Development and evaluation of a method for segmentation of cardiac, subcutaneous, and visceral adipose tissue from Dixon magnetic resonance images

Jon D. Klingensmith
Addison L. Elliott
Amy H. Givan
Zechariah D. Faszold
Cory L. Mahan
Adam M. Doedtman
Maria Fernandez-del-Valle

Development and evaluation of a method for segmentation of cardiac, subcutaneous, and visceral adipose tissue from Dixon magnetic resonance images

Jon D. Klingensmith,^{a,*} Addison L. Elliott,^a Amy H. Givan,^b Zechariah D. Faszold,^a Cory L. Mahan,^b Adam M. Doedtman,^b and Maria Fernandez-del-Valle^b

^aSouthern Illinois University Edwardsville, Department of Electrical and Computer Engineering, Edwardsville, Illinois, United States

^bSouthern Illinois University Edwardsville, Department of Applied Health, Edwardsville, Illinois, United States

Abstract. Magnetic resonance imaging (MRI) has evolved into the gold standard for quantifying excess adiposity, but reliable, efficient use in longitudinal studies requires analysis of large numbers of images. The objective of this study is to develop and evaluate a segmentation method designed to identify cardiac, subcutaneous, and visceral adipose tissue (VAT) in Dixon MRI scans. The proposed method is evaluated using 10 scans from volunteer females 18- to 35-years old, with body mass indexes between 30 and 39.99 kg/m². Cross-sectional area (CSA) for cardiac adipose tissue (CAT), subcutaneous adipose tissue (SAT), and VAT, is compared to manually-traced results from three observers. Comparisons of CSA are made in 191 images for CAT, 394 images for SAT, and 50 images for VAT. The segmentation correlated well with respect to average observer CSA with Pearson correlation coefficient (R^2) values of 0.80 for CAT, 0.99 for SAT, and 0.99 for VAT. The proposed method provides accurate segmentation of CAT, SAT, and VAT and provides an option to support longitudinal studies of obesity intervention. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.6.1.014004](https://doi.org/10.1117/1.JMI.6.1.014004)]

Keywords: magnetic resonance imaging; adipose tissue; obesity; segmentation; image processing.

Paper 18118R received Jun. 1, 2018; accepted for publication Jan. 18, 2019; published online Feb. 7, 2019.

1 Introduction

The increased prevalence of obesity is one of the most important health issues of recent times.¹ It is associated with several comorbidities and noncommunicable diseases, including type-2 diabetes, high blood pressure, cardiovascular disease, and some forms of cancer.² In addition to its impact on health, the estimated costs attributed to obesity-related issues are between \$70 and \$100 billion.¹ Increased adiposity in the human body is divided into two categories: subcutaneous adipose tissue (SAT), located below the skin, and visceral adipose tissue (VAT), located in the abdominal cavity and surrounding internal organs. VAT is present in various deposits around the internal organs and has been demonstrated to be a metabolically-active organ with both paracrine and endocrine actions. It can secrete hormones and molecules that have impact on other tissue and organs, including proinflammatory cytokines.³ Ectopic fat depots are located in areas not traditionally associated with fat storage and can interact with important internal organs and potentially lead to the development of noncommunicable diseases (i.e., heart, liver, and kidney disease).⁴ Ectopic fat also develops around the heart, both inside the visceral pericardium (epicardial fat), and between the visceral and parietal pericardium (paracardial fat). Due to its direct relationship with the development of cardiovascular disease and cardiovascular risk,^{5,6} identifying increased cardiac adipose tissue (CAT) has also emerged as an important goal in understanding the associated cardiovascular risk.

Magnetic resonance imaging (MRI) and computed tomography (CT) have emerged as imaging modalities capable of identifying and facilitating quantification of VAT and SAT,² but can also provide assessment of the CAT.^{7–12} MRI is capable of differentiating adipose tissue from water¹³ and has emerged as a primary tool for imaging studies in obesity and the quantification of SAT and VAT.^{14–18} In addition, as the relationship between cardiovascular disease and volume of CAT has become more pronounced, cardiac MRI has been used to visualize and quantify CAT.^{12,19,20}

MRI and CT provide noninvasive, three-dimensional (3-D) visualization and volumetric assessment of adipose tissue, but require segmentation of the adipose tissue in a large number of image slices for volumetric analysis. Manual segmentation of the adipose tissue in the image slices of each volumetric scan is subject to interobserver variability and requires a large amount of time for analysis.²¹ Fully automated methods remove the user-dependence of the results but are subject to the challenges associated with anatomical variation and artifacts in the images.^{14,16,18,19} Semiautomated methods attempt to balance this tradeoff, potentially reducing the time required for analysis and some of the user-dependence.^{12,17} They typically allow for the user interaction in one of two ways—either by facilitating user-review and correction of automated segmentation results via direct manipulation of contours or manually adjusting threshold values, or by facilitating user guidance early on in the process and utilizing automated approaches to complete the segmentation (i.e., a supervised method).

*Address all correspondence to Jon D. Klingensmith, E-mail: jokling@siue.edu

Regardless of the approach taken to reduce user-dependence, inherent interobserver variability in image interpretation may remain.

The quantitative results achieved by automated and semiautomated segmentation techniques are subject to variability in anatomy and image quality. In addition, it is challenging to assess the accuracy and validity of the segmentation approaches because there is no known gold-standard. Even if using semi-automated techniques, allowing for correction or adjustment of results, the reliability of the data and the interobserver variability are important to understand for longitudinal studies.²¹ If the objective is to replace manual segmentation with a computer-based, fully-automated or semiautomated approach, then the outcome of the computer-based method should be at least as good as the expert observers. In the other words, the average variability between the computer method and the human observers should be no larger than the average variability among the human observers themselves.^{22–24} Results from previous segmentation approaches for VAT, SAT, and CAT have been compared to a single expert observer, and volume has been the main measurement used for assessment.^{14,16–19} However, calculating volume alone may obscure errors in the segmentation output in individual image slices—Dice similarity coefficient²⁵ (DSC), a comparison between overlapping areas from two different segmentations was used previously for comparison. Despite some reported inaccuracies in slice-based comparisons, VAT volume only differed by 7%.¹⁸ The objective of this work was to develop a segmentation algorithm and use a rigorous framework for evaluating the performance of the approach within the context of interobserver variability, providing understanding for its use in longitudinal studies of obesity intervention.

2 Materials and Methods

2.1 Subjects and Magnetic Resonance Image Acquisition

The image data used for algorithm development and evaluation were acquired for a pilot study on the effects of short-term (3 weeks) resistance training on VAT, SAT, and CAT volumes. This project was approved by the institutional review board (ID: #504792) and all volunteer subjects provided informed consent. Participants included in this study were females, 18- to 35-years old, with BMI between 30 and 39.99 kg/m². Individuals with cardiovascular, metabolic, or pulmonary diseases, those who were pregnant, smokers, those taking medications that affect endocrine or cardiovascular function, or engaged in low-intensity strength training more than two times per week or moderate-high intensity training of any type, were excluded. For each scan, the subjects laid in a supine, head-first position with arms extended above their head and a triangular support pillow under their knees. Two body coils were laid length-wise over the thoracic and abdominal regions with a small overlap. The scans were acquired using a 3.0 T Skyra system (Siemens, Munich, Germany), taken with a 20-s breath-hold. Each scan involved the acquisition of 120, 2.5-mm-thick axial images with a 20% gap, resulting in a separation of 3.0 mm between slice centers and total height of 36 cm. The image dimensions, therefore, were 320 × 260 in-plane (widest FOV = 45 cm), yielding an effective voxel size of 1.40625 × 1.40625 × 3.0 mm³. Imaging was performed using a multiecho T1-weighted volumetric gradient echo sequence known as the volumetric interpolated breath-hold examination Dixon sequence (TR/TE = 3.97/

1.23 ms, flip angle = 9 deg) along with the default factory settings. This sequence yielded a set of registered in-phase and out-of-phase image volumes, from which the fat-only and water-only image volumes were reconstructed. The algorithm was evaluated using 10 scans from this data set.

2.2 Segmentation Algorithm

2.2.1 Preprocessing

The segmentation algorithm was developed and implemented in custom software²⁶ using Python 3.5.4 and GitHub for revision control. Additional open-source libraries, including SciPy,²⁷ NumPy,²⁷ scikit-image,²⁸ matplotlib,²⁹ OpenCV,³⁰ SimpleITK,³¹ pynrrd, and PyQt5, were used in the project. Figure 1 depicts the block diagram associated with the preprocessing steps of the algorithm.

The initial step in the algorithm was concatenating the upper and lower image volumes for the fat-only and water-only images to remove the overlap between the two scans and provide a continuous volume for analysis. Then, the N4ITK bias correction algorithm was applied, separately to the fat and water image slices, to remove inhomogeneities in the image data resulting from any nonuniformities in the magnetic field during image acquisition.³² The N4ITK bias correction algorithm is an improvement upon the widely accepted nonparametric non-uniform normalization (N3) algorithm that iteratively seeks the smooth multiplicative field that maximizes high-frequency content in the MRI images using B-splines. Its operation requires identification of appropriate parameters, chosen empirically in the current study based on qualitative inspection of the resulting bias-corrected images. The number of histogram bins used was 200, the bias field full-width-at-half-max value was 0.15, the Wiener filter noise level was 0.01, the shrink factor was 4, the spline order was 3, the convergence threshold was 0.001, and the number of iterations was limited to 50.

The remaining portion of the segmentation algorithm was performed separately on each axial slice of the MRI volume. The bias-corrected images were segmented into fat and water binary masks from the fat-only and water-only images using a *k*-means clustering algorithm that splits the grayscale pixel values into clusters such that the variance between clusters is minimized.³³ Two clusters were used to identify the fat and water from the fat-only and water-only images. The resulting binary images (i.e., masks) were combined via a logical OR operation to create a body mask. Then, a binary morphological closing operation was performed using a 3-pixel-radius disk structuring element to close gaps and fill holes in the body mask. Using the fat, water, and body masks as input, the algorithm proceeded in two different parts: the abdominal cavity algorithm to identify SAT and VAT and the thoracic cavity algorithm to identify SAT and CAT. The top of the diaphragm was used as the boundary used to demarcate the two body cavities and the application of the two algorithms.

2.2.2 Abdominal cavity algorithm

For the abdominal cavity, the two relevant targets of adipose tissue were SAT, defined as adipose tissue between the skin and the aponeuroses and fasciae of the muscles, and the VAT, defined specifically in this case as the VAT in the abdominal cavity. The flowchart for the algorithm to identify these two tissue types is shown in Fig. 2.

Fig. 1 Illustration of the preprocessing steps of the segmentation algorithm. The original fat-only and water-only images, shown in (a), are input to a bias-correction algorithm. The output of the N4ITK bias correction algorithm (b) is used to create fat-only and water-only masks using the *k*-means algorithm. The fat and water masks, shown in (c), are combined to create the body mask (d). The body mask is used as input to the abdominal and thoracic algorithms.

The boundary between the SAT and VAT was identified using an active contour model, initially described in Kass et al.³⁴ The active contour model used is an iterative algorithm that minimizes the energy along the proposed boundary contour, defined as

$$E_{\text{snake}} = \int_0^1 \{E_{\text{int}}[\mathbf{v}(s)] + E_{\text{image}}[\mathbf{v}(s)]\} ds, \quad (1)$$

where

$$E_{\text{int}} = \frac{[\alpha |v_s(s)|^2 + \beta |v_{ss}(s)|^2]}{2}, \quad (2)$$

$$E_{\text{image}} = w_{\text{edge}} E_{\text{edge}}, \quad (3)$$

and

$$E_{\text{edge}} = -|\nabla I(x, y)|^2, \quad (4)$$

where $I(x, y)$ is the grayscale pixel intensity at coordinates x and y . Here, $\mathbf{v}(s)$ is the boundary parameterized from $s = 0$ to 1, ∇ is the gradient operator, and α , β , and w_{edge} are the parameters. The initial boundary used in the active contour model is the outline of the body mask. One of the challenges associated with using active contour models is selecting the optimal parameters.

In this case, there are multiple values of the parameters that yield acceptable results and a wide range of values that yield contours resulting in problems in subsequent steps of the algorithm. The parameters in this study were chosen qualitatively by inspecting intermediate results on several axial slices. The alpha parameter controls how quickly the contour contracts or expands. The beta parameter controls the smoothness of the contour, removing sharp edges or corners. In this application, the beta parameter is less significant because the shapes of interest lack sharp edges or corners. The edge weight parameter controls how closely the contour will follow the edge content of the image. The selected parameters for the active contour were $\alpha = 0.70$, $\beta = 0.01$, and $w_{\text{edge}} = 5$. In addition, the algorithm iterated until the cost function was below a threshold of 0.1 or reached 2500 iterations. An example set of intermediate results with different parameter selections is shown in Fig. 3 to demonstrate the significance of selecting optimal parameters. The red contour in each image is the initial contour, i.e., the outline of the body mask, and the blue contour is the final contour. The middle image [Fig. 3(b)] shows the parameters selected.

The boundary determined by the active contour was filled to create an abdominal mask. Any voxels identified as fat were classified as VAT if inside the abdominal mask and SAT otherwise, as shown by the following logical operations:

Fig. 2 The body mask from Fig. 1 is used to initialize an active contour algorithm. The resulting contour is overlaid on the fat mask (a) to identify the inner boundary of the subcutaneous fat (b). The inner boundary is filled (c) and fat outside this boundary is classified as subcutaneous (d) and fat inside this boundary is classified as visceral (e).

Fig. 3 Example intermediate images with different parameter selections for the active contour algorithm. The parameters for (a) were $\alpha = 0.01$, $\beta = 0.01$, and $w_{\text{edge}} = 5$. The image in (b) uses the parameters that were selected: $\alpha = 0.70$, $\beta = 0.01$, and $w_{\text{edge}} = 5$. The parameters in (c) were $\alpha = 2.0$, $\beta = 0.01$, and $w_{\text{edge}} = 5$.

$$\text{SCAT} = \overline{\text{abdominal_mask}} \cdot \text{fat_mask} \quad (5)$$

and

$$\text{VAT} = \text{abdominal_mask} \cdot \text{fat_mask}. \quad (6)$$

As a final processing step, small objects in the VAT and SAT were removed by labeling objects (using four-connectivity) and removing those with an area $< 5^2$ pixels, which was empirically determined.

2.2.3 Thoracic cavity algorithm

The tissue targets in the thoracic region are SAT, again defined as adipose tissue between the skin and the aponeuroses and fasciae of the muscles, and CAT, defined as adipose tissue layered

around the heart. The flowchart for this algorithm is shown in Fig. 4.

The same active contour model defined above in Sec. 2.2.2 was used to identify the boundary between the SAT and the internal fat. The same parameters were also used ($\alpha = 0.7$, $\beta = 0.01$, and $w_{\text{edge}} = 5.0$, max iterations = 2500, and convergence = 0.1).³⁴ This boundary was filled to create a thoracic mask. Using the fat mask as a reference, any pixels identified as fat were classified as intrathoracic adipose tissue (ITAT) if inside the thoracic mask and SAT otherwise, as defined by the following binary logical operations:

$$\text{SCAT} = \overline{\text{thoracic_mask}} \cdot \text{fat_mask} \quad (7)$$

and

Fig. 4 The active contour that is initialized from the body mask is used to identify the inner boundary of the subcutaneous fat. The fat mask is shown in (a) and an example result of the active contour algorithm is shown overlaid on the fat mask in (b). The resulting inner boundary is filled (c) and fat outside that region is labeled as subcutaneous fat (e). The two largest voids inside the body mask were identified as the lung mask (d). Fat inside the body mask and framed in by the lung mask and user-identified landmarks was labeled as cardiac fat (f).

$$\text{ITAT} = \text{thoracic_mask} \cdot \text{fat_mask}. \quad (8)$$

The ITAT image was used as an intermediate step and further segmented to identify the CAT. Image slices corresponding to the top of the diaphragm and aortic arch were identified by the user as start and end slices for CAT segmentation. In addition to the start and stop slices, anterior and posterior constraints were identified by the user to form a boundary to separate CAT from ITAT. This was only required for three image slices across the heart and the remaining anterior–posterior boundaries were linearly interpolated, providing bounds for each image slice. The left and right constraints for the bounds were identified from the lung mask. The lungs, being composed primarily of air, were identified as voids in the fat and water masks of the Dixon images as defined by the following relationship:

$$\text{lung_mask} = \text{body_mask} \cdot \overline{\text{fat_mask}} \cdot \overline{\text{water_mask}}. \quad (9)$$

The lung mask was further refined to remove small objects that may be connecting the left and right lungs by performing a binary morphological opening using a 10-pixel-radius disk structuring element. Following the morphological opening, holes were filled and objects labeled using four-connectivity. All objects were removed except the two largest, leaving the left and right lungs as the segmented objects in the binary

image. The lung mask formed the left and right constraints for the bounds of the CAT mask [see Fig. 4(d)]. Fat pixels from the fat mask that were inside the specified bounds were classified as CAT. As a final processing step, small objects in the CAT and SAT were removed by labeling objects (using four-connectivity) and removing those with an area $< 5^2$ pixels, which was empirically determined.

For each scan, there were three user-entered landmarks required for algorithm initialization. As an aid to this process, custom software was developed to facilitate the process (Fig. 5). The items required included an axial slice denoting the top of the diaphragm to divide the abdominal and thoracic body cavities, a line across the umbilicus to close the contour and create a valid body mask, and a set of posterior and anterior bounding box constraints for the CAT. The full set of posterior and anterior CAT constraints were linearly interpolated between axial slices from three entered by the user.

2.3 Method Evaluation

The evaluation of the segmentation algorithm was performed in the context of interobserver variability. A custom cross-platform software package for manual tracing was developed for the study, as shown in Fig. 6. The tool was developed using Qt and C++ and provided complete control for the user, including

Fig. 5 A sample screenshot of the software developed for supervision and guidance of the algorithm. After identification of the landmarks, the remainder of the algorithm steps were fully automated.

Fig. 6 A sample screenshot of the software developed for manual tracing of the adipose tissue. It provided the ability to blend the fat-only and water-only images, with user-selected colormaps, for optimal visualization. The example demonstrates the fat only image displayed with a “green” colormap, blended with the water-only image, shown with a “bone” colormap. The user manually traced fat deposits in these images and could save contour files for further processing or editing.

mechanisms to blend the water-only image and the fat-only image together to facilitate visual interpretation and tracing. The user could independently select the colormap and transparency level for the fat-only and water-only images. In addition, the software also provided an erasing tool and an “undo” feature, allowing users to make additional manual corrections to their traced fat depots as necessary based on their interpretation. Three trained observers manually delineated the CAT, SAT, and VAT using the custom software.

Due to the laborious nature of the process, 10 scans were randomly chosen to create the data set for evaluation. The CAT was traced in all the image slices from the aortic arch to the diaphragm, as identified by each observer, resulting in 191 image slices for CAT segmentation comparisons. The SAT was traced across the entire scan, beginning at the bottom of the abdominal cavity, using every third image slice for manual analysis. This yielded 394 total image slices for SAT comparison. Due to the complexity and time required for manual tracing of the VAT, every 20th image slice was used for manual tracing from the bottom of the abdominal cavity to the diaphragm, resulting in 50 images for evaluation. In total, there were 635 image slices, and three observers manually analyzed each. In addition to the image slice comparisons, volumes were calculated using trapezoidal integration and comparisons were made between the computer algorithm and the volumes computed by the three observers.

2.4 Statistical Analysis

Comparisons made on a per-slice basis included difference in cross-sectional area (CSA), Hausdorff distance, and DSC. The difference in CSA was represented in Bland–Altman charts, which demonstrate the difference in CSA versus the mean CSA.³⁵ The basis for comparison in the Bland–Altman analysis (i.e., the manual expert measurement) was the average of the CSA for the three observers. Pearson correlation coefficient was also calculated to assess the agreement between the computer-derived data and the observer-derived data.

CSA is a measurement derived from a segmented contour, but Hausdorff distance is a more strict evaluation of agreement. With the contours represented as points, the Hausdorff distance is the maximum Euclidean distance between the contours, i.e., given point sets $A = \{a_1, \dots, a_p\}$ and $B = \{b_1, \dots, b_q\}$, the Hausdorff distance³⁶ is

$$H = \max[h(A, B), h(B, A)], \quad (10)$$

where the operator $h(A, B)$ is defined as

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|. \quad (11)$$

A small Hausdorff distance means the contours are in good agreement and a large Hausdorff distance means there is substantial disagreement between contours. DSC was also used to assess agreement between segmentation results between the computer algorithm and the observers. Given two overlapping regions A and B , DSC is defined as

$$\text{DSC} = \frac{2N(A \cap B)}{N(A) + N(B)}, \quad (12)$$

where $N()$ is calculated as the number of pixels in a region. A DSC close to 1.0 means excellent agreement and DSC close to

0.0 means there is little agreement between the regions. To assess the algorithm performance within the context of interobserver variability, a modified version of Williams’ index (WI) was used.²² The average agreement between the computer and the observers is computed as

$$P_0 = \frac{1}{n} \sum_{j=1}^n P_{0,j}, \quad (13)$$

where $P_{0,j}$ is the level of agreement between the computer and the j ’th observer. The average agreement between the observers is computed as

$$P_n = \frac{2}{n(n-1)} \sum_j \sum_{j': j' \neq j} P_{j,j'}. \quad (14)$$

With these two values, the WI is computed as the ratio of the two:

$$I = \frac{P_0}{P_n}. \quad (15)$$

The 95% confidence intervals for WI were computed using a jackknife nonparametric sampling technique.²² If WI is close to 1.0 (i.e., if 1.0 falls within the 95% confidence interval), then it can be assumed that the computer does not produce results different than a typical observer and is within the normal observer variation. One advantage of WI is that any metric evaluating the level of disagreement and that produces multivariate data can be used for the computation. In this study, WI was computed using three different metrics indicating variation between computer algorithm and observer results. Because DSC measures similarity on a scale from 0.0 to 1.0, with 0.0 being entirely dissimilar and 1.0 being perfect agreement, the inverse of DSC ($1.0 - \text{DSC}$) was used for the WI calculation. In addition, absolute difference in CSA and Hausdorff distance was also used to compute WI.

For volume-based assessment, VD_A was used to evaluate the agreement in volume between the computer algorithm and the observers. It was computed as the absolute difference between the automated volume and the average observer volume divided by the average observer volume. Pearson correlation coefficient was computed to assess the correlation between computer-derived volumes and observer-derived volumes. Bland–Altman charts were used to visualize the agreement. In addition to the traditional assessment using comparisons of volume calculations, WI based on volume was also calculated as the ratio of observer-observer variability to computer-observer variability.

3 Results

The segmentation was successful on all scans involved in the study. Manual tracing was performed on 10 scans by three observers, resulting in 635 image slices for comparison and evaluation. An example set of images with segmented results, comparing computer output to that of one of the observers is shown in Fig. 7, and a summary of the single-slice comparison data for CAT, SAT, and VAT are given in Table 1.

The Bland–Altman charts visualizing the relationships between the computer algorithm areas and the average area of the observers are shown in Fig. 8. Strong correlations were found for each of the adipose tissue types. In addition, with respect to CSA, the 95% confidence interval ranges for

Fig. 7 An example fat-only image of the CAT is shown in (a). Example segmentation results are shown in (b), with user-identified fat from one of the observers shown in blue, the computer-identified fat shown in red, and overlapping fat shown in magenta. The Hausdorff distance for this example was 16.5 mm, and the DSC was 0.75. An example segmentation result for the subcutaneous fat is shown in (c) and (d) with the same color-coding for convenient comparison. The Hausdorff distance for this example was 5.6 mm and the DSC was 0.93. Finally, example results for visceral fat segmentation are shown in (e) and (f). The Hausdorff distance for the visceral fat segmentation example was 57.4 mm and the DSC was 0.78.

each adipose tissue type contained 1.0. With respect to Hausdorff distance, the 95% confidence interval contained 1.0 for the CAT. The WI for Hausdorff distance did indicate slight increases in variation between the computer and the observers compared to the observers themselves. The same

was true for the DSC, when WI was computed based on single slice comparison of results.

A summary of the volume comparison data is shown in Table 2. Strong correlations and good agreement were found when comparing volume measurements from each of the three tissue types, as shown in the Bland–Altman charts and correlation plots in Fig. 9. In addition, the 95% confidence intervals for WI all contained 1.0, indicating that the computer results did not introduce more variation than exists between the observers.

4 Discussion

The purpose of this study was to develop and evaluate a segmentation method that could be used to quantify obesity and excess adiposity in longitudinal studies of intervention. The data presented demonstrate that the segmentation method is accurate when compared to a group of observers, providing confidence that this method can be used in future studies of the effects of obesity interventions.

The segmentation approach presented here was a hybrid method based on optimal thresholding via *k*-means clustering and active contours for process guidance. A supervisory step was used at the beginning of the process to control and guide the segmentation of the CAT. A full evaluation was performed on the algorithm using interobserver variability based on manual results from three observers. Based on this evaluation, the approach produced results that agree well with previously published data for the SAT and VAT. Addeman et al.¹⁴ reported on an automated technique, based on thresholding the fat-fraction and incorporating a correction for partial volume effects, for SAT and VAT identification. Their mean VD_A was 0.9% for SAT and 5.0% for VAT (termed IAAT in their study). The cohort was more limited in the present study, but mean VD_A for SAT (−2.0%) and VAT (5.4%) were similar. Data presented in other studies were also in a comparable range—mean VD_A of 0.9% for SAT and 4.2% for VAT was reported by Wald et al.¹⁸ Thormer et al.¹⁷ did not present VDA directly, but the Bland–Altman charts suggest that their results were in the 4% to 6% range for SAT and VAT.

The previously published studies of segmentation algorithms for identification of SAT and VAT^{14,16–18,37} did not provide segmentation of the CAT. However, the objective here was to develop a robust scanning protocol and segmentation algorithm for efficient quantification of all three tissue types, including CAT, SAT, and VAT. The CAT presents additional challenge beyond the SAT and VAT, as it is smaller in scope and more difficult to identify visually. Segmentation algorithms can vary with respect to the degree of user-interaction. Fully manual

Table 1 Summary of statistical results for single slice comparisons.

Adipose tissue type (number of image comparisons)	Pearson correlation coefficient (R^2)	AD_A (%) (+/- standard deviation)	WI-CSA (95% confidence interval)	WI-DSC (95% confidence interval)	WI-HD (95% confidence interval)
Cardiac ($n = 191$)	0.80*	−2.1 (+/ − 21.2)	1.07 (0.71, 1.42)	0.83 (0.69, 0.96)	0.82 (0.56, 1.09)
Subcutaneous ($n = 394$)	0.99**	5.9 (+/ − 6.0)	0.72 (0.27, 1.17)	0.78 (0.58, 0.98)	0.62 (0.36, 0.88)
Visceral ($n = 50$)	0.96**	7.2 (+/ − 10.0)	1.13 (0.32, 1.93)	0.84 (0.80, 0.89)	0.71 (0.46, 0.95)

AD_A , mean difference in area compared to average of manual observer areas; WI-CSA, Williams' index based on cross-sectional area; WI-DSC, Williams' index based on Dice similarity coefficient; WI-HD, Williams' index based on Hausdorff distance.

*Significance level $P < 0.05$.

**Significance level $P < 0.001$.

Fig. 8 The Bland–Altman chart, demonstrating agreement between the computer-derived CSA for the CAT and the average observer-derived CSA, is shown in (a). Mean (solid) and 95% confidence interval (dashed) lines are shown on the chart. The correlation between computer-derived CSA and average observer-derived CSA is shown in (b) and includes the trend line (dashed). Corresponding comparisons for SAT CSA are shown in (c) and (d) and comparisons for VAT CSA are shown in (e) and (f).

segmentation, as was performed for the basis of comparison here, is most accurate, but laborious and time-consuming. Fully automated segmentation relieves the tedious nature of manual segmentation, but is not as accurate and is particularly susceptible to errors from anatomical variation or imaging artifacts.²¹ User-interaction in the process can be leveraged to explore this tradeoff, eliminating some of the time required for fully manual segmentation, but also allowing the user to apply expertise to the process to improve accuracy. Interactions can be applied at various times in the process, but often involve automated segmentation followed by user review.^{17,21} To aid in the CAT segmentation, a supervised approach was used. As outlined in Sec. 2.2.3, a user-defined boundary was created to guide the segmentation of the CAT, but the subsequent steps were automated.

The present study produced results aligned with those from other published studies when comparing volume calculations, but the method for manual determination of the adipose tissue was notably different. The manual identification of the adipose tissue in previous studies each included the user selection of some type of threshold.^{14,16–18} This manually determined threshold then played a role in producing the gold standard for comparison. However, the manual identification for the present study was built in custom software, providing a means for the user to have complete control over the segmentation process and manually outline all adipose tissue deposits directly. No processing of pixel data was performed to create the reference segmentations for comparison. This created a completely independent basis for comparison and evaluation of the automated approach, not reliant on calculations based on the image data.

Table 2 Summary of statistical results for volume comparisons.

Adipose tissue type (number of image comparisons)	Pearson correlation coefficient (R^2)	VD _A (%) (+/- standard deviation)	WI-VOL (95% confidence interval)
Cardiac ($n = 10$)	0.94*	-5.0 (+/- 16.1)	1.20 (0.63, 1.77)
Subcutaneous ($n = 10$)	1.00**	-2.0 (+/- 2.1)	0.79 (0.22, 1.35)
Visceral ($n = 10$)	0.99*	5.4 (+/- 4.4)	1.09 (0.00, 2.31)

VD_A, Mean difference in volume compared to average of manual observer volumes; WI-VOL, Williams' index based on volume.

*Significance level $P < 0.05$.

**Significance level $P < 0.001$.

The objective of an automated segmentation technique is to produce the clinical quantity of interest—in this case, the volume of adipose tissue across a specific section of the body. This was, indeed, the goal of the current study. Volume calculations and the underlying CSA computed on a slice-by-slice basis compare well to the measurements made by human observers. The 95% confidence intervals for the WI values computed for area all contain 1.0, indicating that the computer algorithm does not introduce additional variability as compared to that already present among a group of observers, i.e., the computer algorithm is at least as good as a human observer. In addition, the WI for both CAT and VAT was slightly above 1.0, indicating that the small deposits of adipose tissue around the heart and the small and disparate deposits of adipose tissue in the abdominal cavity, as represented by the Dixon MRI images, result in substantial variability among user interpretations. This being the case, a supervised, yet automated approach may in fact reduce some of this variability making it well suited for longitudinal studies. This same effect is also apparent in the WI data for volume calculations.

However, volume and CSA are derived metrics and can obscure errors in the specific segmentation results for any given slice. In addition, integrating CSA to compute volume across a series of slices can average out errors that occur on a single-slice basis that can be used as feedback to further improve the segmentation process. This effect is evident in the current results, as indicated by the stronger correlation in CAT volume versus CAT area—some of the variability in area agreement is neutralized by the process of integrating area into a single volume calculation. An additional objective of this study was to evaluate the performance of the segmentation algorithm in a manner identifying specific errors in single-slice results that can be utilized for improvement and evolution of the algorithm. Pursuant to that goal, additional evaluation metrics beyond area and volume were used for specific slice-to-slice comparisons, including DSC and Hausdorff distance. The WI values based on DSC suggest that comparisons by specific adipose pixel location, and the number of those pixels in agreement, result in slightly more variability in comparing the computer algorithm to the observer results than when comparing observer results to themselves. The same is true for Hausdorff-distance-based WI calculations, except for the CAT, where the

95% confidence interval did contain 1.0. This is evident of the increased variability in CAT manual tracing among the observers. The interobserver variability for SAT and VAT is reduced compared to that of CAT, resulting in slightly lower WI values. With respect to area and volume, the correlation between computer algorithm and average observer results are strongest for the SAT, indicating that the computer detects SAT consistently, but with a very small bias. When looking at the computer detection of SAT compared to the variability among the observers specifically, this small bias is apparent in the 95% confidence interval of the WI value which is slightly below 1.0. The correlation of computer-detected VAT and manually-traced VAT was strong, yet did indicate a small bias as the computer overestimated the area and volume. The computer algorithm identified small areas of fat deposits for the VAT that the observers did not interpret as bright enough or large enough to identify as a small deposit. This effect may be more a byproduct of the evaluation technique used—direct manual tracing—than the computer algorithm. While the observers were instructed to perform their interpretation and tracing of the images in such a way that if they were doubtful, to not include it in the traced area, some variability remains. These differences in VAT were also apparent in the WI values for slice-to-slice comparisons using DSC and Hausdorff distance. However, the strong correlations and good WI values for area and volume instill confidence in the technique for use in longitudinal studies. Figure 10 aids in visualization of the comparison between interobserver variability and computer-observer variability for all three adipose tissue types.

The present study used an approach similar to previously developed segmentation algorithms^{14,16–18,37} for identification of SAT and VAT, but extended the process to include detection of CAT. The CAT is of interest as the link between increased CAT volume and increased cardiovascular risk continues to clarify.^{4–6} Other preliminary work targeting CAT has been pursued using cardiac specific scans.^{12,19} The segmentation algorithm implemented and evaluated in the present study was targeted for a single, Dixon scan, useful for quantifying SAT, VAT, and CAT all concurrently, without the need for custom cardiac scans. More specific cardiac scans may aid in the CAT detection in the future, but no electrocardiogram-gating was used here. In addition, CAT is often divided into two types—epicardial (inside the visceral pericardium) and paracardial (between the visceral and parietal pericardia). However, it remains unclear whether MRI has the resolution to reliably differentiate the two³⁸ and the CAT was treated as one deposit in the current study.

While the segmentation results for the CAT were within the amount of variability of the group of three observers, the CAT data presented unique challenges for the algorithm and the observers related to the fat deposits near the base of the heart and the major blood vessels. Substantial deposits of adipose tissue were present around these vessels, but did not always appear at similar intensities as other deposits and were often more amorphous in shape (i.e., not simply a thin layer of fat on the surface of the myocardium, as in cross sections further toward the apex). Adipose tissue in this area was more difficult for the users to interpret, resulted in more variability among the observers, and presented challenges for the algorithm. This effect is apparent in Figs. 8(a) and 8(b). Studies on adipose tissue reduction via exercise or other intervention may implement protocols that exclude this region. In addition, future work on

Fig. 9 The Bland–Altman chart, demonstrating agreement between the computer-derived volume for the CAT and the average observer-derived volume, is shown in (a). Mean (solid) and 95% confidence interval (dashed) lines are shown on the chart. The correlation between computer-derived volume and average observer-derived volume is shown in (b) and includes the trend line (dashed). Corresponding comparisons for SAT volume are shown in (c) and (d) and for VAT volume comparisons in (e) and (f).

the algorithm itself, including its extension to a third dimension, may mitigate this effect.

For the binary morphological operations used throughout the algorithm, the size of the structuring elements was empirically determined based on a qualitative inspection of the 10 scans available for analysis. The largest necessary structuring element was chosen to ensure appropriate intermediate results. For example, a 10-pixel-radius disk structuring element was used on the lung mask to disconnect the lungs into two separate objects. A 10-pixel-radius was selected to ensure appropriate identification of two objects for each of the 10 subjects. Future work will include investigating the identification of the optimal size of the structuring element based on the

image data, which may improve the robustness and accuracy of the algorithm as more scans become available.

A number of other limitations of the current study are noteworthy. For volume-based comparisons, only 10 scans were used for direct evaluation. The number of scans was limited because they were collected as part of a small pilot study. Additional data acquisition is in process and the results from the evaluation of the segmentation algorithm presented here will be used to improve the approach in the larger data set. However, assuming the Dixon protocol is maintained, the *k*-means clustering algorithm may mitigate these issues and continue to identify the appropriate fat and water intensities as designed. Future work will investigate this idea and

Fig. 10 Box and whisker charts for comparison of interobserver variability and computer-observer variability. Data for DSC and Hausdorff distance (HD) are presented for each adipose tissue type. The dark gray boxes illustrate DSC and HD values for the computer algorithm and each of the three observers. The light gray boxes illustrate DSC and HD values for comparisons between observers one and two, observers one and three, and observers two and three. CAT data for DSC and HD are in (a) and (b), respectively. SAT data for DSC and HD are in (c) and (d). Finally, VAT data for DSC and HD are in (e) and (f).

modifications will be made to the algorithm accordingly. In addition, the manual tracing process was laborious and time intensive, particularly the VAT. Due to the time required for manual tracing of the VAT, single frames every 20 slices were selected for evaluation. The VAT volumes, therefore, are integrated from a relatively small number of area samples. With respect to the algorithm itself, no attempt was made to correct for partial volume effects. Other studies have assessed the effect and included corrections,¹⁴ but it was not included in the current work. Also, potential artifacts, such as phase error resulting from Eddy currents or Dixon-specific fat-water swap³⁹ were not examined. In addition, the parameters used for the segmentation algorithm, in the active contour, for the morphology, and for the N4ITK bias correction algorithm, were chosen empirically.

The main sources of error in the algorithm are likely attributable to the fact that the algorithm is based on the image data alone and does not include higher-level shape information. The addition of 3-D shape information may improve its robustness as the number of scans in the study grows. The addition of *a priori* shape information will be a main focus as the algorithm is evolved and improved over time. Other areas of future work will include a more robust analysis of the sensitivity to a wider array of image data and sensitivity to parameter selection in each component of the algorithm. Finally, no information was recorded regarding the timing of the processes to directly compare the time required for the computer algorithms versus manual tracing. The segmentation algorithm was implemented using an interpreted programming language designed for prototyping (Python) and was therefore not designed for high-performance. Upon determination of adequate accuracy, the algorithm can be implemented in a compiled language (e.g., C++) designed for high-performance.

In conclusion, we have developed a segmentation algorithm to identify CAT, SAT, and VAT from Dixon MRI scans and evaluated it within the context of interobserver variability. The results indicate that the segmentation method produces results similar to that of an observer when manually tracing the adipose tissue deposits. The method can be used for future longitudinal studies of obesity intervention.

Disclosures

The authors have no conflicts of interest relating to this paper.

Acknowledgments

This work was supported by an SIUE Seed Grants for Transitional and Exploratory Projects (STEP) grant.

References

1. S. J. Olshansky et al., "A potential decline in life expectancy in the United States in the 21st century," *N. Engl. J. Med.* **352**(11), 1138–1145 (2005).
2. E. L. Thomas et al., "Whole body fat: content and distribution," *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 56–80 (2013).
3. M. J. Lee, Y. Wu, and S. K. Fried, "Adipose tissue heterogeneity: implication of depot differences in adipose tissue for obesity complications," *Mol. Aspects Med.* **34**(1), 1–11 (2013).
4. K. A. Britton and C. S. Fox, "Ectopic fat depots and cardiovascular disease," *Circulation* **124**(24), e837–e841 (2011).
5. M. A. Cornier et al., "Assessing Adiposity: a scientific statement from the American Heart Association," *Circulation* **124**(18), 1996–2019 (2011).

6. J. P. Despres, "Body fat distribution and risk of cardiovascular disease," *Circulation* **126**(10), 1301–1313 (2012).
7. G. Bastarrika et al., "Relationship between coronary artery disease and epicardial adipose tissue quantification at cardiac CT: comparison between automatic volumetric measurement and manual bidimensional estimation," *Acad. Radiol.* **17**(6), 727–734 (2010).
8. R. Nakazato et al., "Interscan reproducibility of computer-aided epicardial and thoracic fat measurement from noncontrast cardiac CT," *J. Cardiovasc. Comput. Tomogr.* **5**(3), 172–179 (2011).
9. J. V. Spearman et al., "Automated quantification of epicardial adipose tissue using CT angiography: evaluation of a prototype software," *Eur. Radiol.* **24**(2), 519–526 (2014).
10. J. V. Spearman et al., "Prognostic value of epicardial fat volume measurements by computed tomography: a systematic review of the literature," *Eur. Radiol.* **25**(11), 3372–3381 (2015).
11. P. Bornert et al., "Water/fat-resolved whole-heart Dixon coronary MRA: an initial comparison," *Magn. Reson. Med.* **71**(1), 156–163 (2014).
12. R. Homsí et al., "3D-Dixon MRI based volumetry of peri- and epicardial fat," *Int J Cardiovasc Imaging* **32**(2), 291–299 (2016).
13. W. T. Dixon, "Simple proton spectroscopic imaging," *Radiology* **153**(1), 189–194 (1984).
14. B. T. Addeman et al., "Validation of volumetric and single-slice MRI adipose analysis using a novel fully automated segmentation method," *J. Magn. Reson. Imaging* **41**(1), 233–241 (2015).
15. A. Alabousi et al., "Evaluation of adipose tissue volume quantification with IDEAL fat-water separation," *J. Magn. Reson. Imaging* **34**(2), 474–479 (2011).
16. V. Positano et al., "An accurate and robust method for unsupervised assessment of abdominal fat by MRI," *J. Magn. Reson. Imaging* **20**(4), 684–689 (2004).
17. G. Thormer et al., "Software for automated MRI-based quantification of abdominal fat and preliminary evaluation in morbidly obese patients," *J. Magn. Reson. Imaging* **37**(5), 1144–1150 (2013).
18. D. Wald et al., "Automatic quantification of subcutaneous and visceral adipose tissue from whole-body magnetic resonance images suitable for large cohort studies," *J. Magn. Reson. Imaging* **36**(6), 1421–1434 (2012).
19. A. Cristobal-Huerta et al., "Automated quantification of epicardial adipose tissue in cardiac magnetic resonance imaging," in *Int. Conf. IEEE Eng. Med. and Biol. Soc. (EMBC)*, pp. 7308–7311 (2015).
20. S. Flüchter et al., "Volumetric assessment of epicardial adipose tissue with cardiovascular magnetic resonance imaging," *Obesity* **15**(4), 870–878 (2007).
21. S. Bonekamp et al., "Quantitative comparison and evaluation of software packages for assessment of abdominal adipose tissue distribution by magnetic resonance imaging," *Int. J. Obes. (Lond)* **32**(1), 100–111 (2008).
22. V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Trans. Med. Imaging* **16**(5), 642–652 (1997).
23. J. D. Klingensmith, R. Shekhar, and D. G. Vince, "Evaluation of three-dimensional segmentation algorithms for the identification of luminal and medial-adventitial borders in intravascular ultrasound images," *IEEE Trans. Med. Imaging* **19**(10), 996–1011 (2000).
24. J. D. Klingensmith et al., "Validation of an automated system for luminal and medial-adventitial border detection in three-dimensional intravascular ultrasound," *Int. J. Cardiovasc. Imaging* **19**(2), 93–104 (2003).
25. W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imaging* **25**(11), 1451–1461 (2006).
26. A. L. Elliott and J. D. Klingensmith, "SIUE-BiomedicalImaging ResearchLab/SIUE-Dixon-Fat-Segmentation-Algorithm: v1.0.1 Released," Version 1.0.1 (03 April 2018).
27. S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: a structure for efficient numerical computation," *Comput. Sci. Eng.* **13**(2), 22–30 (2011).
28. S. van der Walt et al., "scikit-image: image processing in Python," *Peer J.* **2**, e453 (2014).
29. J. D. Hunter, "Matplotlib: a 2D graphics environment," *Comput. Sci. Eng.* **9**(3), 90–95 (2007).
30. G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media, Inc., Beijing (2008).

31. Z. Yaniv et al., "SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research," *J. Digital Imaging* **31**(3), 290–303 (2018).
32. N. J. Tustison et al., "N4ITK: improved N3 bias correction," *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010).
33. T. Kanungo et al., "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 881–892 (2002).
34. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. J. Comput. Vision* **1**(4), 321–331 (1988).
35. J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **327**(8476), 307–310 (1986).
36. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(9), 850–863 (1993).
37. J. Kullberg et al., "Automated and reproducible segmentation of visceral and subcutaneous adipose tissue from abdominal MRI," *Int. J. Obes. (Lond)* **31**(12), 1806–1817 (2007).
38. J. D. Klingensmith et al., "Automated segmentation of cardiac adipose tissue in Dixon magnetic resonance images," *J. Biomed. Graphics Comput.* **8**(1), 1–13 (2018).
39. B. Glocker et al., "Correction of fat-water swaps in Dixon MRI," *Lect. Notes Comput. Sci.* **9902**, 536–543 (2016).

Jon D. Klingensmith is an assistant professor in the Department of Electrical and Computer Engineering at Southern Illinois University Edwardsville. He received his PhD in biomedical engineering from Case Western Reserve University in 2003. He was a project staff at the Cleveland Clinic Foundation from 2003 to 2004 and held various positions in R&D at Volcano Corporation from 2004 to 2012. His current research interests include 3-D medical image analysis and ultrasonic tissue characterization.

Addison L. Elliott is currently working on his MS degree in electrical engineering at Southern Illinois University Edwardsville, where he

obtained his BS degree in 2017. Currently, he is a research assistant in the biomedical imaging research lab at SIUE. His research interests include magnetic resonance imaging, ultrasound, and medical image processing.

Amy H. Givan is currently working on her MS degree in exercise physiology at Southern Illinois University Edwardsville. She received her BS degree in kinesiology in 2017. Her current research interests include adipose tissue, metabolic disorders, and magnetic resonance imaging.

Zechariah D. Faszold is an electrical engineering student at Southern Illinois University Edwardsville. His research interests include cardiac adipose segmentation using magnetic resonance imaging and spectral analysis using ultrasonography. He is a licensed paramedic in Southwestern Illinois.

Cory L. Mahan received his BS degree in exercise science from Southern Illinois University Edwardsville. Currently, he is pursuing his MS degree in exercise physiology, serving as a research assistant and lab coordinator. His research interests include the effects of resistance training and aerobic exercise on cardiovascular function.

Adam M. Doedtman received his BS degree in exercise science from Southern Illinois University Edwardsville in 2018. He is a first-year student enrolled at Southern Illinois School of Medicine in Carbondale, Illinois. His research interests include the effects of resistance training and aerobic exercise on cardiovascular function.

Maria Fernandez-del-Valle is an assistant professor in the Department of Applied Health at Southern Illinois University Edwardsville. She received her degree in exercise science in 2004 and her PhD with an emphasis in clinical exercise physiology from the Universidad Europea de Madrid in 2012. Her research interests include exercise, body composition, and cardio-metabolic health with a strong focus on understanding the effects of different interventions on reducing fat deposits and improving physical fitness.