physiol. 97 (6) (1985) 2333–2338. Dec 2004.

[16] E.Y. Kim, Y.S. Kim, I. Park, et al., Evaluation of sarcopenia in small-cell lung cancer patients by routine chest CT, Support. Care Cancer 24 (11) (2016) 4721–4726, 11.

[17] A. Recio-Boiles, J.N. Galeas, B. Goldwasser, et al., Enhancing evaluation of sarcopenia in patients with non-small cell lung cancer (NSCLC) by assessing skeletal muscle index (SMI) at the first lumbar (L1) level on routine chest computed tomography (CT), Support. Care Cancer 26 (7) (Jul 2018) 2353–2359.

[18] M.T. Paris, P. Tandon, D.K. Heyland, et al., Automated body composition analysis of clinically acquired computed tomography scans using neural networks, *Clin Nutr.* 39 (10) (2020) 3049–3055, 10.

[19] M. Rozynek, I. Kucybała, A. Urbanik, W. Wojciechowski, Use of artificial intelligence in the imaging of sarcopenia: a narrative review of current status and perspectives, Nutr. 89 (2021), 111227, 09.

[20] S.L. Gomez-Perez, Y. Zhang, C. Byrne, et al., Concordance of computed tomography regional body composition analysis using a fully automated open-source neural network versus a reference semi-automated program with manual correction, Sensors (9) (Apr 27 2022) 22.

[21] A. Kalra, Decoding the bland-altman plot: basic review, J. Pract. Cardiovas. Sci. 3 (2017) 36–38.

[22] J.E. Burns, J. Yao, D. Chalhoub, J.J. Chen, R.M. Summers, A machine learning algorithm to estimate sarcopenia on abdominal CT, *Acad Radiol.* 27 (3) (2020) 311–320, 03.