Automated Segmentation of Visceral, Deep Subcutaneous, and Superficial Subcutaneous Adipose Tissue Volumes in MRI of Neonates and Young Children

Yeshe Manuel Kway, BEng • Kashthuri Thirumurugan, MSc • Mya Thway Tint, PhD • Navin Michael, PhD • Lynette Pei-Chi Shek, MMed • Fabian Kok Peng Yap, MMed • Kok Hian Tan, MMed • Keith M. Godfrey, FMedSci • Yap Seng Chong, MD • Marielle Valerie Fortier, MD • Ute C. Marx, Dr rer nat habil • Johan G. Eriksson, DMSc • Yung Seng Lee, PhD • S. Sendhil Velan, PhD • Mengling Feng, PhD* • Suresh Anand Sadananthan, PhD*

From the Singapore Institute for Clinical Sciences (Y.M.K., K.T., M.T.T., N.M., L.P.C.S., Y.S.C., M.V.F., J.G.E., Y.S.L., S.S.V., S.A.S.) and Institute of Bioengineering and Bioimaging (S.S.V.), Agency for Science Technology and Research, 30 Medical Dr, Singapore 117609; Departments of Medicine (Y.M.K., J.G.E.), Obstetrics and Gynaecology (M.T.T., Y.S.C.), and Pediatrics (L.P.C.S., Y.S.L.), Yong Loo Lin School of Medicine, National University of Singapore, Singapore; Khoo Teck Puat-National University Children's Medical Institute, National University Hospital, Singapore (L.P.C.S., Y.S.L.); Departments of Pediatric Endocrinology (F.K.P.Y.), Obstetrics and Gynaecology (K.H.T.), and Diagnostic and Interventional Imaging (M.V.F.), KK Women's and Children's Hospital, Singapore; Pediatrics Academic Clinical Programme (F.K.P.Y.), Academic Medicine (K.H.T.), Duke-National University of Singapore Medical School, Singapore (F.K.P.Y., K.H.T.); Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore (F.K.P.Y.); Medical Research Council Lifecourse Epidemiology Unit and National Institute for Health Research Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton National Health Service Foundation Trust, Southampton, England (K.M.G.); School of Engineering, Pforzheim University, Pforzheim, Germany (U.C.M.); Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland (J.G.E.); Folkhälsan Research Center, Helsinki, Finland (J.G.E.); and Saw Swee Hock School of Public Health, National University Health System and National University of Singapore, Singapore (M.F.); Institute of Data Science, National University of Singapore, Singapore (M.F.). Received December 31, 2020; revision requested February 9, 2021; revision received June 1; accepted July 12. **Address correspondence to** S.A.S. (e-mail: suresh@scics.a-star.edu.sg).

Supported in part by the Singapore National Research Foundation under its Translational and Clinical Research (TCR) Flagship Programme and administered by the Singapore Ministry of Health National Medical Research Council, Singapore (NMRC/TCR/004-NUS/2008; NMRC/TCR/012-NUHS/2014). Additional funding is provided by the Singapore Institute for Clinical Sciences, Agency for Science Technology and Research, Singapore.

* M.F. and S.A.S. are co-senior authors.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2021; 3(5):e200304 • <https://doi.org/10.1148/ryai.2021200304> • Content codes: **AI** **PD**

Purpose: To develop and evaluate an automated segmentation method for accurate quantification of abdominal adipose tissue (AAT) depots (superficial subcutaneous adipose tissue [SSAT], deep subcutaneous adipose tissue [DSAT], and visceral adipose tissue [VAT]) in neonates and young children.

Materials and Methods: This was a secondary analysis of prospectively collected data, which used abdominal MRI data from Growing Up in Singapore Towards healthy Outcomes, or GUSTO, a longitudinal mother-offspring cohort, to train and evaluate a convolutional neural network for volumetric AAT segmentation. The data comprised imaging volumes of 333 neonates obtained at early infancy (age ≤ 2 weeks, 180 male neonates) and 755 children aged either 4.5 years ($n = 316$, 150 male children) or 6 years ($n = 439$, 219 male children). The network was trained on images of 761 randomly selected volumes (neonates and children combined) and evaluated on 100 neonatal volumes and 227 child volumes by using 10-fold validation. Automated segmentations were compared with expert-generated manual segmentation. Segmentation performance was assessed using Dice scores.

Results: When the model was tested on the test datasets across the 10 folds, the model had strong agreement with the ground truth for all testing sets, with mean Dice similarity scores for SSAT, DSAT, and VAT, respectively, of 0.960, 0.909, and 0.872 in neonates and 0.944, 0.851, and 0.960 in children. The model generalized well to different body sizes and ages and to all abdominal levels.

Conclusion: The proposed segmentation approach provided accurate automated volumetric assessment of AAT compartments on MR images of neonates and children.

Clinical trial registration no. NCT01174875

Supplemental material is available for this article.

© RSNA, 2021

Childhood obesity, characterized by increased fat accumulation and accelerated weight gain in early life, is associated with adulthood obesity and elevated cardiometabolic risks (1–4). In adults, individual abdominal adipose tissue (AAT) depots, namely superficial subcutaneous adipose tissue (SSAT), deep subcutaneous adipose tissue (DSAT), and visceral adipose tissue (VAT), exhibit differences in terms of their anatomic location, adipocyte size, and association with cardiometabolic risks (5–9). Levels of VAT

and DSAT are strongly associated with type 2 diabetes and cardiovascular disease in adults, although SSAT has been reported to be a putatively distinct protective fat storage depot (5,7,9). However, in children, some studies have reported subcutaneous adipose tissue (SAT) to be more strongly linked to cardiometabolic risk factors than VAT (10). Furthermore, sex and ethnic differences in abdominal fat partitioning observed in adults have been shown to manifest as early as infancy and early childhood (11–13).

Abbreviations

AAT = abdominal adipose tissue, DSAT = deep subcutaneous adipose tissue, GUSTO = Growing Up in Singapore Towards healthy Outcomes, SAT = subcutaneous adipose tissue, SSAT = superficial SAT, 2D = two dimensional, VAT = visceral adipose tissue

Summary

Accurate and automated segmentation of abdominal subcutaneous (deep and superficial) and visceral fat depots in neonates and young children was performed by using a deep learning model on two-dimensional MR images.

Key Points

- The proposed segmentation approach showed strong agreement with expert-generated manual segmentation, with Dice similarity coefficients for superficial subcutaneous adipose tissue (SSAT), deep subcutaneous adipose tissue (DSAT), and visceral adipose tissue (VAT), respectively, of 0.960, 0.909, and 0.872 in neonates and 0.944, 0.851, and 0.960 in children.
- The model segmentation resulted in mean overestimation rates of 1.6% and 0.6% and a mean underestimation rate of 1% in neonates and resulted in mean underestimation rates of 3.7% and 0.3% and a mean overestimation rate of 7.2% in children for SSAT, VAT, and DSAT, respectively.
- Segmentation accuracy for SSAT and DSAT was higher in neonates (96% and 91%, respectively) than in children (94% and 85%, respectively; $P < .001$ for both); the accuracy for VAT segmentation was higher in children than in neonates (0.96 vs 0.87, $P < .001$).

Keywords

Pediatrics, Deep Learning, Convolutional Neural Networks, Water-Fat MRI, Image Segmentation, Deep and Superficial Subcutaneous Adipose Tissue, Visceral Adipose Tissue

Hence, quantification of individual AATs in early life is important for understanding the metabolic sequelae of obesity.

The anatomy of the AAT compartments in neonates and young children differs from one another, as there are rapid changes in body composition early in life. Earlier work predominantly focused on segmentation of total AAT, SAT, VAT (14,15), and muscle volumes (16) by using conventional image processing techniques such as clustering, thresholding, volume registration, active contours, and multiatlas-based methods. A comprehensive overview of imaging and segmentation methods is described elsewhere (17,18). Segmentation of DSAT and SSAT is challenging, as the fascia separating them has a very subtle and discontinuous appearance. Hence, to our knowledge, there have been only two attempts at automatic segmentation of these depots, which were made by using dynamic programming to delineate DSAT only in the posterior abdominal region (19) and level sets over the entire abdomen (20).

Convolutional neural networks have become the state of the art for image segmentation (21). Two-dimensional (2D) models on abdominal CT (22) and MRI (23,24) and three-dimensional models on whole-body MRI (25) volumes have shown their successful application for SAT-VAT quantification. FatSegNet uses three-dimensional convolutions to aggregate 2D convolutional outputs from coronal, axial, and sagittal views to address poor single-view resolution and misclassification of the upper extremities (23). Although these deep learning methods focused

on AAT segmentation in adults, automated applications in the pediatric population are limited because of challenges in pediatric MRI. This work presents a deep learning method that uses a convolutional neural network for volumetric segmentation of SSAT, DSAT, and VAT in neonates and young children.

Materials and Methods

Study Participants

This was a secondary analysis of prospectively collected data. Data were obtained from children in the Growing Up in Singapore Towards healthy Outcomes (GUSTO), a prospective longitudinal mother-offspring cohort (26). Briefly, the GUSTO study recruited pregnant women in their first trimester between June 2009 and September 2010. A total of 1201 babies were born. Abdominal MR images (liver dome to upper sacrum) were obtained with parental consent in neonates (age ≤ 2 weeks) ($n = 333$) and children aged 4.5 years ($n = 316$) and 6 years ($n = 439$); participants had a wide weight range and were of three ethnicities (Table 1). The study flowchart is shown in Figure E1 (supplement). A total of 337 participants underwent MRI at more than one imaging time point, and 91 children underwent MRI at all three time points. The study was approved by the institutional review board of the Singapore National Health Care Group and by the centralized institutional review board of SingHealth. Parents of all participants gave written consent.

MRI Protocol

While being monitored by a neonatologist, neonates were placed within an immobilization device while naturally asleep on an adult head coil of a GE Signa 1.5-T MR imager. Axial abdominal images were obtained by using a T1-weighted water-suppressed fast spin-echo sequence (13). The data at ages 4.5 and 6 years were acquired by using a water-suppressed half-Fourier acquisition single-shot axial turbo spin-echo sequence and a body matrix coil on a Siemens Skyra 3-T imager (Table 1).

Ground Truth Labeling

Earlier attempts at segmentation of these data (in 333 neonates and 316 children aged 4.5 years) used watershed-based (13) and graph-theoretical (12) techniques to segment SAT and VAT. DSAT-SSAT delineation was performed by manually tracing the boundary along the fascia superficialis and was based on anatomic prior knowledge in regions where the fascia has a subtle and discontinuous appearance. Figure E2 (supplement) shows the fat depots in a neonate and a 6-year-old child. The resulting masks were manually edited for missegmentations and were used as the ground truth. Manual refinements were performed by two trained MRI readers (M.T.T. with 10 years of experience in MRI and a physician with 20 years of experience in MRI) for neonatal volumes and by one reader (K.T., 2 years of experience in MRI) for child volumes by using ITK-Snap software (www.itksnap.org). The mean interrater coefficients of variation

Table 1: Participant Characteristics at Different Imaging Time Points and Corresponding MRI Information in GUSTO Cohort Study

Parameter	≤2 Weeks of Age (<i>n</i> = 333)	4.5 Years of Age (<i>n</i> = 316)	6 Years of Age (<i>n</i> = 439)
Participant characteristics			
Ethnicity			
Chinese	145 (44)	158 (50)	226 (51)
Malay	127 (38)	94 (30)	140 (32)
Indian	61 (18)	64 (20)	73 (17)
Age at MRI	9.7 days ± 2.8	4.6 years ± 0.1	6.0 years ± 0.1
Male sex	180 (54)	150 (47)	219 (49)
Weight (kg)	3.2 ± 0.4	17.5 ± 3.0	21.0 ± 4.5
Height (cm)	50.0 ± 2.1	105.5 ± 4.2	115.4 ± 5.0
BMI (kg/m ²)	12.8 ± 1.2	15.7 ± 1.9	15.7 ± 2.4
zBMI*	-0.13 ± 1.1	0.24 ± 1.2	0.06 ± 1.3
Abdominal circumference (cm)	29.3 ± 2.5	51.8 ± 4.9	54.1 ± 6.8
Imaging information			
Imaging system	GE Signa 1.5-T MR imager	Siemens Skyra 3-T MR imager	Siemens Skyra 3-T MR imager
Imaging sequence	T1-weighted water-suppressed axial fast SE	T2-weighted water-suppressed half-Fourier RARE	T2-weighted water-suppressed half-Fourier RARE
Repetition time (msec)	600	1000	1000
Echo time (msec)	7	95	95
Coil type	Adult head coil	Body matrix coil	Body matrix coil
Imaging volume (liver dome to upper sacrum)	15–24 images	30–45 images	30–45 images
Image dimensions	256 × 256 to 512 × 512 (interpolated)	320 × 320	320 × 320
Image resolution (mm)	0.43 × 0.43 × 5	0.94 × 0.94 × 5	0.94 × 0.94 × 5
Acquisition time (min)	4 minutes 37 seconds	1 minute	1 minute

Note.—Data are shown as the mean ± standard deviation or the number of participants with the percentage in parentheses. BMI = body mass index, GUSTO = Growing Up in Singapore Towards healthy Outcomes, RARE = rapid acquisition with relaxation enhancement, SE = spin echo, zBMI = BMI *z* score.

*BMI standardized for age and sex based on World Health Organization child growth standards (27).

were 1.57, 3.23, and 2.06 for neonates (13), and the mean intrarater Dice similarity coefficients were 0.98, 0.95, and 0.99 for children aged 4.5 years (12) and 0.92, 0.96, and 0.99 for children aged 6 years for SSAT, DSAT, and VAT, respectively.

Segmentation Model

A U-Net–based (28) network that used a VGG-16 (University of Oxford) (29) encoder was adopted and trained (Fig 1). Weights for all nonpretrained layers were initialized by using Glorot uniform initialization (30). The segmentation task was defined as a voxelwise classification into background, SSAT, DSAT, and VAT. The final convolution was designed to output a tensor with the dimensions of 512 × 512 × 4, which was followed by SoftMax activation. The input image was additionally concatenated to the output of the second-to-last convolutional layer to achieve finer reconstruction. Batch normalization was added before applying all nonlinear transformations (31). To preserve the original image dimensions, zero padding was applied for all convolutions.

Model Evaluation

To ensure statistical stability, the following experiment was performed 10 times (10 folds). For every fold, the test set was formed by first extracting 30% of the child (227 of 755) and neonatal (100 of 333) volumes. Subsequently, training and validation sets (70:30) were created by extracting all images from the remaining volumes (total, 22618; children, 17992; neonates, 2426) for each experimental fold. Children's images were zero padded to match the neonatal image dimensions. All images were individually normalized to a scale of 0–1. The model was implemented in TensorFlow, version 1.14.0, and Keras, version 2.2.5, and was trained on an Nvidia Tesla V100-SXM2 graphics processing unit. The inference runtime was assessed with an Intel Core i7 processor. The batch size and learning rate were tuned empirically. Models were trained by optimizing a summation of soft Dice losses over all individual classes, $\text{Loss} = -\sum_c \left(\frac{2 \times |P_c \cap GT_c| + \epsilon}{|P_c| + |GT_c| + \epsilon} \right)$, where P_c contains pixelwise prediction probabilities for class c , GT_c represents the complementary ground truth, and ϵ is a constant (0.001) to prevent division by zero (32). All models were trained over several epochs with a batch size of 14 and a

Figure 1: U-Net-based architecture using a pretrained VGG-16 encoder (University of Oxford) in its path. The input image is copied three times to the encoder requirements. Additionally, the input image is concatenated to the output of the second-to-last convolutional layer to achieve finer reconstruction. Batch normalization is applied to all layers before nonlinear transformations are performed, and zero padding is applied for all convolutional operations. Bilinear interpolation is used as the upsampling operation. Conv = convolution, DSAT = deep subcutaneous adipose tissue, ReLu = rectified linear unit, SSAT = superficial subcutaneous adipose tissue, VAT = visceral adipose tissue.

learning rate of 0.0001 by using an Adam optimizer (33). Training was stopped when the validation loss did not improve over four consecutive epochs (Fig E3 [supplement]). Volumes of individual fat depots were computed by summing the number of voxels in each class and multiplying by the voxel resolution. Figure 2 shows the entire workflow.

Statistical Analysis

In consideration of the differences in abdominal shape and size between neonates and children, evaluation was performed by computing the mean performance over the 10 folds separately for neonatal ($n = 100$) and child ($n = 227$) volumes. The segmentation overlaps between the predicted and ground truth volumes were evaluated by computing the Dice scores, with 0 indicating no overlap and 1 representing 100% overlap. The false-positive rate (the number of predicted voxels assigned to a class that have a true label belonging to another class) and false-negative rate (the proportion of true labeled ground truth voxels for which the model predicted a different class) were also computed. Metrics were computed on entire volumes for all individual fat classes. To assess volumetric agreement with the ground truth, Bland-Altman plots (the mean of ground truth and predicted volumes plotted against their difference) were computed for one randomly selected model (model 1). One-way analysis of variance and a Tukey post hoc test were performed to evaluate whether there were differences in Dice

scores of DSAT, SSAT, and VAT within neonates and children. An independent-sample t test was performed to check for Dice score differences in each fat depot between neonates and children. A P value of less than .05 was considered to indicate a significant difference. Analyses were performed by using Python (version 3.6) and SPSS Statistics (version 26, IBM).

Code Availability

The code is publicly available at https://github.com/YesheKway/Pediatric_Abdominal_Fat_Seg.

Results

Participant and Imaging Characteristics

Abdominal MR image volumes from 333 neonates (180 male [54%]) aged 2 weeks or younger, 316 children (150 male [47%]) aged 4.5 years, and 439 children (219 male [50%]) aged 6 years were used for algorithm development and testing. Detailed characteristics of the participants and the imaging parameters are listed in Table 1.

Overall Performance

Evaluation of the proposed method against the ground truth is summarized in Table 2 and Table E1 (supplement). The Dice scores across all 10 folds had narrow ranges and small standard deviations over all AAT classes in both neonates and children.

Figure 2: Workflow of the proposed method for volumetric assessment of individual abdominal adipose tissue compartments, namely visceral adipose tissue (VAT), deep subcutaneous adipose tissue (DSAT), and superficial subcutaneous adipose tissue (SSAT). **(1)** Extraction of region of interest. **(2)** Application of the model on all individual image sections of the extracted volume. **(3)** Deriving binary segmentation masks for each class by using the ArgMax application. **(4)** Volume computation by summing the number of segmented voxels in each class and multiplying by the voxel resolution.

The presented model achieved performance identical to that of models that were trained solely on individual time point data (Table E2 [supplement]).

The mean Dice scores over the 10 folds for SSAT, DSAT, and VAT, respectively, were 0.960, 0.909, and 0.872 for neonates and 0.944, 0.851, and 0.960 for children; the differences among the mean Dice scores for all the fat depots within neonates and children and for each fat depot as compared between neonates and children were significant ($P < .001$ for all). The performance variation between the models is illustrated by the Dice score ranges in Table 2. Overall, all models achieved good performance for all AAT classes. Bland-Altman plots for model 1 are shown in Figure 3. Overall, the under- and overestimation rates for SSAT, DSAT, and VAT in neonates as well as in children were within the 95% limits of agreement. Figure 4 shows a qualitative representation of the accurate segmentation from participants across a range of body sizes and shapes.

Accurate segmentation was observed on images with substantial motion artifacts (Fig 5C, 5D) and images in which the upper extremities were located in close proximity to the abdominal wall (Fig 5E, 5F), with the latter being a well-known challenge for accurate abdominal fat segmentation. The results also indicated that the model had learned important anatomical features to make correct predictions for highly ambiguous regions where, because of motion artifacts, the intraabdominal content was shifted toward the anterior portion into the umbilicus (Fig 5A)

or overlapped with the SSAT layer (Fig 5B). The model correctly distinguished intraabdominal content from SSAT and estimated its actual curve progression.

The inference time for processing an abdominal MRI volume took approximately 27 seconds on the Tesla V100-SXM2 graphics processing unit and took 73 seconds on the standard Intel Core i7 processor, whereas manual segmentation is highly dependent on the rater and can take up to several hours.

Performance on Neonates

The segmentation accuracy for VAT (Dice score, 0.872) was lower than the accuracies for DSAT (Dice score, 0.909; $P < .001$) and SSAT (Dice score, 0.960; $P < .001$) and demonstrated slightly increased false-positive and false-negative rates of 13.8% and 12.3%, respectively. Bland-Altman plots showed marginal overestimation of SSAT (mean difference, $1.47 \text{ cm}^3 \pm 1.59$) and VAT (mean difference, $0.17 \text{ cm}^3 \pm 1.57$). The mean overestimation rates of SSAT and VAT were 1.6% and 0.6% of their mean fat volumes. The model tended to slightly underestimate DSAT (mean difference, $0.14 \text{ cm}^3 \pm 1.28$) volumes. The mean underestimation rate for DSAT was 1% of the total DSAT volume. Across participants with high variation in body fat accumulation, most of the computed differences were within the 95% limits of agreement, with only five outliers being identified for VAT and SSAT and four outliers being identified for DSAT.

Table 2: Mean Evaluation Results and Segmentation Performance

Parameter	FP (%)	FN (%)	Dice Similarity Coefficient	GT (cm ³)	P (cm ³)	Difference (%)*
Tenfold validation model						
Neonates						
SSAT	4.7 ± 0.3	3.3 ± 0.3	0.960 ± 0.001 (0.959–0.961)	NA	NA	NA
DSAT	9.1 ± 0.6	9.0 ± 0.7	0.909 ± 0.002 (0.906–0.913)	NA	NA	NA
VAT	13.8 ± 1.4	12.3 ± 1.0	0.872 ± 0.002 (0.868–0.876)	NA	NA	NA
Children						
SSAT	6.0 ± 0.9	5.3 ± 0.9	0.944 ± 0.002 (0.940–0.947)	NA	NA	NA
DSAT	15.0 ± 3.3	14.6 ± 1.8	0.851 ± 0.010 (0.840–0.857)	NA	NA	NA
VAT	3.9 ± 0.6	4.1 ± 0.5	0.960 ± 0.004 (0.950–0.963)	NA	NA	NA
Single model						
Neonates (n = 100)						
SSAT	4.9 ± 1.2	3.2 ± 1.2	0.960 ± 0.009	91.0 ± 24.2	92.5 ± 24.7	1.6
DSAT	8.8 ± 3.9	9.2 ± 4.2	0.911 ± 0.024	14.7 ± 6.4	14.6 ± 6.2	-0.7
VAT	14.0 ± 9.6	12.0 ± 3.3	0.872 ± 0.034	29.5 ± 10.2	29.7 ± 9.9	0.7
Children (n = 227)						
SSAT	3.7 ± 2.6	7.5 ± 2.7	0.943 ± 0.013	498.7 ± 428.2	480.1 ± 417.4	-3.8
DSAT	23.2 ± 13.1	10.3 ± 5.9	0.844 ± 0.049	224.5 ± 316.0	240.8 ± 322.3	7.0
VAT	4.2 ± 3.0	4.8 ± 3.6	0.955 ± 0.025	210.2 ± 111.9	209.7 ± 116.5	-0.2

Note.—The GT volumes, P volumes, and the differences between the two are only shown for the single model. For the 10-fold validation model, the Dice scores are shown as the mean ± standard deviation, with the range in parentheses. All other values are shown as mean ± standard deviation, unless otherwise specified. DSAT = deep subcutaneous adipose tissue, FN = false-negative rate, FP = false-positive rate, GT = ground truth, NA = not applicable, P = prediction, SSAT = superficial subcutaneous adipose tissue, VAT = visceral adipose tissue.

*Difference between P and GT volumes.

Performance on 4.5-Year-Old and 6-Year-Old Children

High segmentation accuracy was seen for SSAT and VAT, with mean Dice scores of 0.944 and 0.960, respectively. There was a slightly reduced accuracy for DSAT (Dice score, 0.85 [$P < .001$]) compared with the accuracy for SSAT and VAT, with false-positive and false-negative rates of 15.0% and 14.6%, respectively. The highest performance variation was found for DSAT predictions, with a difference of 0.017 in the Dice score between the highest (Dice score, 0.857) and lowest (Dice score, 0.840) level of performance being shown. Bland-Altman plots indicated a slight underestimation of SSAT volumes (mean difference, $18.39 \text{ cm}^3 \pm 27.16$) and VAT volumes (mean difference, $0.53 \text{ cm}^3 \pm 14.01$) and a slight overestimation of DSAT volumes (mean difference, $16.1 \text{ cm}^3 \pm 26.27$). The mean underestimation rates for SSAT and VAT were 3.7% and 0.3% of their total fat volumes. The mean DSAT overestimation rate was 7.2% of the total DSAT volume. Across all AAT classes, almost all data points were identified as being within the 95% limits of agreement. There were a total of eight, nine, and 10 outliers for DSAT, SSAT, and VAT, respectively. However, absolute errors for all outli-

ers were mainly found below 12% with respect to the mean of the predicted volume and ground truth volume.

Discussion

We present a convolutional neural network for volumetric segmentation of SSAT, DSAT, and VAT in neonates and young children. The proposed method shows strong agreement with the ground truth for all adipose tissue depots. Furthermore, the model demonstrates good generalization with respect to high population heterogeneity and different imaging protocols.

The developed model enabled delineation of SAT into DSAT and SSAT. Differential segmentation of these two fat compartments is important because they exhibit different biological and morphological characteristics (7–9). As underscored by other studies, delineation of DSAT and SSAT is challenging because of the subtle and discontinuous appearance of the fascia superficialis on most MR images (34). Our results showed that the segmentation accuracy (Dice score) for SSAT and DSAT was higher in neonates (0.960 [SSAT] and 0.909 [DSAT]) than in children (0.944 [SSAT] and 0.851 [DSAT]) ($P < .001$ for both comparisons). These differences are likely attributable to the fact that DSAT

Figure 3: Bland-Altman plots show the volume difference (in cubic centimeters) between the ground truth (GT) and model prediction (P) for visceral adipose tissue (VAT), deep subcutaneous adipose tissue (DSAT), and superficial subcutaneous adipose tissue (SSAT) on the neonatal (left) and child (right) testing sets (100 and 227 volumes, respectively). mean diff = mean difference, SD = standard deviation.

delineation is more challenging in child volumes, as imaging volumes comprise around twofold more image sections, over which DSAT is distributed with increased variation. Additionally, the high standard deviation of DSAT in children (mean, $224.5 \text{ cm}^3 \pm 316.0$) indicates high volumetric interparticipant variation, which makes it more difficult for the model to generalize. There was higher segmentation accuracy for VAT in children than in neonates (Dice score, 0.960 vs 0.872; $P < .001$), which may be due to the fact that neonates have lower levels of VAT than children. The differences in VAT levels in neonates may affect the segmentation accuracy, as their fat depots appear as very small globs with varying intensities that are sparsely spread over the abdominal cavity.

The reference standard Dixon water-fat imaging (35) works only under breath-hold conditions; hence, it is not feasible in young children. Free-breathing imaging results in a decreased contrast difference among fat classes and introduces respiratory artifacts. Earlier fat segmentation approaches relied on

conventional image processing techniques, such as intensity thresholding for fat identification (22) or removal of arms (24), before applying convolutional neural networks. Hence, these approaches were affected by the limitations of conventional image processing methods, such as difficulty in correctly discriminating adjacent tissue compartments with similar intensities or identifying sharp intensity changes. These limitations make these approaches unsuitable for pediatric MRI, as they are prone to failure in the presence of increased respiratory artifacts, as well as being prone to contrast and intensity variation.

Misclassification of closely located upper extremities is a well-known challenge for AAT segmentation. As neonates and children may move a lot during image acquisition, the appearance of arms is common in pediatric imaging. FatSegNet was the first deep learning approach to achieve AAT segmentation on MRI data without preprocessing steps (23). FatSegNet addressed the problem of the misclassification of the arms by providing their

Figure 4: Abdominal adipose tissue segmentation into visceral (yellow), deep subcutaneous (blue), and superficial subcutaneous (red) adipose tissue depots. **(A–D)** Segmentation examples of neonatal abdomens (age ≤ 2 weeks) of different shapes. **(E–H)** Representative MR image sections at the level of umbilicus and the corresponding segmented output for participants aged 4.5 and 6 years with different amounts of abdominal adipose tissue accumulation (low to high).

model with more spatial context information with the inclusion of axial, coronal, and sagittal views into the segmentation pipeline. In this work, we showed that it is likewise possible to train a model solely on axial image sections for scenarios when fast imaging (2D imaging sequence) is required in pediatric studies (Fig 5E, 5F). On the basis of our results, we propose that 2D segmentation can be successfully applied for volumetric fat assessment. Qualitative evaluation of the segmentation results showed that the proposed model achieved accurate fat assessment, correctly excluding nonfat structures, such as bone marrow, spleen, and fat content within the digestive tract. The model also successfully identified and segmented ambiguous regions (eg, separating VAT and SSAT in the umbilicus; Fig 5A, 5B).

As is the case for any segmentation algorithm, this model also comes with limitations. Visual investigations showed that in some cases, the very prominent appearance of blood vessels or other structures within the DSAT compartment could cause modest misclassified patches (Fig 5G). Additional performance reduction was identified when the image quality was drastically affected by intensity inhomogeneities of the B1 field or receiver-coil sensitivity (Fig 5H). Further fine-tuning of the model on reduced-quality images could help address this limitation. The use of Dixon fat-water images could potentially improve the segmentation performance, as it enables improved tissue contrast through more accurate fat-water separation. Three-dimensional models could potentially produce a more consistent volumetric segmentation and address

Figure 5: Strengths and limitations of the proposed approach for abdominal adipose tissue segmentation into visceral (yellow), deep subcutaneous (blue), and superficial subcutaneous (red) adipose tissues in **(A–D)** neonates (age ≤ 2 weeks) and **(E–H)** children (aged 4.5 and 6 years). **(A, B)** Ambiguous regions (overlapping of visceral and subcutaneous layers), **(C, D)** images with motion artifacts, **(E, F)** images in which the arms are located in close proximity to the abdominal wall, **(G)** misclassified patch within the deep subcutaneous adipose tissue compartment because of the very prominent appearance of blood vessels or other structures within the fat layer, and **(H)** image affected by intensity inhomogeneities of the B1 field or receiver coil sensitivity.

some of the limitations but are not applicable on images obtained by using 2D acquired image sequences. However, 2D models also come with benefits, such as reduced model complexity, reduced processing time, and reduced memory consumption.

Despite rapid changes in AAT partitioning between neonates and children, our method achieved accurate volumetric assessments of AAT depots over different ages from the neonatal period to early childhood. Our findings imply that the proposed method achieves results that are in strong agreement with expert-generated manual segmentation and can therefore greatly reduce the manual workload in large cohorts and longitudinal studies.

Considering the fat distributional differences during the neonatal period and the disparity of pathophysiologic findings for individual AAT depots in pediatric studies, this work provides opportunities to enhance the current understanding of AAT distribution and growth dynamics with respect to the development of cardiometabolic disease. Understanding fat distributions could help with the early identification of individuals at risk for adverse cardiometabolic health and with implementation and evaluation of personalized interventions.

Acknowledgments: The Growing Up in Singapore Towards healthy Outcomes study group includes Airu Chia, Allan Sheppard, Amutha Chinnadurai, Anna

Magdalena Fogel, Anne Eng Neo Goh, Anne Hin Yee Chu, Anne Rifkin-Graboi, Anqi Qiu, Arijit Biswas, Bee Wah Lee, Birgit Froukje Philipp Broekman, Bobby Kyungbeom Cheon, Boon Long Quah, Candida Vaz, Chai Kiat Chng, Cheryl Shufen Ng, Choon Looi Bong, Christiani Jayakumar Henry, Ciaran Gerard Forde, Claudia Chi, Daniel Yam Thiam Goh, Dawn Xin Ping Koh, Desiree Y. Phua, Doris Ngiuk Lan Loh, E. Shyong Tai, Elaine Kwang Hsia Tham, Elaine Phaik Ling Quah, Elizabeth Huiwen Tham, Evelyn Chung Ning Law, Evelyn Xiu Ling Loo, Fabian Kok Peng Yap, Faidon Magkos, Falk Mueller-Riemenschneider, George Seow Heong Yeo, Hannah Ee Juen Yong, Helen Yu Chen, Heng Hao Tan, Hong Pan, Hugo P. S. van Bever, Hui Min Tan, Iliana Magiati, Inez Bik Yun Wong, Ives Yubin Lim, Ivy Yee-Man Lau, Izzuddin Bin Mohd Aris, Jeannie Tay, Jeevesh Kapur, Jenny L. Richmond, Jerry Kok Yen Chan, Jia Xu, Joana Dawn Holbrook, Joanne Su-Yin Yoong, Joao Nuno Andrade Requicha Ferreira, Johan Gunnar Eriksson, Jonathan Tze Liang Choo, Jonathan Y. Bernard, Jonathan Yinhao Huang, Joshua J. Gooley, Jun Shi Lai, Karen Mei Ling Tan, Keith M. Godfrey, Kenneth Yung Chiang Kwek, Keri McCrickerd, Kok Hian Tan, Kothandaraman Narasimhan, Krishnamoorthy Naiduva, Kuan Jin Lee, Leher Singh, Li Chen, Lieng Hsi Ling, Lin Lin Su, Ling-Wei Chen, Lourdes Mary Daniel, Lynette Pei-Chi Shek, Marielle V. Fortier, Mark Hanson, Mary Foong-Fong Chong, Mary Rauff, Mei Chien Chua, Melvin Khee-Shing Leow, Michael J. Meaney, Michelle Zhi Ling Kee, Min Gong, Mya Thway Tint, Navin Michael, Neerja Karnani, Ngee Lek, Oon Hoe Teoh, P. C. Wong, Paulin Tay Straughan, Peter David Gluckman, Pratibha Keshav Agarwal, Priti Mishra, Queenie Ling Jun Li, Rob Martinus van Dam, Salome A. Rebello, Sambasivam Sendhil Velan, Seang Mei Saw, See Ling Loy, Seng Bin Ang, Shang Chee Chong, Sharon Ng, Shiao-Yng Chan, Shirong Cai, Shu-E Soh, Sok Bee Lim, Stella Tsotsi, Stephen Chin-Ying Hsu, Sue-Anne Ee Shioh Toh, Suresh Anand Sadananthan, Swee Chye Quek, Varsha Gupta, Victor Samuel Rajadurai, Walter Stunkel, Wayne Cutfield, Wee Meng Han, Wei Wei Pang, Wen Lun Yuan, Yanan Zhu, Yap Seng Chong, Yin Bun Cheung, Yiong Huak Chan, and Yung Seng Lee.

Author contributions: Guarantor of integrity of entire study, S.A.S.; study concepts/ study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, Y.M.K., K.T., K.M.G., U.C.M., S.S.V.; clinical studies, N.M., L.P.C.S., F.K.P.Y., Y.S.C., J.G.E., Y.S.L., S.S.V., M.F., S.A.S.; statistical analysis, Y.M.K., K.T., U.C.M., M.F., S.A.S.; and manuscript editing, Y.M.K., M.T.T., N.M., L.P.C.S., K.H.T., K.M.G., M.V.F., U.C.M., J.G.E., Y.S.L., S.S.V., M.F., S.A.S.

Disclosures of Conflicts of Interest: Y.M.K. disclosed no relevant relationships. K.T. disclosed no relevant relationships. M.T.T. disclosed no relevant relationships. N.M. disclosed no relevant relationships. L.P.C.S. is on the Takeda Pharmaceuticals advisory board for vaccines. F.K.P.Y. disclosed no relevant relationships. K.H.T. receives a consulting fee or honorarium from SingHealth. K.M.G. is part of an academic consortium that has received research funding from Abbott Nutrition, Nestec, Benevolent AI Bio, and Danone; has been reimbursed for speaking at conferences sponsored by companies selling nutritional products. Y.S.C. disclosed no relevant relationships. M.V.F. disclosed no relevant relationships. U.C.M. disclosed no relevant relationships. J.G.E. disclosed no relevant relationships. Y.S.L. disclosed no relevant relationships. S.S.V. received a consulting fee or honorarium from SBIC. M.F. disclosed no relevant relationships. S.A.S. disclosed no relevant relationships.

References

- Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med* 1997;337(13):869–873.
- Ayer J, Charakida M, Deanfield JE, Celermajer DS. Lifetime risk: childhood obesity and cardiovascular risk. *Eur Heart J* 2015;36(22):1371–1376.
- Bjerregaard LG, Jensen BW, Ångquist L, Osler M, Sørensen TIA, Baker JL. Change in overweight from childhood to early adulthood and risk of type 2 diabetes. *N Engl J Med* 2018;378(14):1302–1312.
- Woo JG. Infant growth and long-term cardiometabolic health: a review of recent findings. *Curr Nutr Rep* 2019;8(1):29–41.
- Smith SR, Lovejoy JC, Greenway F, et al. Contributions of total body fat, abdominal subcutaneous adipose tissue compartments, and visceral adipose tissue to the metabolic complications of obesity. *Metabolism* 2001;50(4):425–435.
- Lee MJ, Wu Y, Fried SK. Adipose tissue heterogeneity: implication of depot differences in adipose tissue for obesity complications. *Mol Aspects Med* 2013;34(1):1–11.
- Marinou K, Hodson L, Vasan SK, et al. Structural and functional properties of deep abdominal subcutaneous adipose tissue explain its association with insulin resistance and cardiovascular risk in men. *Diabetes Care* 2014;37(3):821–829.
- Kim SH, Chung JH, Song SW, Jung WS, Lee YA, Kim HN. Relationship between deep subcutaneous abdominal adipose tissue and metabolic syndrome: a case control study. *Diabetol Metab Syndr* 2016;8(1):10.
- Golan R, Shelef I, Rudich A, et al. Abdominal superficial subcutaneous fat: a putative distinct protective fat subdepot in type 2 diabetes. *Diabetes Care* 2012;35(3):640–647.
- González-Álvarez C, Ramos-Ibáñez N, Azpíroz-Leehan J, Ortiz-Hernández L. Intra-abdominal and subcutaneous abdominal fat as predictors of cardiometabolic risk in a sample of Mexican children. *Eur J Clin Nutr* 2017;71(9):1068–1073.
- Khoo CM, Leow MK, Sadananthan SA, et al. Body fat partitioning does not explain the interethnic variation in insulin sensitivity among Asian ethnicity: the Singapore adults metabolism study. *Diabetes* 2014;63(3):1093–1102.
- Sadanathan SA, Tint MT, Michael N, et al. Association between early life weight gain and abdominal fat partitioning at 4.5 years is sex, ethnicity, and age dependent. *Obesity (Silver Spring)* 2019;27(3):470–478.
- Tint MT, Fortier MV, Godfrey KM, et al. Abdominal adipose tissue compartments vary with ethnicity in Asian neonates: Growing Up in Singapore Toward Healthy Outcomes birth cohort study. *Am J Clin Nutr* 2016;103(5):1311–1317.
- Sun J, Xu B, Freeland-Graves J. Automated quantification of abdominal adiposity by magnetic resonance imaging. *Am J Hum Biol* 2016;28(6):757–766.
- Addeman BT, Kutty S, Perkins TG, et al. Validation of volumetric and single-slice MRI adipose analysis using a novel fully automated segmentation method. *J Magn Reson Imaging* 2015;41(1):233–241.
- Karlsson A, Rosander J, Romu T, et al. Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI. *J Magn Reson Imaging* 2015;41(6):1558–1569.
- Borga M. MRI adipose tissue and muscle composition analysis—a review of automation techniques. *Br J Radiol* 2018;91(1089):20180252.
- Hu HH, Chen J, Shen W. Segmentation and quantification of adipose tissue by magnetic resonance imaging. *MAGMA* 2016;29(2):259–276.
- Mosbech TH, Pilgaard K, Vaag A, Larsen R. Automatic segmentation of abdominal adipose tissue in MRI. In: Heyden A, Kahl F, eds. *Image analysis. SCIA 2011. Vol 6688, Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2011; 501–511.
- Sadanathan SA, Prakash B, Leow MK-S, et al. Automated segmentation of visceral and subcutaneous (deep and superficial) adipose tissues in normal and overweight men. *J Magn Reson Imaging* 2015;41(4):924–934.
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2015; 3431–3440.
- Wang Y, Qiu Y, Thai T, Moore K, Liu H, Zheng B. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. *Comput Methods Programs Biomed* 2017;144:97–104.
- Estrada S, Lu R, Conjeti S, et al. FatSegNet: a fully automated deep learning pipeline for adipose tissue segmentation on abdominal Dixon MRI. *Magn Reson Med* 2020;83(4):1471–1483.
- Langner T, Hedström A, Mörwald K, et al. Fully convolutional networks for automated segmentation of abdominal adipose tissue depots in multicenter water-fat MRI. *Magn Reson Med* 2019;81(4):2736–2745.
- Küstner T, Hepp T, Fischer M, et al. Fully automated and standardized segmentation of adipose tissue compartments via deep learning in 3D whole-body MRI of epidemiologic cohort studies. *Radiol Artif Intell* 2020;2(6):e200010.
- Soh SE, Tint MT, Gluckman PD, et al. Cohort profile: Growing Up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study. *Int J Epidemiol* 2014;43(5):1401–1409.
- World Health Organization. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age—methods and development. World Health Organization Web site. <http://www.who.int/childgrowth/standards/Technical%5Freport.pdf>. Published 2006. Accessed August 19, 2021.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Vol 9351, Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2015; 234–241.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv 1409.1556v6* [preprint] <https://arxiv.org/abs/1409.1556v6>. Posted September 4, 2014. Accessed August 19, 2021.
- Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proc Mach Learn Res* 2010;9:249–256.
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *ArXiv 1502.03167* [preprint] <https://arxiv.org/abs/1502.03167>.

- arxiv.org/abs/1502.03167. Posted February 11, 2015. Accessed August 19, 2021.
32. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. ArXiv 1606.04797v1 [preprint] <https://arxiv.org/abs/1606.04797v1>. Posted June 15, 2016. Accessed August 19, 2021.
 33. Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd International Conference on Learning Representations. ArXiv 1412.6980 [preprint] <https://arxiv.org/abs/1412.6980>. Posted December 22, 2014. Accessed August 19, 2021.
 34. Kullberg J, Hedström A, Brandberg J, et al. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. *Sci Rep* 2017;7(1):10425.
 35. Dixon WT. Simple proton spectroscopic imaging. *Radiology* 1984;153(1):189–194.