Impact of Algorithm-Driven Slice Selection on Body Composition Analysis: A Comparison of Two Deep Learning Models Using CT Images

Jason R. Klug, PhD, Data Scientist, Radiology, Mayo Clinic
Adam Dachowicz, PhD; Timothy Kline, PhD, MS; Panos Korfiatis, PhD; Daniel Blezek, PhD; Bill Ryan; Eric Williamson, MD; Steve Langer, PhD, FSIIM; Jeremy Collins, MD; Francis Baffour, MD; Andy Missert, PhD; Gian Marco Conte, MD, PhD

Introduction

Advancements in deep learning have enabled automated algorithms for body composition analysis on CT, including autonomous slice selection for segmentation. While this enhances automation, it raises concerns about consistency and comparability of outputs. This study evaluates the agreement of two body composition algorithms at the L3 vertebra level.

Hypothesis

We hypothesized that independent slice selection could introduce variability in body composition metrics, impacting consistency.

Methods

CT scans from 1076 patients across 2009 abdomen/pelvis series were analyzed using an internally developed model and Comp2Comp, an open-source tool. Both algorithms autonomously selected slices for L3 segmentation, including skeletal muscle (SKM), visceral fat (VAT), subcutaneous fat (SAT), and intermuscular fat (IMAT). The selected slice index and segmentation outputs were compared using Bland-Altman analysis and interclass correlation coefficients (ICCs). Differences were evaluated across demographic and scanner variables.

Results

The mean slice index difference between algorithms was 1.37 ± 6.27, with identical slices selected in 17.4% of cases. The absolute slice index difference was mildly correlated with the difference in SKM area (spearman rank 0.22; r2=0.03). The segmentation agreement was high, with mean differences (LOA) in SKM = -0.23cm2 (18.17, -18.63), SAT = -19.95cm2 (30.27, -70.16), VAT = 10.40cm2 (36.51, -15.72), and IMAT = -16.44cm2 (1.78, -34.65). ICCs for area agreement were excellent for SKM, SAT, and VAT but lower for IMAT (ICC: 0.983, 0.982, 0.995, 0.595, respectively). Absolute differences in SKM area between the algorithms were larger in males compared to females (F: 4.94 =/- 7.22 cm2; M 6.16 +/- 8.26 cm2; Mann Whitney U-test; p = 5.4e-0.5). Absolute differences in SKM area between AI model segmentation (Kruskal-Wallis test) for BMI (p = 1.4e-0.3), age (p = 4.1e-0.5), patient race group (p = 1.1e-0.2), and site (p = 5.4e-0.5) were significantly different. Absolute differences in SKM area were not significantly different for scanner manufacturer (p = 5.8e-0.2) and slice thickness (p = 6.1e-0.1).

Conclusion

Allowing algorithms to select slices autonomously did not significantly affect SKM, VAT, and SAT areas, while IMAT showed only moderate agreement. Absolute differences in SKM area between the models show differential effects on demographic features, including age, sex, BMI and site. Scanner variables did not show significant effects. Overall, automated model selection of the L3 slice may depend on the quality of vertebrae segmentation, but the L3 slice selected may play a minor role in body composition segmentation variability. The clinical impact of these differences will be evaluated in the future.

Figure(s)

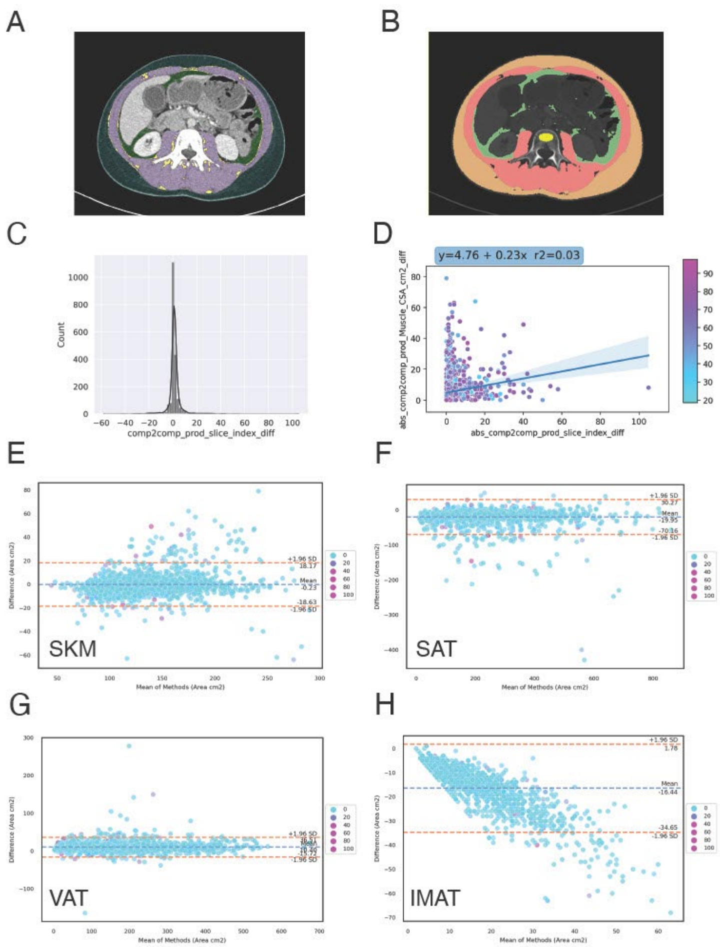

Figure 1. Example output and summary agreement results. Examples of in-house (A) and comp2comp (B) algorithm segmentation. (C) Histogram of deltas between axial slices chosen for segmentation by each algorithm. (D) Regression of difference between SKM area between algorithms on delta between chosen axial slices. (E)-(F) Bland-Altman plots for differences in segmented areas between algorithms for skeletal muscle (SKM), subcutaneous fat (SAT), visceral fat (VAT), and intermuscular fat (IMAT) segmentation outputs. Point colors correspond to the difference in L3 slice index selected between the algorithms.

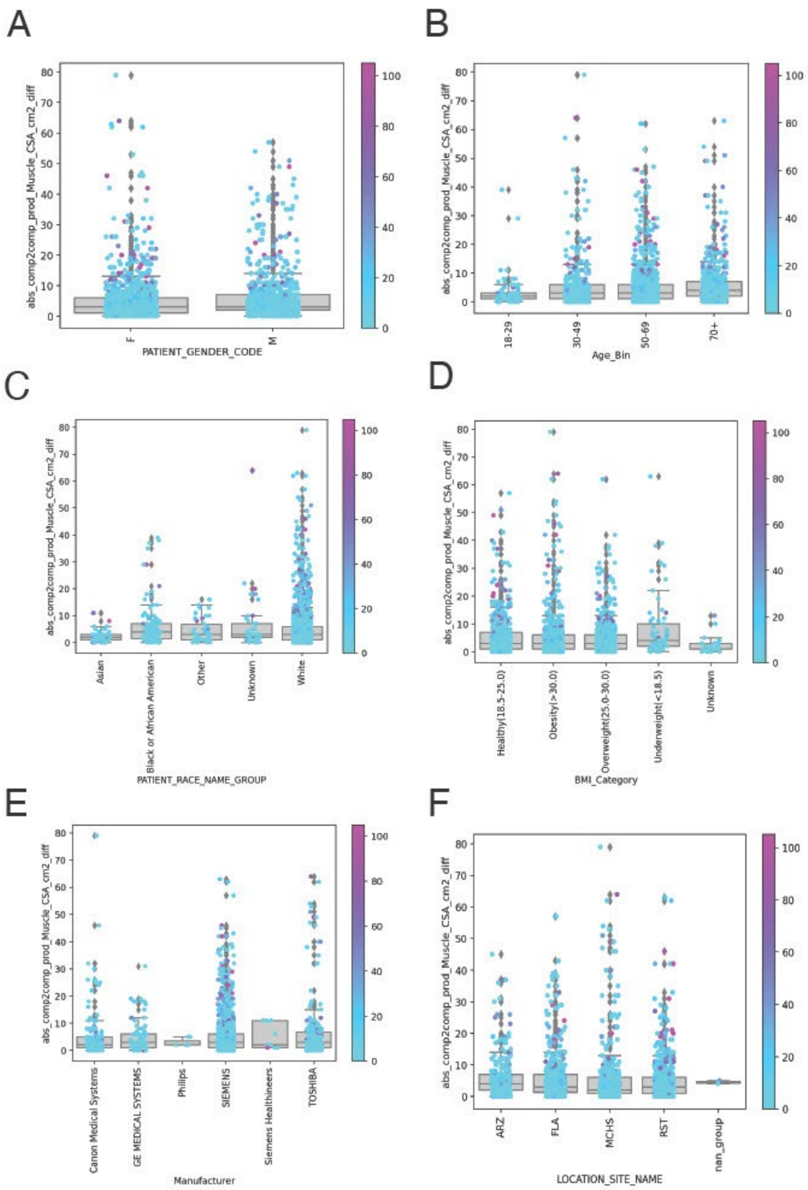

Figure 2. Difference in segmented SKM areas between algorithms, across multiple demographic variables. Point colors correspond to the difference in L3 slice index selected between the algorithms. Differences are plotted across patient sex (A), age (B), race/ethnicity (C), BMI (D), scanner manufacturer (E), and performing site location within the health system (F).

Keywords

Artificial Intelligence/Machine Learning; Emerging Technologies; Imaging Research