



OPEN

## Large-scale analysis of iliopsoas muscle volumes in the UK Biobank

Julie A. Fitzpatrick<sup>1,3</sup>, Nicolas Basty<sup>1,3</sup>✉, Madeleine Cule<sup>2</sup>, Yi Liu<sup>2</sup>, Jimmy D. Bell<sup>1</sup>, E. Louise Thomas<sup>1</sup> & Brandon Whitcher<sup>1</sup>

Psoas muscle measurements are frequently used as markers of sarcopenia and predictors of health. Manually measured cross-sectional areas are most commonly used, but there is a lack of consistency regarding the position of the measurement and manual annotations are not practical for large population studies. We have developed a fully automated method to measure iliopsoas muscle volume (comprised of the psoas and iliacus muscles) using a convolutional neural network. Magnetic resonance images were obtained from the UK Biobank for 5000 participants, balanced for age, gender and BMI. Ninety manual annotations were available for model training and validation. The model showed excellent performance against out-of-sample data (average dice score coefficient of  $0.9046 \pm 0.0058$  for six-fold cross-validation). Iliopsoas muscle volumes were successfully measured in all 5000 participants. Iliopsoas volume was greater in male compared with female subjects. There was a small but significant asymmetry between left and right iliopsoas muscle volumes. We also found that iliopsoas volume was significantly related to height, BMI and age, and that there was an acceleration in muscle volume decrease in men with age. Our method provides a robust technique for measuring iliopsoas muscle volume that can be applied to large cohorts.

The iliopsoas muscles, predominantly made up of slow-twitch fibers, are a composite of the psoas major and iliacus muscles; they are anatomically separate in the abdomen and pelvis but are merged together in the thigh. The iliopsoas is engaged during most day to day activities, including posture, walking and running. Together these muscles serve as the chief flexor of the hip and a dynamic stabiliser of the lumbar spine<sup>1</sup>, with the psoas uniquely having role in the movement of both the trunk and lower extremities<sup>2</sup>. Given the key involvement of the iliopsoas muscles in daily activities, there is increasing interest in its potential as a health biomarker. This has most commonly taken the form of a cross-sectional area (CSA) through one (generally the right) or both iliopsoas muscles, with the most common measurement taken through the psoas muscle. This CSA can be used either as an independent measurement or as a ratio to vertebral body size<sup>3,4</sup> or in the form of the psoas muscle index, calculated as the psoas muscle major CSA divided by the height squared<sup>5</sup>. Indeed, psoas CSA has been suggested as a predictor of sarcopenia<sup>6</sup>, surgical outcome and length of hospital stay post surgery<sup>7-9</sup>, poor prognosis in response to cancer treatment<sup>10</sup>, morbidity following trauma<sup>4</sup>, a surrogate marker of whole body lean muscle mass<sup>11</sup>, cardiovascular fitness<sup>12</sup>, changes in cardiometabolic risk variables following lifestyle intervention<sup>13</sup> and even risk of mortality<sup>14,15</sup>.

Measurements of the psoas major muscle are most commonly made from CSA of axial MRI or CT images<sup>7,12</sup>, with most studies generally relying on manual annotation of a single slice, through the abdomen, these tend to be retrospectively repurposed from clinical scans rather than a specific acquisition<sup>16-18</sup>. However, the CSA of the psoas muscle varies considerably along its length<sup>2</sup> therefore small differences in measurement position can potentially have a significant effect on its overall measured size. Moreover, there is a lack of consistency within the literature regarding the precise location at which measurement of the psoas CSA should be made, with researchers using a variety of approaches including: the level of the third lumbar vertebrae (L3)<sup>6,9,10,17,18</sup>, L4<sup>3,4,14,16</sup>, between L4-L5<sup>11,13</sup>, as well at level of the umbilicus<sup>7,8,19</sup> the precise position of which is known to vary with obesity/ascites. There is further discrepancy between studies regarding whether the measurements should comprise of one single<sup>10</sup> or both psoas muscles<sup>17</sup>, with the majority of publications combining the areas of both muscles.

This lack of consistency together with the relatively low attention given to robustness and reproducibility of its measurement, and the reliance on images from retrospective clinical scans have led many to question its validity as a biomarker<sup>20</sup>. A more objective proposition may be to measure total psoas muscle volume<sup>21-25</sup>, from dedicated images. A variety of approaches have been used thus far: inclusion of muscle between L2-L5<sup>21</sup>, psoas muscle volume from L3 and approximately the level of the iliopectineal arch (end point estimated from

<sup>1</sup>Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK. <sup>2</sup>Calico Life Sciences LLC, South San Francisco, CA, USA. <sup>3</sup>These authors contributed equally: Julie A. Fitzpatrick and Nicolas Basty. ✉email: n.basty@westminster.ac.uk

	Female	Male
Participants	2496 (49.9)	2504 (50.1)
<b>Ethnicity of total cohort</b>		
White European	2422 (48.44)	2416 (48.32)
Asian	18 (0.36)	35 (0.70)
Black	17 (0.34)	12 (0.24)
Other	15 (0.30)	9 (0.18)
Chinese	11 (0.22)	10 (0.20)
Not reported	7 (0.14)	12 (0.24)
Mixed	6 (0.12)	10 (0.20)
Age (years)	63.3 ± 8.3	63.3 ± 8.4
BMI (kg/m <sup>2</sup> )	26.2 ± 4.7	27.0 ± 3.9
Height (cm)	162.5 ± 6.1	176.2 ± 6.8
Weight (kg)	69.3 ± 13.3	83.9 ± 13.5

**Table 1.** Demographics of the subjects ( $n = 5000$ ). Reported values are counts with percentage (%) for categorical variables and average  $\pm$  standard deviation (SD) for continuous variables.

images in publications<sup>22,23</sup>, from the origin of the psoas at lumbar vertebrae (unspecified) to its insertion in the lesser trochanter<sup>24</sup>, or with no anatomical information provided at all<sup>25</sup>. Whilst all of these approaches include substantially more muscle than is included in simple CSA measurements, these are still incomplete volume measurements. Moreover, measuring the entire psoas muscle volume as a single entity is challenging, since even with 3D volumetric scans it is difficult to differentiate between composite iliacus and psoas muscles once they merge at the level of the inguinal ligament. Therefore, to measure psoas volume as an independent muscle it is necessary to either assign an arbitrary cutoff and not include a considerable proportion of the psoas muscle (estimated to be approximately 50% in some studies<sup>22</sup>) or simply include the iliacus muscle and measure the iliopsoas muscle volume in its entirety.

Convolutional Neural Networks (CNNs) have become a strong tool for automated image segmentation, especially architectures such as the U-Net<sup>26</sup> for two-dimensional (2D) data or the V-Net<sup>27</sup> for three-dimensional (3D) data. These techniques owe their popularity to the modest amount of training data required, robustness and fast execution speed. CNNs have been applied for automated muscle segmentation in computed tomography<sup>28–30</sup>, specifically for 2D segmentation of the psoas major muscle<sup>29</sup>, as well as MRI<sup>31,32</sup>.

The increasing use of whole body imaging<sup>33</sup> in large cohort studies such as the UK Biobank (UKBB), which plans to acquire MRI scans from the neck to the knee in 100,000 individuals<sup>34</sup>, requires different approaches to image analysis. Manual image segmentation is time consuming and infeasible in a cohort as large as the UKBB. However, this dataset provides a unique opportunity to measure iliopsoas muscles volume in a large cross-sectional population. Therefore, development of a robust and reliable automated method is essential. In this paper, we present an automated method to segment iliopsoas muscle volume using a CNN and discuss results arising from 5000 participants from the UKBB imaging cohort, balanced for BMI, age, and gender.

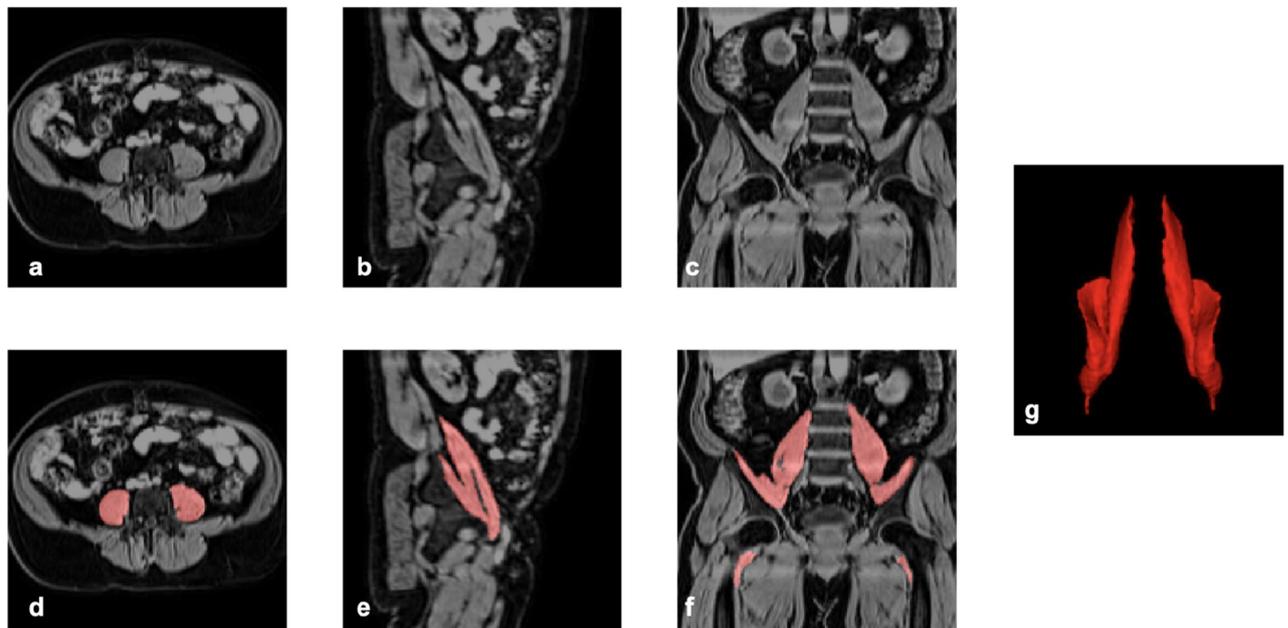
## Methods

**Data.** A total of 5000 subjects were randomly selected for this study, while controlling for BMI, age, and gender from the UKBB imaging cohort. Age was discretised into four groups: 44–53, 54–63, 63–72 and 73–82 years. The eight strata were defined to cover both age and gender. Weights were used to maintain the proportion of subjects within each age group to match that of the larger UKBB population.

Demographics for the study population (Table 1) were balanced for gender (female:male ratio of 49.9:50.1). The average age of the male subjects was  $63.3 \pm 8.4$  years and the female subjects was  $63.3 \pm 8.3$  years. The average BMI of the male subjects was  $27.0 \pm 3.9$  kg/m<sup>2</sup> (range 17.6–50.9 kg/m<sup>2</sup>) and for female subjects  $26.2 \pm 4.7$  kg/m<sup>2</sup> (range 16.1–55.2 kg/m<sup>2</sup>), with the mean for both groups being categorised as overweight. The self-reported ethnicity was predominantly White European (96.76%). As per the whole UKBB population, the sub-cohort in the current study was significantly healthier than the UK general population. The most common ailment were related to arthropathies, with smaller proportion reporting a variety of neoplasms, ranging from skin melanomas to benign neoplasms (Supplementary Table S1).

Participant data from the UKBB cohort was obtained as previously described<sup>34</sup> through UKBB Access Application number 23889. The UKBB has approval from the North West Multi-Centre Research Ethics Committee (REC reference: 11/NW/0382). All methods were performed in accordance with the relevant guidelines and regulations, and informed consent was obtained from all participants. Researchers may apply to use the UKBB data resource by submitting a health-related research proposal that is in the public interest. More information may be found on the UKBB researchers and resource catalogue pages (<https://www.ukbiobank.ac.uk/>).

Raw MR images were obtained from the UKBB Abdominal Protocol<sup>35</sup> and preprocessed as previously reported<sup>36,37</sup>. The data were acquired on the same model, a Siemens Aera 1.5 T scanner (Syngo MR D13) (Siemens, Erlangen, Germany), across three sites (Stockport, Newcastle, Reading, UK). The Dixon sequence involved six overlapping series that were acquired using a common set of parameters: TR = 6.67 ms, TE = 2.39/4.77 ms, in-plane voxel size  $2.232 \times 2.232$  mm, FA = 10° and bandwidth = 440 Hz. The first series, over the neck, consisted



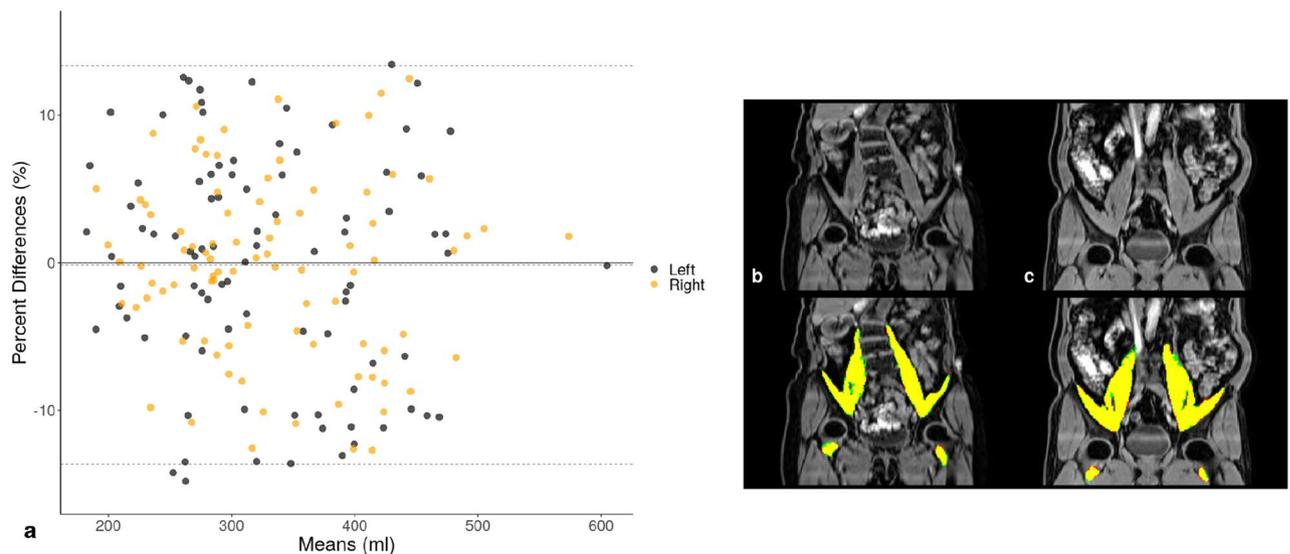
**Figure 1.** Iliopsoas muscle manual annotations: (a) axial, (b) sagittal, and (c) coronal views, (d–f) showing of the segmentation (red) overlaid on the anatomical reference data, and (g) 3D rendering of manual segmentation.

of 64 slices, slice thickness 3.0 mm and  $224 \times 168$  matrix; series two to four (covering the chest, abdomen and pelvis) were acquired during 17 second expiration breath holds with 44 slices, slice thickness 4.5 mm and  $224 \times 174$  matrix; series five, covering the upper thighs, consisted of 72 slices, slice thickness 3.5 mm and  $224 \times 162$  matrix; series six, covering the lower thighs and knees, consisted of 64 slices, slice thickness 4 mm and  $224 \times 156$  matrix. During preprocessing the data were resampled to voxel size  $2.232 \times 2.232 \times 3.0$  mm.

**Manual annotation.** A single expert radiographer manually annotated both iliopsoas muscles for 90 subjects using the open-source software MITK<sup>38</sup>. Each axial slice of the water images was examined, the iliopsoas identified, and the borders of the psoas and iliopsoas manually drawn for 90 subjects. On average, manual annotation of both muscles took five to seven hours per subject. The annotated data covered a broad range of age and BMI from male and female UKBB participants. A typical Dixon abdominal dataset, centred on the iliopsoas muscles, is shown in Fig. 1, manual iliopsoas muscle annotations are overlaid on the anatomical reference volume in red and a 3D rendering of the manual annotation.

**Model.** We trained a model able to predict both muscles individually. The preprocessing steps for the training data, where the cropping is also needed for applying the model to unseen data, are as follows. Two arrays of size  $96 \times 96 \times 192$  were cropped around the hip landmarks<sup>36</sup>, to approximate the location of the muscles in order to perform the segmentation (an example of the cropped regions may be found in Supplementary Fig. S1). After cropping, each volume was normalised such that the signal intensities lie between zero and one, where the 99<sup>th</sup> percentile was used instead of the maximum to avoid possible spikes in signal intensity. That is, all signal intensities above the 99<sup>th</sup> percentile were mapped to one. Two sets of 16 training samples were generated for every subject by separating the right and the left muscles, introducing reflections exploiting the symmetry of the structures. Further data augmentation included seven random transformations consisting of translations by up to six voxels in-plane, up to 24 voxels out-of-plane, and random scaling ranging from  $-50$  to  $+50\%$  out-of-plane and from  $-25$  to  $+25\%$  in-plane, in addition to the original data. We chose larger factors for out-of-plane transformations to account for the skewed variability in shape and position of the muscles, reflecting the fact that there is more variation in height than width in the population. After data augmentation, 2880 training samples were produced from the original 90 manually annotated pairs of iliopsoas muscles.

The model used for 3D iliopsoas muscle segmentation closely follows a similar architecture to the U-Net<sup>26</sup> and the V-Net<sup>27</sup>, with a contracting path and an expansive path connected by skip connections at each resolution level. These network architectures have been established as the gold standard for image segmentation over the last few years, as they require modest amounts of training data as a consequence of operating at multiple resolution levels while providing excellent results within seconds. Several convolution blocks are used in our model architecture. An initial block ( $I$ ) contains a  $5 \times 5 \times 5$  convolution with eight filters followed by a  $2 \times 2 \times 2$  convolution with 16 filters and stride two. The down-sampling blocks in the contraction ( $D_{i,m}$ ) consist of  $i$  successive  $5 \times 5 \times 5$  convolutions with  $m$  filters followed by a  $2 \times 2 \times 2$  convolution of stride with stride two, used to decrease the resolution. In the expansion, the up-sampling blocks ( $U_{j,n}$ ) mirror the ones in the contraction where there are transpose convolutions instead of stride two convolutions. The block ( $L$ ) at the lowest resolution level of the architecture consist of three successive  $5 \times 5 \times 5$  convolutions with 128 filters followed by a  $2 \times 2 \times 2$



**Figure 2.** Bland–Altman plot (a) of iliopsoas muscle volumes determined with CNN-based and manual segmentations ( $n = 90$ ), using a six-fold cross-validation experiment. Dotted lines represent the average bias ( $-0.2\%$ ) and the 95% limits of agreement. Overlays of the CNN-based and manual segmentations for two subjects (b,c), where the manual annotation is red, the CNN segmentation is green and the overlap is yellow.

transpose convolution of stride two and 64 filters. The final block ( $F$ ) contains a  $5 \times 5 \times 5$  convolution with 16 filters followed by a single  $1 \times 1 \times 1$  convolution and a final sigmoid activation classification layer. All blocks incorporate skip connections between their input and output, resulting in residual layers. The architecture follows:  $I \rightarrow D_{2,32} \rightarrow D_{3,64} \rightarrow D_{3,128} \rightarrow L \rightarrow U_{3,128} \rightarrow U_{3,64} \rightarrow U_{3,32} \rightarrow F$  with skip connections between blocks at equivalent resolution levels. Padding is used for the convolutions throughout the network and a stride of one, unless otherwise specified, when moving between the resolution levels. Other than the final sigmoid activation, scaled exponential linear units (SELU) are used throughout the network. The SELU activation function has recently been proposed<sup>39</sup>, where the self-normalising properties allow it to bypass batch normalisation layers enabling higher learning rates that lead to more robust and faster training.

The model was trained minimising the Dice score coefficient (DSC) loss<sup>27</sup> with a batch size of three using the Adam optimiser and a learning rate of  $1e-4$  until convergence at 100 epochs. The learning rate was determined through a parameter sweep ( $1e-1$  to  $1e-6$ ). We performed all of the CNN development, learning, and predictions using Keras (TensorFlow backend)<sup>40</sup> on a NVIDIA Titan V 12 GB GPU. We limited the batch size to three due to the GPU memory.

**Validation.** A common metric used to evaluate segmentation performance is the DSC, also known as the F1 score. It is defined as twice the intersection of the labels divided by the total number of elements. Intersection of labels can also be seen as a True Positive (TP) outcome. The total number of elements can also be seen as the sum of all False Positives (FP), False Negatives (FN) and twice the number of TPs.

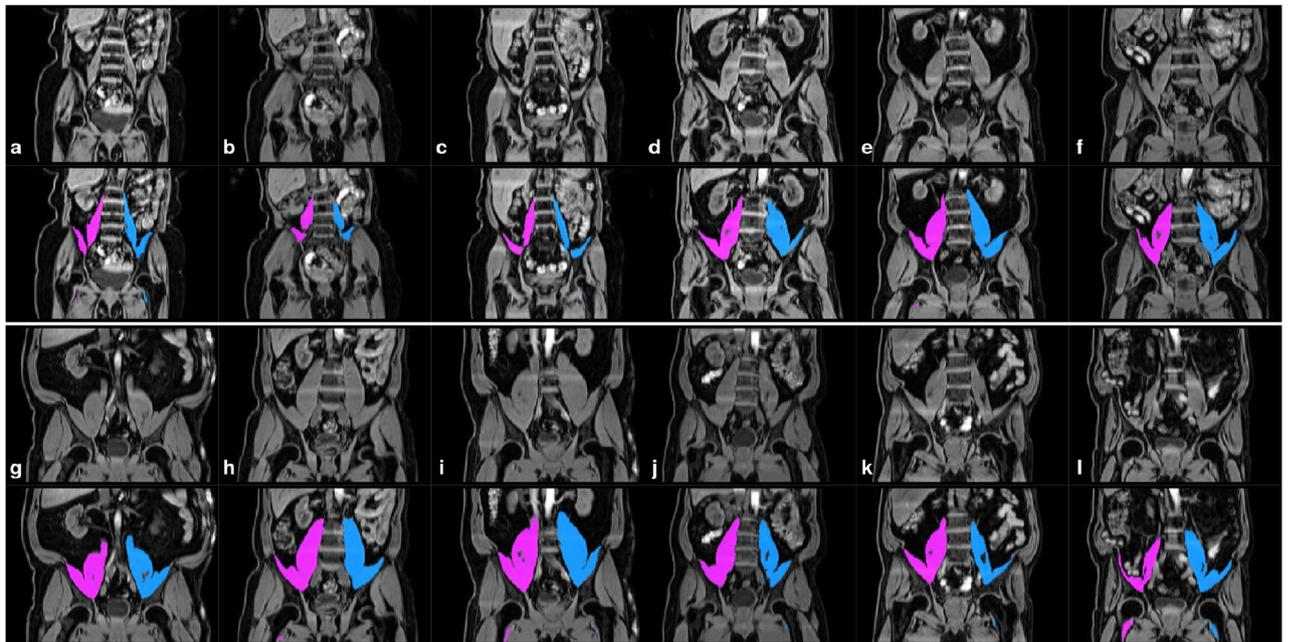
$$\text{DSC} = \frac{2 \text{ TP}}{\text{FP} + 2 \text{ TP} + \text{FN}} \quad (1)$$

For validation of the model, we performed a six-fold cross-validation experiment, where in a single iteration 75 of the manually annotated images (approximately 83%) were used to train the model and the performance was evaluated on the remaining 15 out-of-sample images (approximately 17%).

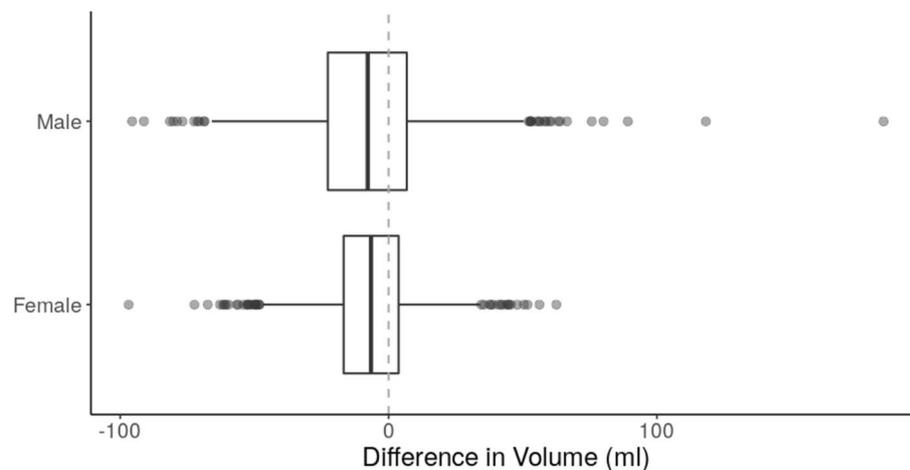
**Statistical analysis.** All summary statistics, hypothesis tests and figures have been performed using the R software environment for statistical computing and graphics<sup>41</sup>. Variables were tested for normality using the Shapiro–Wilk’s test, the null hypothesis was rejected in all cases. Spearman’s rank correlation coefficient ( $\rho$ ) was used to assess monotonic trends between variables. The Wilcoxon rank-sum test was used to compare means between groups, and the Wilcoxon signed-rank test with paired observations. Methods for segmenting the iliopsoas muscle volume were compared using the Bland–Altman plot. Given the exploratory nature of the research,  $p$ -values  $< 0.05$  were judged to be statistically significant.

## Results

**Validation.** A summary of the cross-validation experiment may be found in Supplementary Table S2. The average bias was  $-0.2\%$  with upper and lower limits of agreement being  $13.3\%$  and  $-13.7\%$ , respectively (Fig. 2). The overlap between the CNN-based and manual segmentations for two subjects is also provided in Fig. 2, where the DSCs are 0.85 (left) and 0.90 (right) for (b) and 0.96 in both for (c). With consistent DSCs from the cross-



**Figure 3.** CNN segmentations of the left and right iliopsoas muscles overlaid in purple (right) and blue (left) from a range of body types and iliopsoas muscle volumes: (a–c) small, (d–f) average, (g–i) large and (j–l) asymmetric. The top row for each subject displays the signal intensities without the segmentation result, the bottom row includes the iliopsoas muscle segmentations.



**Figure 4.** Difference in volume (ml) between the left and right iliopsoas muscles, separated by gender. Negative values indicate the right iliopsoas muscle is larger than the left.

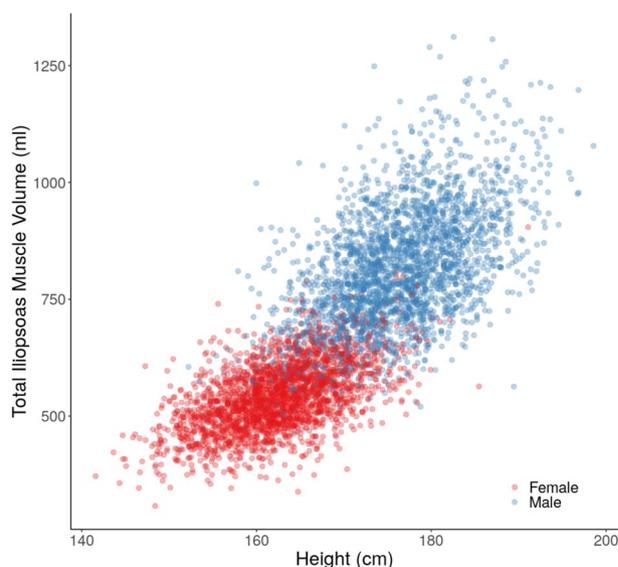
validation experiment showing a robust model performance on both muscles, we trained a final model using the entire 90 available manual annotations.

Example segmentations from our method are provided in Fig. 3, displaying a sample of 12 subjects covering a variety of body sizes and habitus. The first three subjects (a–c) have some of the smallest iliopsoas muscles (total volume  $\approx 346$  ml), the next three subjects (d–f) have typical iliopsoas muscles (total volume  $\approx 800$  ml) and the third set of three subjects (g–i) have some of the largest iliopsoas muscles (total volume  $\approx 1300$  ml). The final set of three subjects represent subjects whose left and right iliopsoas muscles differ in volume (difference in volume  $\approx 93$  ml for j and k, difference in volume = 182 ml for l). We can see that the model performs well for all of them, with additional details regarding model validation provided in Supplementary Fig. S2.

**Iliopsoas muscle volume.** In each gender there was a small (approximately 2%) yet statistically significant asymmetry between left and right iliopsoas muscles (Wilcoxon signed-rank test; male:  $d = -7.3$  ml; female:  $d = -6.5$  ml; both  $p < 10^{-15}$ ) (Fig. 4). These differences were not significantly associated with the handedness

	Female		Male		Significance
	Mean $\pm$ SD	Range	Mean $\pm$ SD	Range	
Total volume (ml)	542.3 $\pm$ 72.1	307.5, 904.2	814.5 $\pm$ 125.4	467.3, 1311.5	$p < 10^{-15}$
Average volume (ml)	271.2 $\pm$ 36.0	153.8, 452.1	407.2 $\pm$ 62.7	233.7, 655.8	$p < 10^{-15}$
Left volume (ml)	267.9 $\pm$ 36.8	134.5, 457.2	403.6 $\pm$ 63.5	247.0, 639.2	$p < 10^{-15}$
Right volume (ml)	274.4 $\pm$ 37.0	159.4, 447.0	410.9 $\pm$ 64.0	220.3, 675.1	$p < 10^{-15}$
L-R volume difference (ml)	-6.5 $\pm$ 16.1	-96.9, 62.5	-7.3 $\pm$ 22.8	-95.6, 184.4	$p = 0.040$
Iliopsoas muscle index (ml/cm <sup>2</sup> )	205.1 $\pm$ 22.6	124.2, 304.1	261.8 $\pm$ 34.2	157.6, 417.2	$p < 10^{-15}$

**Table 2.** Iliopsoas muscle volumes ( $n = 5000$ ). Significance refers to the  $p$ -value for a Wilcoxon rank-sums test, where the null hypothesis is the medians between the two groups (male and female subjects) being equal.



**Figure 5.** Scatterplot of total iliopsoas muscle volume (ml) by height (cm), separated by gender.

of the participants. Significantly larger iliopsoas muscle volumes were measured in male compared with female subjects (Table 2).

**Relationship between iliopsoas muscle volume and physical characteristics.** Significant correlations were observed between the total iliopsoas muscle volume and height in both genders (male:  $\rho = 0.51$ ; female:  $\rho = 0.54$ , both  $p < 10^{-15}$ ) (Fig. 5).

To account for the potential confounding effect of height on iliopsoas muscle volume, an iliopsoas muscle index (IMI) was defined

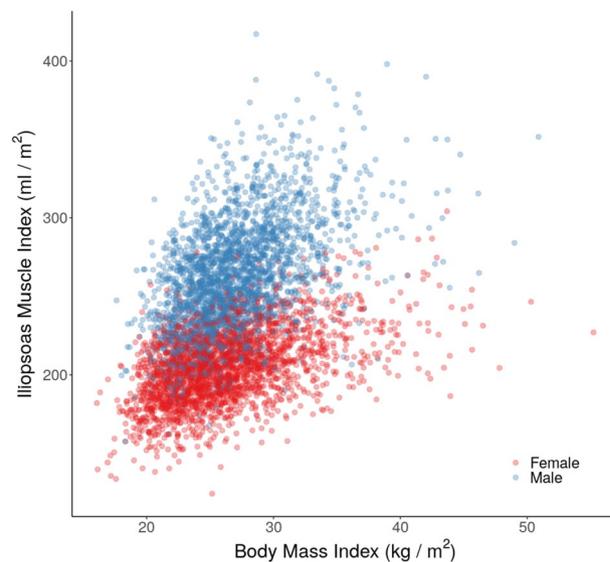
$$\text{IMI} = \frac{\text{total iliopsoas muscle volume}}{\text{height}^2}, \quad (2)$$

with units ml/m<sup>2</sup>. Significant correlations were observed between the IMI and BMI in both genders (male:  $\rho = 0.48$ ; female:  $\rho = 0.47$ , both  $p < 10^{-15}$ ) (Fig. 6).

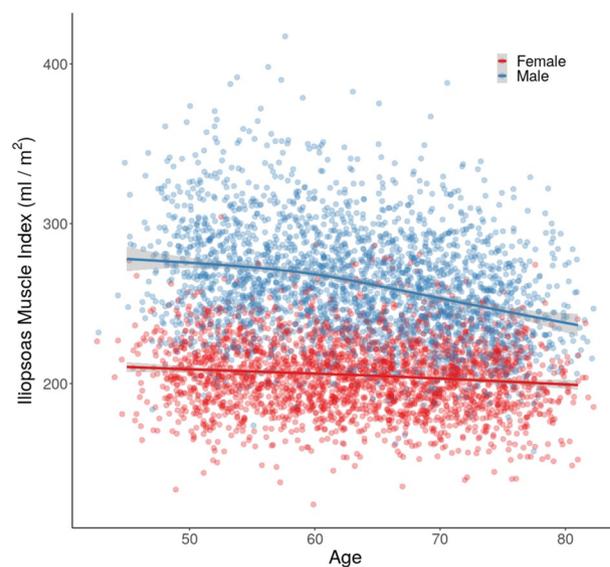
A significant negative correlation was observed between IMI and age in both genders (male:  $\rho = -0.31$ ,  $p < 10^{-15}$ ; female:  $\rho = -0.11$ ,  $p < 10^{-7}$ ). However, the relationship could not be easily explained by a simple linear method (Fig. 7). In fact the decrease in IMI as a function of age accelerates for men, starting in their early 60s, while for women it remains relatively constant.

## Discussion

There is considerable interest in measuring psoas muscle size, primarily related to its potential as a sarcopenic marker, thereby making it an indirect predictor of conditions influenced by sarcopenia and frailty, including health outcomes such as morbidity, and mortality<sup>4,6-10,14,15</sup>. The complexity in measuring total muscle directly, particularly in a frail population has necessitated the reliance on easily measured surrogates and the psoas muscle CSA is increasingly used for this purpose. However there is little consistency in the field regarding how the psoas muscle is measured, with considerable variation between publications. An automated approach to analysis will



**Figure 6.** Scatterplot of iliopsoas muscle index ( $\text{ml}/\text{m}^2$ ) by BMI ( $\text{kg}/\text{m}^2$ ), separated by gender.



**Figure 7.** Scatterplot of iliopsoas muscle index ( $\text{ml}/\text{m}^2$ ) by age at recruitment (years), separated by gender. The curves are fit to the data using a generalised additive model with cubic splines.

reduce the need for manual annotation, allowing more of the muscle to be measured and enable much larger cohorts to be studied, this is particularly important as large population based biobanks are becoming more common. In this paper we have described a CNN-based method to automatically extract and quantify iliopsoas muscle volume from MRI scans for 5000 participants from the UKBB. Excellent agreement was obtained between automated measurements and the manual annotation undertaken by a trained radiographer as demonstrated by the extremely high DSC with testing data.

CNNs have been established as the gold standard in automated image segmentation. The results, which can be produced with a modest amount of manual annotations as training data and smart data augmentation, are highly accurate, fast, and reproducible. Manual annotations become a bottleneck for large-scale population studies, when the number of participants exceeds many thousand such as with the UKBB. Applying automated methods to vast amounts of data requires a thorough set of quality-control procedures beyond just out-of-sample testing data, which is often used to validate new methods in machine learning studies. Large-scale quality control can be done by steps such as looking at maximum and minimum values, asymmetric values (for symmetric structures such as the iliopsoas muscles), outliers, and overall behaviour of the results.

The vast majority of previous studies investigating psoas muscle size have relied on CSA measurements primarily because of data availability and time constraints<sup>3,4,6–11,13,14,16–19</sup>. Analysis of CSA is considerably less labour intensive than manually measuring tissue volumes, furthermore, many studies have repurposed clinical CT or MRI scans<sup>16–18</sup> which typically will not have been acquired in a manner to enable volume measurements. This has led to psoas muscle CSA being measured at a variety of positions relating to lumbar landmarks including L3, L4 and between L4–5, as well as more unreliable soft tissue landmarks such as the umbilicus, with the CSA measurements used alone, relative to lumbar area, height, height squared or total abdominal muscle within the image at the selected level. While lumbar landmarks should provide a relatively consistent CSA in longitudinal studies, comparison between studies and cohorts becomes almost impossible. This is further compounded by studies that have shown considerable variation in psoas CSA along its length<sup>2,42</sup>, and that regional differences in psoas CSA have been observed in athletes<sup>43</sup>, following exercise training or inactivity<sup>44</sup>. This appears to suggest that CSA at a fixed position may not accurately reflect changes in the psoas size elsewhere in response to health related processes. It is clear that to overcome these confounding factors, it is essential to measure total psoas volume.

In this study, we have trained a CNN to segment iliopsoas muscles, applied it to 5000 UKBB subjects and measured their total volume. This measurement includes the psoas major and iliacus muscles, and as mentioned in the proceeding section, the psoas minor muscle (if present). This reflects the practical difficulties of isolating the entire psoas muscle in images in a consistent and robust manner. The merging of the iliacus and psoas muscles below the inguinal ligament makes their separation not only impractical, but unachievable with standard imaging protocols. Similarly, it is not possible to separate the psoas major and minor muscles under these conditions, even if CSA measurements were to be made. Therefore, a standard operating procedure was required, either measure a partial psoas volume, selecting an anatomical cut-off before the junction with the iliacus muscle, or to include the iliacus and measure the iliopsoas muscle volume in its entirety. In this study we have opted for the latter, as selecting an arbitrary set point would clearly introduce a significant confounding factor with unforeseeable impact on the subsequent results. Thus, we have measured the entire iliopsoas muscle, and although literature comparisons are limited, as there is a paucity of comparable volumetric studies within the general population, our average reported values for male subjects ( $407.2 \pm 62.7$  ml) were within the range 351.1–579.5 ml in a cohort which included male athletes and controls<sup>43</sup>.

Furthermore, our CNN-based method performs very well, with a small but systematic underestimation of  $-0.2\%$  when compared with manual annotations. Incremental improvement of the model is possible using straightforward techniques, such as increasing the number and variety of training data or expanding the breadth of data augmentation<sup>45</sup>. These are currently under investigation.

We observed a small (approximately 2%) but significant asymmetry in iliopsoas muscle volume, with the right muscle being larger in both male and female subjects. Previous studies have looked at the muscle asymmetry in tennis players, and found that the iliopsoas muscle was 13% smaller on the non-dominant compared with the dominant side of the body, whereas inactive controls the dominant size was 4% larger than the non-dominant<sup>43</sup>. Similarly footballers players have significantly larger psoas CSA on their dominant kicking side<sup>46</sup>. The best equivalent to this within the UKBB phenotyping data was handedness, which we found not to be related to left-right differences in iliopsoas volume in the current study. An additional factor which may contribute towards iliopsoas asymmetry relates to the presence or absence of the psoas minor muscle, a long slim muscle typically found in front of the psoas major. This muscle can often fail to develop during embryonic growth<sup>2</sup> and there can be considerable differences in the incidence of agenesis which can be unilateral or bilateral with ethnicity thought to be a factor<sup>47</sup>. Further work is required to understand whether this contributes to the left-right asymmetry observed in the present study, since it is not possible to resolve this muscle on standard MRI images.

In line with previous studies of psoas CSA, male subjects had significantly larger iliopsoas muscles compared to females<sup>6</sup>. This is unsurprising since gender differences in both total muscle and regional muscle volumes are well established<sup>48,49</sup>. Indeed some studies have suggested using gender specific cut-offs of either psoas CSA alone or psoas muscle index to identify patients at risk of poorer health outcomes<sup>10</sup>. Furthermore, some studies have suggested that the magnitude of gender differences in trunk muscle CSA vary depending where are measured. This adds weight to the argument that volumetric measurements are perhaps more robust than CSA measures for this comparison<sup>50</sup>. It has been proposed that the gender differences in psoas volume could in part relate to the impact of height on psoas volume<sup>12</sup>. Indeed, we found a significant correlation between iliopsoas muscle volume and height similar to those previously reported by earlier studies<sup>49</sup>. However, the gender differences observed in our study were still present when correcting for height. Interestingly, it has been reported that the relationship between muscle volume and body weight is curvilinear, since increases in body weight often reflect gain in fat, as well as muscle mass. In the present study we observe a significant correlation between IMI and BMI. This is in agreement with previous studies of psoas CSA which have also shown a significant correlation with BMI<sup>6</sup>, indeed some studies combined both metrics as a prognostic marker<sup>17</sup>. We also found a significant correlation between IMI and age. It is widely reported that muscle mass declines with age, particularly beyond the fifth decade, a fundamental characteristic of sarcopenia<sup>51</sup>. The magnitude of this decline was relatively small, but this may arise by the limited age range within the UKBB data set (44–82 years), compared to other studies that have investigated the impact of age on muscle volume across the entire adult age span (18–88 years), which usually tend to reveal a more dramatic decline in muscle volume<sup>49</sup>.

In conclusion, we have developed a robust and reliable model using a CNN to automatically segment iliopsoas muscles and demonstrated the applicability of this methodology in a large cohort, which will enable future population-wide studies of the utility of iliopsoas muscle as a predictor of health outcomes.

### Code availability

Model weights and instructions for use are available at [https://github.com/recoh/iliopsoas\\_muscle](https://github.com/recoh/iliopsoas_muscle).

Received: 11 August 2020; Accepted: 5 November 2020

Published online: 19 November 2020

## References

- Regev, G. J. *et al.* Psoas muscle architectural design, in vivo sarcomere length range, and passive tensile properties support its role as a lumbar spine stabilizer. *Spine* **36**, E1666–E1674 (2011).
- Hanson, P., Magnusson, S. P., Sorensen, H. & Simonsen, E. B. Anatomical differences in the psoas muscles in young black and white men. *J. Anat.* **194**, 303–307 (1999).
- Swanson, S. & Patterson, R. B. The correlation between the psoas muscle/vertebral body ratio and the severity of peripheral artery disease. *Ann. Vasc. Surg.* **29**, 520–525 (2015).
- Ebbeling, L. *et al.* Psoas: lumbar vertebra index: central sarcopenia independently predicts morbidity in elderly trauma patients. *Eur. J. Trauma Emerg. Surg.* **40**, 57–65 (2014).
- Mourtzakis, M. *et al.* A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl. Physiol. Nutr. Metab.* **33**, 997–1006 (2008).
- Jones, K., Doleman, B., Scott, S., Lund, J. N. & Williams, J. P. Simple psoas cross-sectional area measurement is a quick and easy method to assess sarcopenia and predicts major surgical complications. *Colorectal Dis.* **17**, O20–O26 (2015).
- Durand, F. *et al.* Prognostic value of muscle atrophy in cirrhosis using psoas muscle thickness on computed tomography. *J. Hepatol.* **60**, 1151–1157 (2014).
- Saitoh-Maeda, Y. *et al.* A low psoas muscle volume correlates with a longer hospitalization after radical cystectomy. *BMC Urol.* **17**, 87 (2017).
- Delitto, D. *et al.* A clinically applicable muscular index predicts long-term survival in resectable pancreatic cancer. *Surgery* **161**, 930–938 (2017).
- Kasahara, R. *et al.* A low psoas muscle index before treatment can predict a poorer prognosis in advanced bladder cancer patients who receive gemcitabine and nedaplatin therapy. *BioMed Res. Int.* **2017**, (2017).
- Morrell, G. R. *et al.* Psoas muscle cross-sectional area as a measure of whole-body lean muscle mass in maintenance hemodialysis patients. *J. Renal Nutr.* **26**, 258–264 (2016).
- Fitzpatrick, J. *et al.* Psoas major cross-sectional area: a potential marker of cardiorespiratory fitness. *Int. J. Clin. Exp. Physiol.* **4**, 15–20 (2017).
- Maltais, A. *et al.* One-year lifestyle intervention, muscle lipids, and cardiometabolic risk. *Med. Sci. Sports Exerc.* **51**, 2156–2165 (2019).
- Drudi, L. *et al.* Psoas muscle area predicts all-cause mortality after endovascular and open aortic aneurysm repair. *Eur. J. Vasc. Endovasc. Surg.* **52**, 764–769 (2016).
- Huber, T. C. *et al.* Predictors of all-cause mortality after endovascular aneurysm repair: Assessing the role of psoas muscle cross-sectional area. *J. Vasc. Interv. Radiol.* **30**, 1972–1979 (2019).
- Lee, J.S.-J. *et al.* Frailty, core muscle size, and mortality in patients undergoing open abdominal aortic aneurysm repair. *J. Vasc. Surg.* **53**, 912–917 (2011).
- Hervocho, R. *et al.* Body mass index and total psoas area affect outcomes in patients undergoing pneumonectomy for cancer. *Ann. Thorac. Surg.* **103**, 287–295 (2017).
- Bukvić, M. *et al.* Psoas muscle index as indicator of sarcopenia in patients with colorectal carcinoma during oncological treatment. *Int. J. Sci. Eng. Res.* **10**, 599–602. <https://doi.org/10.14299/ijser.2019.07.01> (2019).
- Gu, D. H. *et al.* Clinical usefulness of psoas muscle thickness for the diagnosis of sarcopenia in patients with liver cirrhosis. *Clin. Mol. Hepatol.* **24**, 319 (2018).
- Baracos, V. E. Psoas as a sentinel muscle for sarcopenia: a flawed premise. *J. Cachexia Sarcopenia Musc.* **8**, 527–528 (2017).
- Modesto, A. E. *et al.* Psoas muscle size as a magnetic resonance imaging biomarker of progression of pancreatitis. *Eur. Radiol.* **1–10**, (2020).
- Valero, V. *et al.* Sarcopenia adversely impacts postoperative complications following resection or transplantation in patients with primary liver tumors. *J. Gastrointestinal. Surg.* **19**, 272–281 (2015).
- Amini, N. *et al.* Impact total psoas volume on short- and long-term outcomes in patients undergoing curative resection for pancreatic adenocarcinoma: a new tool to assess sarcopenia. *J. Gastrointestinal Surg.* **19**, 1593–1602 (2015).
- Zargar, H. *et al.* Change in psoas muscle volume as a predictor of outcomes in patients treated with chemotherapy and radical cystectomy for muscle-invasive bladder cancer. *Bladder Cancer* **3**, 57–63 (2017).
- Suh, J. W. *et al.* Effect of sarcopenic overweight on lung transplant based in three-dimensional reconstructed psoas muscle mass. *Ann. Thorac. Surg.* **107**, 1626–1631 (2019).
- Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, 234–241 (Springer, 2015).
- Millitari, F., Navab, N. & Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571 (IEEE, 2016).
- Hiasa, Y. *et al.* Automated muscle segmentation from clinical ct using bayesian u-net for personalized musculoskeletal modeling. *IEEE Trans. Med. Imaging* **39**, 1030–1040 (2019).
- Hashimoto, F., Kakimoto, A., Ota, N., Ito, S. & Nishizawa, S. Automated segmentation of 2d low-dose ct images of the psoas-major muscle using deep convolutional neural networks. *Radiol. Phys. Technol.* **12**, 210–215 (2019).
- Kamiya, N. *et al.* Surface muscle segmentation using 3d u-net based on selective voxel patch generation in whole-body ct images. *Appl. Sci.* **10**, 4477 (2020).
- Ghosh, S., Boulanger, P., Acton, S. T., Blemker, S. S. & Ray, N. Automated 3d muscle segmentation from mri data using convolutional neural network. In *2017 IEEE International Conference on Image Processing (ICIP)*, 4437–4441 (IEEE, 2017).
- Weber, K. A. *et al.* Deep learning convolutional neural networks for the automatic quantification of muscle fat infiltration following whiplash injury. *Sci. Rep.* **9**, 1–8 (2019).
- Borga, M. *et al.* Validation of a fast method for quantification of intra-abdominal and subcutaneous adipose tissue for large-scale human studies. *NMR Biomed.* **28**, 1747–1753 (2015).
- Sudlow, C. *et al.* UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nat. Commun.* <https://doi.org/10.1038/s41467-020-15948-9> (2020).
- Basty, N. *et al.* Image processing and quality control for abdominal magnetic resonance imaging in the UK Biobank (2020). [arXiv:2007.01251](https://arxiv.org/abs/2007.01251).
- Liu, Y. *et al.* Systematic quantification of health parameters from UK Biobank abdominal MRI using deep learning. [bioRxiv:https://doi.org/10.1101/2020.07.14.187070](https://doi.org/10.1101/2020.07.14.187070) (2020).
- Nolden, M. *et al.* The Medical Imaging Interaction Toolkit: challenges and advances. *Int. J. Comput. Assist. Radiol. Surg.* **8**. <https://doi.org/10.1007/s11548-013-0840-8> (2013).

39. Klambauer, G., Unterthiner, T., Mayr, A. & Hochreiter, S. Self-normalizing neural networks. *Adv. Neural Inf. Process. Syst.* 971–980 (2017).
40. Chollet, F. *et al.* Keras. <https://github.com/fchollet/keras> (2015).
41. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2020).
42. Reid, J. G., Livingston, L. A. & Pearsall, D. J. The geometry of the psoas muscle as determined by magnetic resonance imaging. *Arch. Phys. Med. Rehabil.* 75, 703–708 (1994).
43. Sanchis-Moysi, J., Idoate, F., Izquierdo, M., Calbet, J. A. L. & Dorado, C. Iliopsoas and gluteal muscles are asymmetric in tennis players but not in soccer players. *PLoS ONE* 6, e22858 (2011).
44. Hides, J. A. *et al.* Magnetic resonance imaging assessment of trunk muscles during prolonged bed rest. *Spine* 32, 1687–1692 (2007).
45. Lundervold, A. S. & Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* 29, 102–127. <https://doi.org/10.1016/j.zemedi.2018.11.002> (2019).
46. Stewart, S., Stanton, W., Wilson, S. & Hides, J. Consistency in size and asymmetry of the psoas major muscle among elite footballers. *Br. J. Sports Med.* 44, 1173–1177 (2010).
47. Ojha, P., Prakash, S. & Jain, A. Morphology of psoas minor muscle—a cadaveric study. *Int. J. Curr. Res. Rev.* 8, 35 (2016).
48. Gallagher, D. & Heymsfield, S. B. Muscle distribution: variations with body weight, gender, and age. *Appl. Radiat. Isot.* 49, 733–734 (1998).
49. Janssen, I., Heymsfield, S. B., Wang, Z. M. & Ross, R. Skeletal muscle mass and distribution in 468 men and women aged 18–88 years. *J. Appl. Physiol.* 89, 81–88 (2000).
50. Abe, T., Kearns, C. F. & Fukunaga, T. Sex differences in whole body skeletal muscle mass measured by magnetic resonance imaging and its distribution in young Japanese adults. *Br. J. Sports Med.* 37, 436–440 (2003).
51. Rosenberg, I. H. Sarcopenia: origins and clinical relevance. *J. Nutr.* 127, 990S–991S (1997).

### Author contributions

J.A.F., J.D.B. and E.L.T. conceived the study. J.D.B., E.L.T., N.B. and B.W. designed the study. J.A.F. performed the manual annotations. N.B. implemented the methods and performed data analysis. B.W. performed statistical analysis of the data. E.L.T. and N.B. drafted the manuscript. J.D.B., E.L.T., N.B., Y.L., M.C. and B.W. participated in data preprocessing and editing the manuscript. All authors read and approved the manuscript.

### Competing interests

MC and YL are employees of Calico Life Sciences LLC. NB, JAF, JDB, ELT and BW declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77351-0>.

**Correspondence** and requests for materials should be addressed to N.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020