Neck Fat Estimation from DXA Using Convolutional Neural Networks

Emily Cresswell[1], Fredrik Karpe[1,2,3], and Nicolas Basty[3(✉)]

[1] Oxford Centre for Diabetes, Endocrinology and Metabolism,
University of Oxford, Oxford, UK
[2] NIHR Biomedical Research Centre, OUH Foundation Trust, Oxford, UK
[3] Research Centre for Optimal Health, School of Life Sciences,
University of Westminster, London, UK
`n.basty@westminster.ac.uk`

Abstract. Expansion of neck adipose tissue (NAT) is an understudied trait in obesity biology. NAT can be quantified using dual X-ray absorptiometry (DXA), a tool commonly used in body composition studies. However, the neck region is not defined in standard regional outputs from a body composition DXA acquisition; instead, quantifying NAT relies on a scanner-dependent software, where manual input is required to define a non-standard region of interest. This makes specialised body composition studies at scale very time-consuming. We thus developed an automated pipeline for NAT estimation from DXA using convolutional neural networks. We investigated whether predicting measurements with a prior step of cropping a region of interest, using automatic landmark prediction, was better than directly predicting from the entire image. We then compared our proposed architecture to the ResNet50 architecture, a well known model used as a basis for transfer learning experiments in many tasks including classification and regression. For the direct and the two-step prediction, both models displayed high performance accuracy. The anatomical landmark placement model performed within three pixels accuracy, and NAT estimation errors for both models were within less than 2.5% points of mean absolute error. However, the proposed model outperformed ResNet50 in the direct prediction, where ours had a mean absolute error of 2.42 against 5.58 for ResNet50. To ensure that the direct predictions are specifically focusing on neck fat, and that the results did not arise from chance correlations with other adipose tissue depots, we generated activation maps which highlighted that the network is focusing on the neck region. This work will enable large-scale population analyses of non-standard regions of interests across a variety of DXA analyses.

Keywords: DXA · Convolutional neural networks · Body composition · Adipose tissue · Neck fat

1 Introduction

The expansion of upper body fat depots is strongly associated with adverse metabolic outcomes; opposite to those of the lower body [11]. In certain

adiposity redistribution syndromes, such as Cushing's disease, obesity hypoventilation syndrome and familial partial lipodystrophy, there is a marked accumulation of neck adipose tissue (NAT), also accompanied by elevated cardiometabolic risk [17]. The neck region is an upper body fat depot comprising both white and thermogenic brown adipose tissue (BAT), the latter of which appears positively associated with metabolic health [2]. It has been speculated that abdominal adiposity is linked to BAT transdifferentiation and replacement by white tissue [3,13]. In the absence of easy and reproducible methods to quantify NAT at scale, investigators have used neck circumference (NC) as a proxy measure. NC has demonstrated strong relationships with visceral adipose tissue; an ectopic fat depot associated with adverse cardiometabolic consequences [16], and type 2 diabetes [4]. NC, however, does not delineate tissue composition, confounding direct NAT measurements due to the effect of neck lean tissue. In addition, studies directly assessing NAT using computed tomography (CT) at a single slice neck vertebral level have suggested positive relationships between NAT volume and cardiometabolic risk [1,21]. However, these small studies did not examine the full length of the neck region, and were limited to healthy cohorts. Dual X-ray Absorptiometry (DXA) is a reliable modality for determining tissue composition [22], specifying the weight and ratios of fat to lean tissue, as opposed to CT-based NAT.

DXA is commonly used for body composition analyses, allowing efficient all-around quantification of fat and lean tissue. It is fast, inexpensive and involves a low radiation dose exposure, which has made it a standard tool for measuring body composition in large cohorts such as the Oxford Biobank (OBB) [12] and the UK Biobank [20]. The standard regional body composition output, automatically included in DXA scans, consists of arm, leg, trunk as well as further subdivisions such as the abdominal and gluteofemoral regions. Additional algorithms within the scanner platforms are used for the quantification of visceral fat content. The neck region is not part of any standard output, requiring the manual definition and construction of regions of interest (ROI) for analysis, which renders it difficult and time-intensive to perform NAT estimations at scale.

Although DXA imaging is commonly used in body composition, it is a less commonly studied imaging modality in the image processing literature. DXA image processing studies generally deal with the skeletal system, for example covering bone segmentation [9], scoliosis prediction [10] or osteoporosis detection [8].

In this study, we have developed and evaluated a two-stage automated end-to-end pipeline for estimation of NAT fat to lean mass ratio in % from DXA images acquired from the OBB, enabling automated estimation of NAT at scale. Analysing these values alongside biochemical parameters from the OBB will thence allow for an in-depth investigation into the role of NAT in metabolic health.

2   Methods

We used the DXA images of 495 OBB participants for this study, including 72 type 2 diabetics. Each individual data set consisted of two images, one low energy X-ray (mostly delineating the soft tissue) and one high energy (mostly delineating the skeletal structure), as well as the NAT mass percentage within a manually defined ROI. The quantitative image software output used as labels in this study came from the GE Lunar iDXA enCORE platform. For each set of images, we manually placed the neck ROI defined at four anatomical landmarks and saved their coordinates: first, two at the level of the right and left lung apices and second, two at the lowest margins of the mandible on the right and the left hand sides. An example of the two X-ray images acquired during a DXA scan are shown in Fig. 1A-B, and the respective neck ROIs for the same subject in Fig. 1C-D. For model training, we performed tenfold data augmentation. This included horizontal flipping, translations of a random number of pixels (-16 to 16), and also small random rotations (-2 to $2°$C), in order to realistically simulate different subjects of various heights and positioned in various orientations. We kept the data for 95 subjects completely apart for final evaluation and we performed all prototyping and parameter tuning experiments with the 400 other data using an 80–20 training and validation split.

In this study, we compare two pipelines for NAT estimation, one predicting the measurement from the entire image and one using a prior step of cropping a ROI of the neck region, as previous studies have shown deep learning tasks to improve when an intermediate step of localisation is added into the pipeline [6,14].

The pipeline using cropped data for prediction consists of two parts: first, the ROI placement, and second; estimation of NAT using the cropped neck ROI, as delineated by the four landmarks. For the first part, we trained a U-net [18] to predict the four coordinates of the neck ROI landmarks. We created label ROI images using the neck landmark coordinates, where each coordinate was represented by a point spread function (PSF) consisting of a Gaussian blur with standard deviation of four pixels, centred on the four landmark coordinates, as has been done by others for landmark detection [23]. We also normalised the two channels of input data as well as the ROI label image. We show an example ROI label image in Fig. 2A. Our U-net-based model has six resolution levels, doubling the number of convolution filters from 64 to 512 in the first four encoding levels, and 512 filters for the fifth, sixth, and the bottom. Each convolutional block has a filter size 3, stride of 2 instead of pooling, batch normalisation and a leaky ReLU activation. The decoding blocks at the same resolution levels match the encoding blocks in terms of filter numbers, and have deconvolutions instead of convolutions. The output layer has a sigmoid activation. We trained the models for 15 epochs minimising a F1 score loss function using the Adam optimiser and a learning rate of 0.002 with decay 0.005. We then calculated the centroids of the final output map predicted by the ROI placement model. We repeated training ten times for the ROI placement-model and assessed performance through mean absolute error (MAE) of euclidean distance in pixels for each of the four landmarks compared to the ground truth coordinates on a held out test set of 95 images.
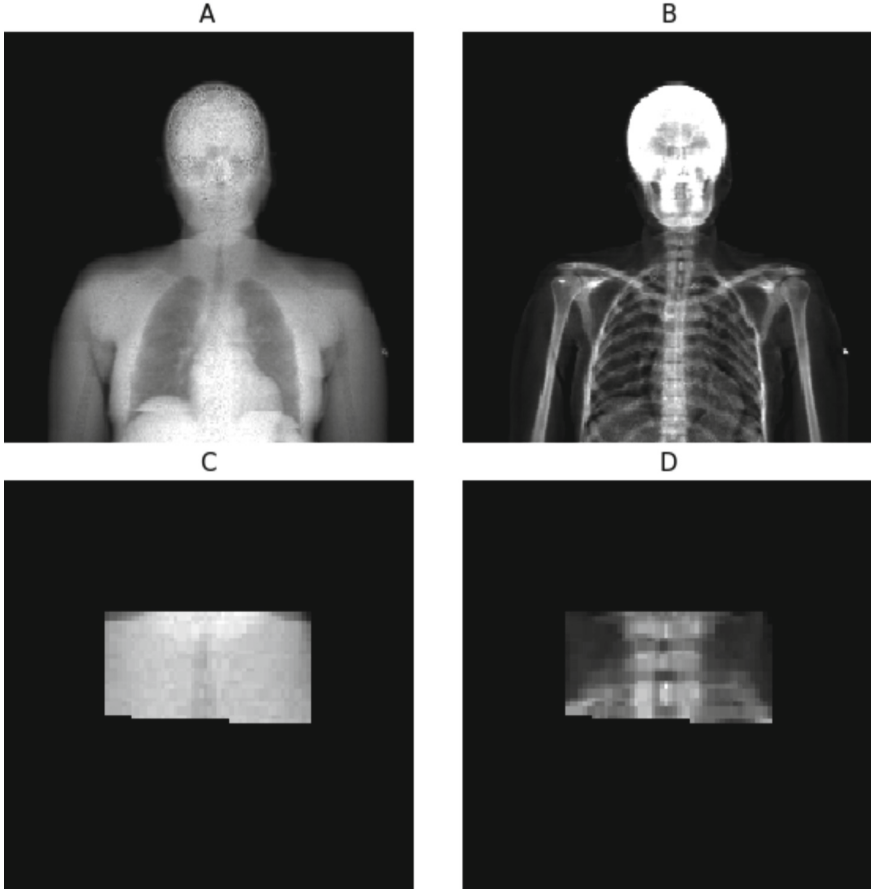
Fig. 1. Example of DXA images and corresponding neck ROI: A) low attenuation energy X-ray, B) high energy X-ray, C) neck ROI of low and D) high attenuation energy X-ray.

For NAT estimation, we adapted a recently published methodology for brain age prediction by Peng et al. [15]. The authors propose a simple fully convolutional neural network that predicts a probability for the age of the subject from 3D brain MRI. Their proposed architecture consists of five convolutional layers followed by two fully connected layers. Our proposed model is constructed using two kinds of convolutional blocks. First: RC_X, a residual block made up of two successive convolutions with X filters of size 3, batch normalisation and ReLU activation where the input and the output of the block are connected by a skip-connection followed by max-pooling. Second: two fully convolutional layers (convolutions of filter size 1), FC_X. The architecture is as follows: $RC_{64} - RC_{128} - RC_{256} - RC_{512} - RC_{1024} - FC_{256} - FC_{65}$. The first FC has ReLU activation, and average pooling as in [15], and the final output has a softmax activation. The network input consists of the neck ROIs of both channels extracted
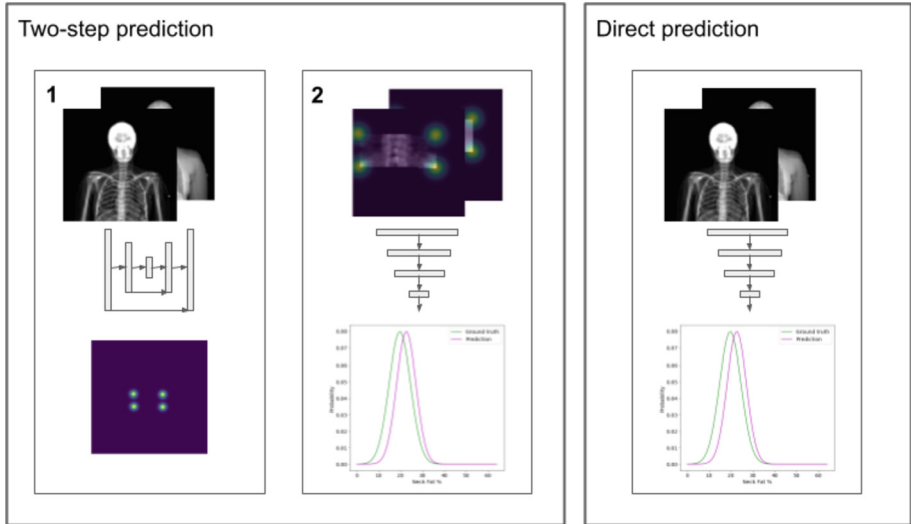
Fig. 2. Graphic representation of the pipelines. Left: two-step prediction showing ROI placement using U-Net (A) and Neural network output estimation of NAT (B). The green distribution represents the true value and magenta for our estimation. Right: direct prediction.

from X-ray images of input size 128×128. The final layer predicts a probability distribution, between 0 and 65, instead of 0 to 50 as in [15]. We selected the range from 0 to 65 to capture the range of the ground truth labels (which ranged from NAT of 7.3% to 53.6%) that we converted into Gaussian distribution centred around the value used as labels. We found the model performance to improve when training was performed minimising categorical cross-entropy and not the Kullback-Leibler divergence loss, as used in Peng et al. [15]. During model prototyping experiments, we explored various values of standard deviation, σ, for the Gaussian distribution around the NAT estimation label (1, 3, 5, 7.5, 10) and found the most robust and repeatable results to be obtained from $\sigma = 5$. We set the number of epochs to 65, after observing that the exploratory training runs, which went on for 100 epochs, converged around epoch 65. We trained the NAT estimation model using the Adam optimiser and a learning rate of 0.01 and decay of 0.001. The output of the network is a probability distribution, and we calculate the final predicted NAT percent values as the mean of the distribution. We repeated training ten times and assessed performance via MAE and root mean square error (RMSE) in percentage points using the 95 held out testing set. Figure 2 shows a graphic describing the pipeline, 1A) showing the ROI placement, 1B) the NAT estimation, and 2) showing the direct prediction pipeline.

We also trained models predicting NAT from the entire top half (containing the neck region) of the DXA images without performing a landmark prediction and ROI cropping step. For both the direct and two-step methods, we compared our proposed architecture against ResNet50 [7]. ResNet50 is a version with 50 layers of the well-known ResNet architecture, which was the first to introduce residual learning. For our ResNet50 comparison experiments, we used transfer learning to retrain the model for NAT prediction. We substituted the final activation layer with the same fully connected layer with 65 outputs as in our proposed model, before retraining it. We also performed visual checks of the activation maps generated using guided back-propagation [19] (adapting code from https://github.com/experiencor/deep-viz-keras/) to ensure that the direct predictions are specifically focusing on neck fat, and that the results are not arising from a chance correlation with other adipose tissue depots. All experiments presented in this paper, including transfer learning experiment using the pre-trained ResNet50 model, were performed using Keras [5].

3   Results

The ROI placement metrics are summarised in Table 1, where we give the MAE for each landmark in pixels (euclidean distance) when compared to the manual placement on the held out test set of 95 data. The results show that the network is able to predict the four landmarks within less than three pixels accuracy. The NAT estimation metrics for both methods using our model and ResNet50 are summarised in Table 2, where we show that the networks in all experiments are able to predict NAT within a few percentage points. We show examples of low, medium, and high NAT out-of-sample subjects in Fig. 3. The first column are the low attenuation X-ray images, the middle columns show the predicted landmark PSF centroids in red superimposed on top of the manually placed ground truth PSFs in white, and the third column shows the network output (magenta) and the distribution centred on the ground truth NAT value (green). We can see in those examples that the network slightly overestimates (by 1.1% points) the low NAT and slightly underestimates (by 2.36% points) the high NAT. We show activation maps using our proposed direct prediction model in Fig. 4.

Table 1. ROI placement performance metrics. The error of the coordinates of the four individual landmarks are given in euclidean distance in pixels (resolution $0.23\,\mathrm{mm}^2$).

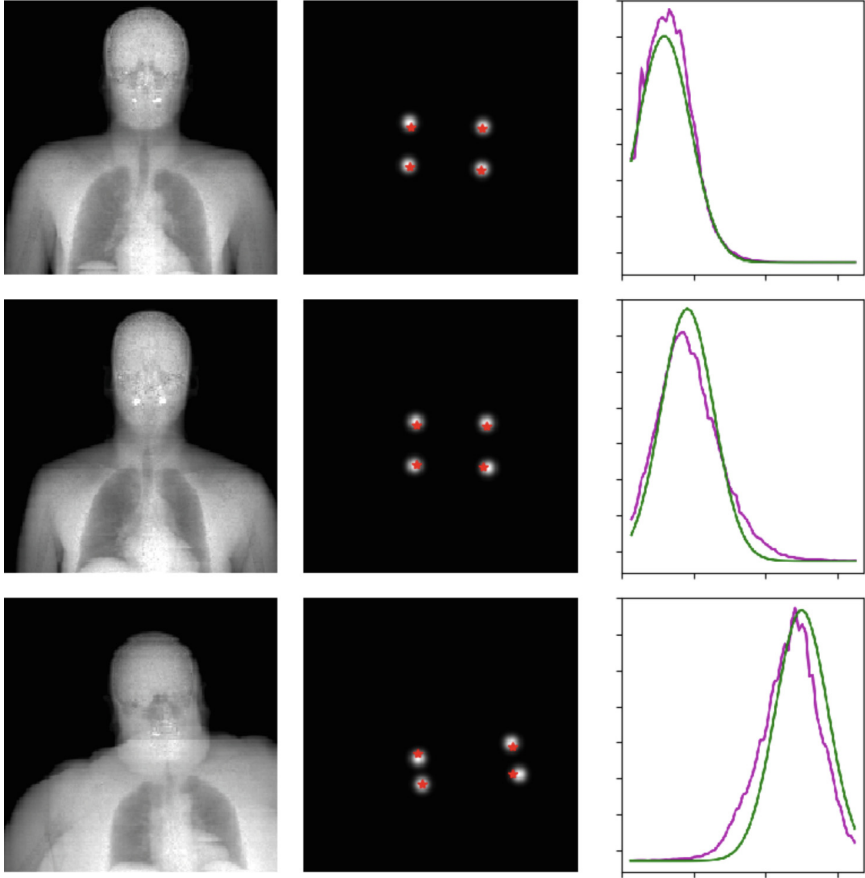| Landmark	Top left	Top right	Bottom right	Bottom left
Mean absolute error	2.0	2.5	1.9	2.1

Fig. 3. ROI and NAT estimation on three out-of-sample data for the two-step model. First column is the low energy X-ray, second the four landmarks (red = estimation, white = PSFs centred on manual landmarks), third the predicted distribution (magenta) and ground truth label (green).

Table 2. Root Mean Square Error and Mean Absolute Error of NAT prediction from two-step and direct methods using the model proposed in this work and ResNet50. The values shown in the table are in percentage points.

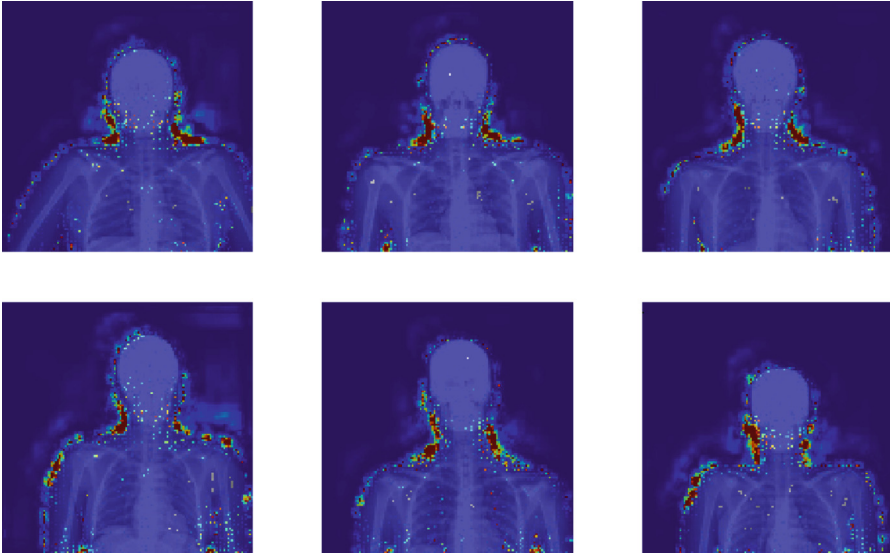| Model	Root mean square error	Mean absolute error
ResNet50 (two-step)	3.03	2.26
This work (two-step)	2.97	2.36
ResNet50 (direct)	6.68	5.58
This work (direct)	3.04	2.42

Fig. 4. Activation maps for six examples of direct NAT estimation using our model, highlighting areas around the neck as the strongest contributors to the model NAT output.

4 Discussion and Conclusion

In this study, we have developed an automatic solution for NAT composition estimation from DXA imaging. We have shown through our experiments that the two proposed methods are able to perform NAT predictions within a small margin of error. Compared to ResNet50, our model to performed better when predicting NAT using the entire image foregoing the cropping step, for which both models performed similarly well. In order to further increase confidence when applying our model to a large cohort, we will increase the training dataset, with a focus on enriching with more extreme NAT values. This method of predicting parameters from DXA images using deep learning will enable a fast way to estimate NAT in the Oxford Biobank at scale, which would have otherwise required many months of manual labour. There is further potential to apply this to other DXA-containing Biobanks, allowing the investigation of specific disease cohorts where NAT is of interest (e.g. Cushing's, familial partial lipodystrophy, obesity hypoventilation syndrome). On top of that, we can easily measure neck diameter using the distance between the predicted landmarks from the ROI model. As most of the existing data is on NC, we will use this to confirm previously established relationships between NC and metabolic health [1,21]. Our research may enable additional parameters to be learnt, and further regions to be predicted in other large-scale image analyses.

References

1. Arias-Tellez, M.J., et al.: Neck adipose tissue accumulation is associated with higher overall and central adiposity, a higher cardiometabolic risk, and a pro-inflammatory profile in young adults (2021)
2. Becher, T., Palanisamy, S., Kramer, D.J., Eljalby, M., Marx, S.J., Wibmer, A.G., Butler, S.D., Jiang, C.S., Vaughan, R., Schöder, H., Mark, A., Cohen, P.: Brown adipose tissue is associated with cardiometabolic health. Nat. Med. **27**(1), 58–65 (2021)
3. Chait, A., den Hartigh, L.J.: Adipose tissue distribution, inflammation and its metabolic consequences, including diabetes and cardiovascular disease. Front. Cardiovasc Med. **7**, 22 (2020)
4. Cho, N.H., et al.: Neck circumference and incidence of diabetes mellitus over 10 years in the Korean genome and epidemiology study (KoGES). Sci. Rep. **5**, 18565 (2015)
5. Chollet, F., et al.: Keras (2015). https://github.com/fchollet/keras
6. Fitzpatrick, J.A., Basty, N., Cule, M., Liu, Y., Bell, J.D., Thomas, E.L., Whitcher, B.: Large-scale analysis of iliopsoas muscle volumes in the UK biobank. Sci. Rep. **10**(1), 1–10 (2020)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Hussain, D., Han, S.M.: Computer-aided osteoporosis detection from DXA imaging. Comput. Methods Programs Biomed. **173**, 87–107 (2019)
9. Hussain, D., Naqvi, R.A., Loh, W.K., Lee, J.: Deep learning in DXA image segmentation (2021)
10. Jamaludin, A., Kadir, T., Clark, E., Zisserman, A.: Predicting scoliosis in DXA scans using intermediate representations (2019)
11. Karpe, F., Pinnick, K.E.: Biology of upper-body and lower-body adipose tissue-link to whole-body phenotypes. Nat. Rev. Endocrinol. **11**(2), 90–100 (2015)
12. Karpe, F., Vasan, S.K., Humphreys, S.M., Miller, J., Cheeseman, J., Dennis, A.L., Neville, M.J.: Cohort profile: the oxford biobank. Int. J. Epidemiol. **47**(1), 21–21g (2018)
13. Kotzbeck, P., et al.: Brown adipose tissue whitening leads to brown adipocyte death and adipose tissue inflammation. J. Lipid Res. **59**(5), 784–794 (2018)
14. Oktay, O., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
15. Peng, H., Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M.: Accurate brain age prediction with lightweight deep neural networks. Med. Image Anal. **68**, 101871 (2021)
16. Preis, S.R., et al.: Neck circumference as a novel measure of cardiometabolic risk: the framingham heart study. J. Clin. Endocrinol. Metab. **95**(8), 3701–3710 (2010)
17. Resende, A.T.P., Martins, C.S., Bueno, A.C., Moreira, A.C., Foss-Freitas, M.C., de Castro, M.: Phenotypic diversity and glucocorticoid sensitivity in patients with familial partial lipodystrophy type 2. Clin. Endocrinol. **91**(1), 94–103 (2019)

18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation (2015)
19. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806 (2014)
20. 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**(3), e1001779 (2015)
21. Torriani, M., Gill, C.M., Daley, S., Oliveira, A.L., Azevedo, D.C., Bredella, M.A.: Compartmental neck fat accumulation and its relation to cardiovascular risk and metabolic syndrome. Am. J. Clin. Nutr. **100**(5), 1244–1251 (2014)
22. Vasan, S.K.: Comparison of regional fat measurements by dual-energy x-ray absorptiometry and conventional anthropometry and their association with markers of diabetes and cardiovascular disease risk. Int. J. Obes. **42**(4), 850–857 (2018)
23. Xue, H., Artico, J., Fontana, M., Moon, J.C., Davies, R.H., Kellman, P.: Landmark detection in cardiac MRI by using a convolutional neural network (2021)