Clinical Radiology 77 (2022) e363–e371

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Fully automated deep-learning section-based muscle segmentation from CT images for sarcopenia assessment

S. Islam [a,b,*], F. Kanavati [a], Z. Arain [a], O. Fadeeva Da Costa [a,b], W. Crum [a], E.O. Aboagye [a], A.G. Rockall [a,b]

[a] Comprehensive Cancer Imaging Centre, Division of Cancer, Dept of Cancer and Surgery, Faculty of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NS, UK
[b] Dept of Radiology, Imperial College Healthcare NHS Trust, Hammersmith Hospital, Du Cane Rd, London, W12 0NS, UK

ARTICLE INFORMATION

Article history:
Received 28 July 2021
Accepted 7 January 2022

U-Net's architecture is developed for biomedical image segmentation. https://paperswithcode.com/method/u-net

AIM: To develop a fully automated deep-learning-based approach to measure muscle area for assessing sarcopenia on standard-of-care computed tomography (CT) of the abdomen without any case exclusion criteria, for opportunistic screening for frailty.

MATERIALS AND METHODS: This ethically approved retrospective study used publicly available and institutional unselected abdominal CT images (n=1,070 training, n=31 testing). The method consisted of two sequential steps: section detection from CT volume followed by muscle segmentation on single-section. Both stages used fully convolutional neural networks (FCNN), based on a UNet-like architecture. Input data consisted of CT volumes with a variety of fields of view, section thicknesses, occlusions, artefacts, and anatomical variations. Output consisted of segmented muscle area on a CT section at the L3 vertebral level. The muscle was segmented into erector spinae, psoas, and rectus abdominus muscle groups. Output was tested against expert manual segmentation.

RESULTS: Threefold cross-validation was used to evaluate the model. Section detection cross-validation error was 1.41 ± 5.02 (in sections). Segmentation cross-validation Dice overlaps were 0.97 ± 0.02, 0.95 ± 0.04, and 0.94 ± 0.04 for erector spinae, psoas, and rectus abdominus, respectively, and 0.96 ± 0.02 for the combined muscle area, with R^2 = 0.95/0.98 for muscle attenuation/area in 28/31 hold-out test cases. No statistical difference was found between the automated output and a second annotator. Fully automated processing took <1 second per CT examination.

CONCLUSIONS: A FCNN pipeline accurately and efficiently automates muscle segmentation at the L3 vertebral level from unselected abdominal CT volumes, with no manual processing step. This approach is promising as a generalisable tool for opportunistic screening for frailty on standard-of-care CT.

* Guarantor and correspondent: S. Islam, Division of Cancer, Department of Cancer and Surgery, Faculty of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NS, UK. Tel.: +44 (0)20 3313 3759.
E-mail address: s.islam@imperial.ac.uk (S. Islam).

https://doi.org/10.1016/j.crad.2022.01.036

Introduction

Single-section computed tomography (CT) analysis has garnered significant clinical interest in the past few years with regard to assessing sarcopenia as a surrogate marker of frailty.[1] Sarcopenia refers to loss of muscle mass and strength and is particularly relevant in oncology where it is typically found to be associated with poor outcomes.[2–5] A section extracted at the L3 vertebra is commonly chosen as a standard landmark by a majority of medical researchers for sarcopenia measurement,[1,6] as muscle and adipose fat areas at L3 have been found to be the most linearly correlated to their whole-body counterparts.[5]

The main motivation for automating the process of single-section CT analysis is to (a) accelerate research for studies that extract single-section sarcopenia-based measurements, and (b) use it as part of a potential opportunistic screening tool that provides prognostic information to clinicians, based on clinically validated measurements, in medical, surgical, or cancer populations alongside the standardised evaluation of CT images at the time of initial staging. Identification of sarcopenia, indicating possible frailty, would allow appropriate stratification of patients for treatment regimes, including pre-habilitation and re-habilitation, with the aim of improving patient outcomes.

The current workflow for single-section sarcopenia measurement includes manual extraction of the L3 section followed by semi-automated segmentation of muscle using software (such as Section-O-Matic or ImageJ), which frequently involves manual refinement. This process takes approximately 10 minutes per image, and it becomes time-consuming to run on large datasets. In addition, due to the manual step, this has not been incorporated routinely into the daily workflow.

Previous work has demonstrated the detection of the L3 vertebral level[6,7] and other studies have described automated segmentation of the abdominal wall musculature.[8–10] Automated segmentation of muscle or body composition has been reported in small single-centre cohorts with single-vendor datasets using the same CT section thickness[11,12]

The objective of the present study was to develop an efficient and generalisable deep-learning method based on fully convolutional neural networks (FCNN) to detect the L3 section automatically and sequentially segment the skeletal muscle at that section into erector spinae, psoas, and rectus abdominus with no manual-processing step, in order to allow a time-efficient tool for fully automated sarcopenia detection in daily practice.[13] A FCNN-based method was motivated by previous work[14–17] that used FCNNs to predict confidence maps for landmark localisation. The intended use would be to provide a fully automatic pipeline for muscle quantification on any CT examination that covers the L3 vertebral level would allow clinical translation of sarcopenia detection as standard, overcoming the current barrier to research and clinical translation caused by the need for manual or semi-automated segmentation.

Materials and methods

This retrospective study aimed to develop and test a model for automated sarcopenia measurement using CT. Ethical approval for retrospective analysis of the data was obtained under the Hammersmith and Queen Charlotte's & Chelsea Research Ethics Committee approval 05/QO406/178, and informed consent was waived, for retrospective analysis of anonymised imaging data. The method included two steps, each of which was tested independently: (1) section detection and (2) muscle segmentation. As a whole, the method takes as input a CT volume and outputs muscle segmentation at the detected L3 section.[6,13]

Dataset

Without any selection criteria, a diverse dataset was collected consisting of CT images of the abdomen in cancer patients from multiple sources, previously de-identified for other studies. Two sets were obtained from the Cancer Imaging Archive (TCIA)[18]: head and neck[19] (n=90) and colon[20,21] (n=674); a liver tumour dataset (n=96) was obtained from the LiTS segmentation challenge.[22] In addition, an ovarian cancer dataset obtained from Hammersmith Hospital, Hammersmith Hospital data was used for model training (n=210) and testing (n=31). This resulted in 1,070 CT volumes from individual patients (Fig 1). No cases were excluded due to metal implants or anatomical variations.

Section detection dataset preparation and annotation

The CT images in the collected datasets had a range of different section thicknesses (range 0.5–7 mm, mean 1.36 mm).[6] The pixel size of the resulting maximal intensity projection (MIP) images was normalised to 1×1 mm^2 to allow consistent input into the section detection algorithm. All the input three-dimensional (3D) CT volumes were first converted into two-dimensional (2D) MIP images along the frontal and sagittal views, similar to;[7] however, a restricted MIP was computed for the sagittal view in order to eliminate the outer edges of the pelvis and to have a clear view of the sacrum vertebrae, which is an essential reference point for determining the position of L3 if the annotator counts the vertebrae bottom-up. As the spine tends to be situated in the middle of the images in the majority of cases, only the restricted sagittal MIP was computed using the range [−20, 20] from the centre of the image; however, for rare cases where the spine is not at the centre of the image, it is potentially possible to use a more elaborate image processing technique to attempt to detect the centre-line of the spine and centre it.

Finally, the images between 100 and 1,500 HU were thresholded in order to eliminate the majority of soft tissue

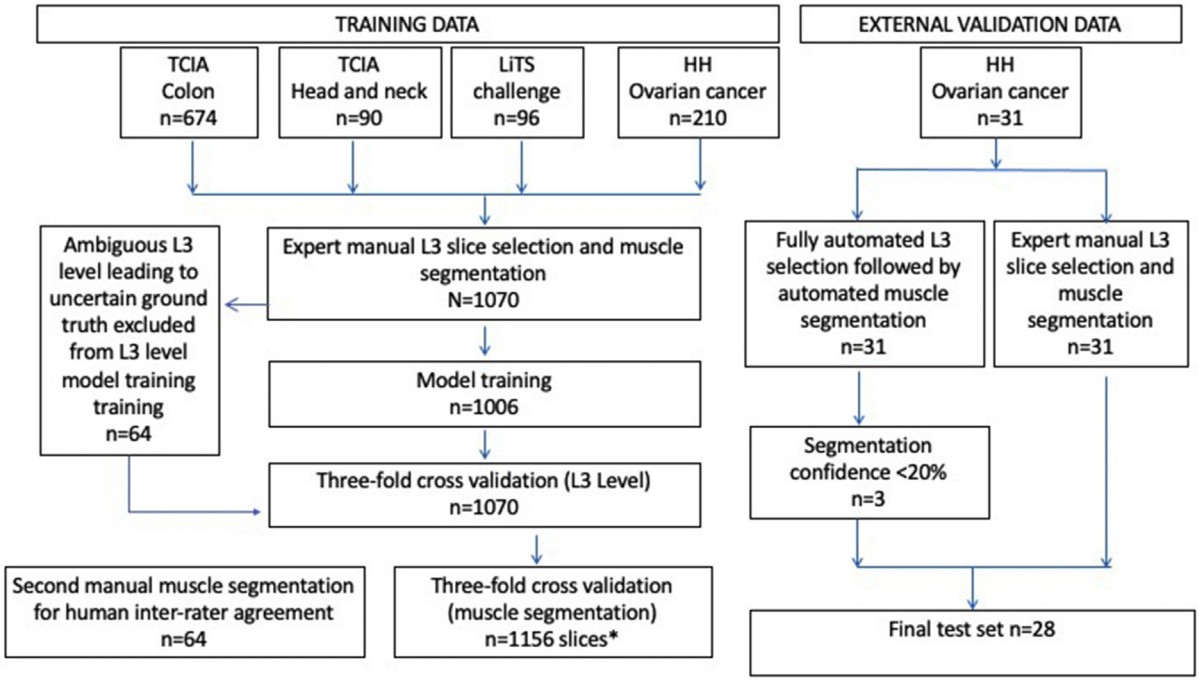

Figure 1 Training and validation dataset curation and workflow. *A total of 1,156 transverse sections were extracted from the 1,070 volumes from around the L3 vertebra, with multiple adjacent sections extracted from some volumes with small section thickness.

at the lower end of attenuation and minimise the effect of metal implants and artefacts above 1,500 HU. The images were then mapped to 8 bit ([−127,127]).

Ground-truth annotations of the L3 section level, using the MIP images were undertaken by two annotators: a radiologist with 8 years of experience and an annotator with 5 years of experience working with CT images. For each image set, the annotator was presented with the frontal and restricted sagittal MIPs side by side, and the annotator clicked on the location of the L3 section. The main landmark was chosen as the middle of the vertebra, lining up with the top edge of the transverse process. Only the position along the *y*-axis was recorded. Fig. 2 shows an example of L3 section annotation. It took about 2–4 seconds to annotate a single image manually. Sixty-four cases had an ambiguous L3 level, with either disagreement or uncertainty in assigning the location of the L3 vertebra. Further inspection by a senior radiologist with 18 years of experience as a consultant revealed that the majority of ambiguous cases consisted of patients with congenital vertebral anomalies, with the principal anomaly being transitional vertebra in the lumbar spine. Transitional vertebrae are those that exhibit ambiguous characteristics; they are relatively common in the population (15–35%),[23,24] and they occur at the junction between spinal segments with various degrees of apparent transition: atlanto-occipital junction, cervicothoracic junction (with a cervical rib from C7), thoracolumbar junction (with lumbar rib at L1 or a 13th rib from T13), and lumbosacral junction (commonly referred to as lumbosacral transitional vertebrae (LSTV)).

Figure 2 Frontal plane MIP with the automated output from the section detection algorithm highlighting the location of the L3 vertebra.

Inaccurate identification of the correct level due to LSTV has led to procedures being carried out at the wrong vertebra level.[25] Correct identification of L3 in ambiguous cases can only be resolved if the image contains a view of the whole spine.[23,26]

The 64 cases with uncertainty or ambiguity for L3 level (transitional vertebra cases (n=57) and seven additional ambiguous cases) were excluded from the training dataset, leaving 1,006 images for the training process; however, the detection algorithm was still evaluated on the transitional cases to verify the output, as in real-world scenarios, such cases are expected to be encountered. The average (rounded down) L3 section location from the two annotators was used as ground truth for training.

Muscle segmentation dataset preparation and annotation

A total of 1,156 transverse sections were extracted from the 1,070 volumes from around the L3 vertebra, with multiple adjacent sections extracted from some volumes with small section thickness. Ground-truth manual segmentation was carried out in ITKsnap[27] by a single radiologist with 8 years of experience; for efficiency while manually segmenting, the extracted sections were stacked up as a single volume to allow loading them all at once within the viewer. The sections were segmented into erector spinae, psoas, and rectus abdominus and the manual segmentations were the segmentation ground truth. Fig. 3 shows examples of the segmentations.

The image intensities were clipped to [−250 HU, 250 HU], so as to allow enough extended range to include the muscle ([−29 HU, 150 HU]). The images were then normalised to the [−1.0, 1.0] range. All the images were of a fixed size of 512 × 512 pixels.

On a subset of the head and neck dataset (n=64), a medical student annotator with 3 years of experience of human anatomy performed an independent section selection at L3 and muscle segmentation. These segmentations were then checked by a senior radiologist. This set was used for evaluation of inter-rater agreement between human annotators.

Deep-learning model

For both the section detection and segmentation, FCNNs that are variants of the UNet architecture were used. The UNet architecture[28] consists of multiple down-sampling (encoder path) and up-sampling (decoder path) blocks, with the latter mirroring the former and with skip connections from the encoder to the decoder path. Each block consists of a set of convolutional units, where each unit is a sequence convolution, batch normalisation, and non-linear activation.

For section detection, the UNet architecture was modified to output one-dimensional (1D) heat maps by introducing a global horizontal max-pooling that reduces the dimension from 2D to 1D by taking the maximum value only along the horizontal line.

Augmentation

Image augmentation was employed to help artificially increase the number of images that the CNN encounters during training. Augmentation was applied randomly in real time to the entire training dataset during the training phase of algorithm development. Augmentation typically helps to improve generalisation performance. Image transformations were applied to the input images create artificial variants. A set of image transformations such as horizontal flipping, scaling, intensity offsets, and piecewise affine deformation were used. For section detection, region drop-outs and over-exposure (to simulate occlusions), and vertical image sub-sampling (to simulate different section thicknesses) were additionally used. In MIP images, occlusions can show up due to the presence of metal implants, bowel content, or contrast agents; region drop-outs and over-exposures can help make the algorithm less susceptible to such occlusions. To simulate images with a variety of section thicknesses (up to 7 mm), images were down-sampled randomly along the vertical axis and then up-sampled back to their original size using linear interpolation.

Data analysis

To assess the performance of the section detection and segmentation, threefold cross-validation was performed in the training dataset. The section detection error was measured using the absolute difference between the ground truth position of L3 and the predicted position. The muscle segmentation accuracy was measured using the Dice score, which is defined as twice the area of the overlap between the predicted segmentation and the ground truth segmentation divided by the sum of both areas. The mean

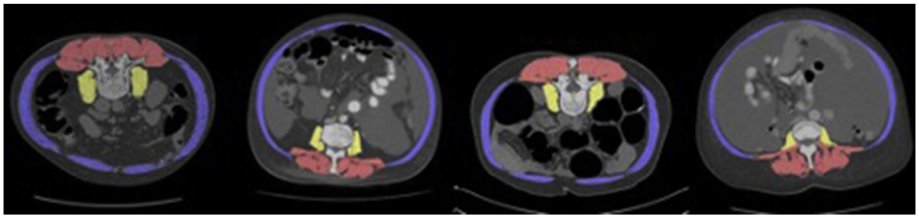

Figure 3 Examples of CT sections with manual segmentations of the muscle into erector spinae, psoas, and rectus abdominus. The colon dataset contained a mix of prone and supine patients.

and standard deviation from the cross-validation were computed.

External validation

Algorithmic performance (section detection and segmentation) was further validated on an external data set (n=31) consisting of whole-body CT from patients with ovarian cancer. A single reader selected the L3 section manually, and performed manual segmentation of the erector spinae, psoas, and rectus abdominus at this level. The muscle area and muscle attenuation were computed from the segmentations using custom MATLAB scripts, applying the same threshold ([−29 HU, 150 HU]) as the automated method to reject voxels of macroscopic fat. The same dataset was then processed by the algorithm to produce L3 section detection and muscle segmentation for comparison. Linear regression was applied to compare the manual and automated methods.

Results

Section detection

The section detection method was evaluated using either the frontal or restricted sagittal MIP as input to the CNN. Additionally, the section detection was applied on the excluded set of transitional vertebra cases. Table 1 summarises the results of threefold cross-validation for section detection (n=1070). The absolute error in position between the automatically detected section and the ground truth section was 1.53 ± 4.22 sections (2.12 ± 4.56 mm, median 1 mm). Outputs in all cases consisted in one of the two adjacent true candidate vertebrae or, in some cases, both. Results reported in Table 2 correspond to the error between human annotators.

Segmentation

dice coeff = F1 score

For muscle segmentation, threefold cross-validation Dice scores were 0.97 ± 0.02, 0.95 ± 0.04, and 0.94 ± 0.04 for erector spinae, psoas, and rectus abdominus, respectively, and 0.96 ± 0.02 for the combined muscle area. Fig. 4 summarises the Dice overlap results. Fig. 5 demonstrates examples of failure cases for segmentation and section detection. Approximately half of segmentation failure cases, considered the worst 5% of Dice score, had a clear reason for the error, such as the presence of abdominal pathology (Fig 5) or transitional lumbosacral vertebra (Electronic Supplementary Material Fig. S1). Other failure cases were predominantly related to (i) inclusion of bowel loops at the periphery of the abdomen or anterior to the vertebra; (ii) inclusion of the tip of right lobe of liver; or (iii) errors due to omission of part of the muscle or extension into adjacent subcutaneous.

Table 1
Summary of results of threefold cross-validation for section detection.

		Mean	SD	Median	Max	> 10
Frontal	Error (mm)	2.12	4.56	1	38	22
	Error (no. of sections)	1.53	4.22	0.67	45.71	15
Sagittal	Error (mm)	1.99	5.41	1	52	28
	Error (no. of sections)	1.41	5.02	0.50	65	23

Threefold cross-validation results for section detection comparing the automatic section detection with the ground truth (n=1,070) with either using the frontal MIP or sagittal MIP as input. The error was obtained by computing the absolute difference in position between the automatically detected section and the ground truth section. The error is reported in millimetres and number of sections, with the latter being computed from the error in millimetres by dividing it by the section thickness for a given image, without rounding. The number of outlier images that have an error >10 is also reported (either in millimetres or number of sections).

Table 2
Summary of error for L3 section level between annotators.

	Mean	SD	Median	Max
Error between A and B (mm)	1.90	1.76	1	9
Error between A and B (number of sections)	1.94	2.36	0.80	11.43
Error between A/B and assigned ground truth (mm)	1.14	0.97	1	5
Error between A/B and assigned ground truth (number of sections)	0.97	1.18	0.40	5.71

Errors are reported in millimetres and in number of sections. To train the section detection algorithm, we used as ground truth the mean section location from both annotators. With the newly assumed ground truth, the error between any of the annotators and the ground truth is reported (the error is the same for both annotators as the absolute difference with their mean is symmetric).

Sarcopenia measurements

Results in Fig 6 shows the Bland–Altman plots for muscle attenuation and muscle area for the subset (n=64) of images that the two annotators independently performed section detection and segmentation on, as well as the independent output of our proposed method. This Bland–Altman test was performed as the exact section selected by an individual annotator or by the automated section level selection may be a nearby section and therefore Dice comparison is not possible in that case. A paired t-test comparing the muscle area and attenuation computed with that of the second annotator showed no statistically significant differences for muscle area (p=0.9503) and muscle attenuation (p=0.822).

Assessment of automated pipeline on external dataset

On the external dataset (n=31), the algorithm rejected three cases due to low section detection confidence of <15%. In the remaining 28 cases, the algorithm correctly identified the correct section (Fig 7a). There was excellent agreement between the external human annotator and automated measurements of muscle attenuation and area with linear trends of R^2 = 0.97 (Fig. 7b) and R^2 = 0.81 (Fig 7c) respectively.

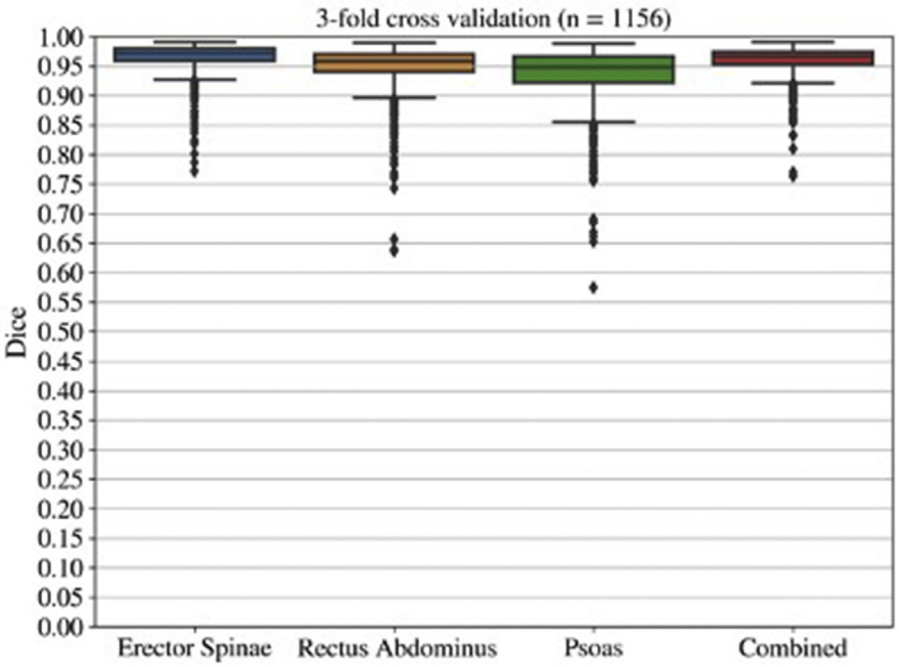

Figure 4 Boxplot of Dice overlap scores between the auto-segmentation output and the ground truth for the different muscle groups as well as combined, obtained via threefold cross-validation.

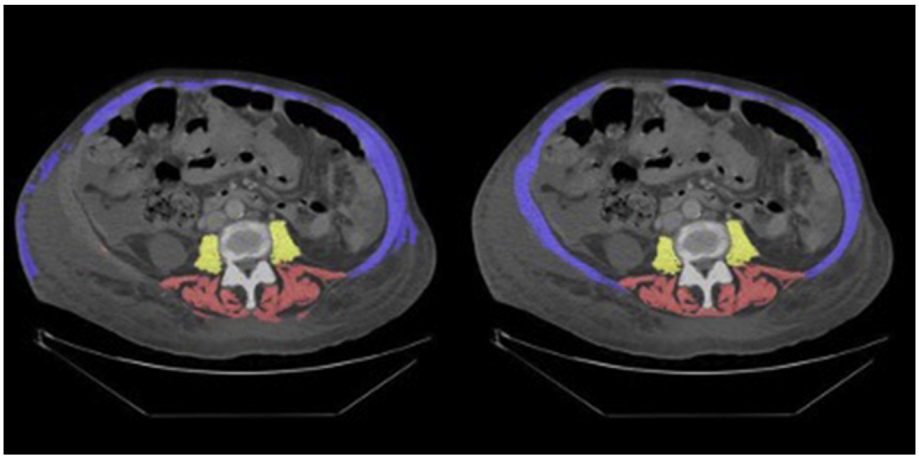

Figure 5 An ovarian cancer patient case with auto-segmentation failure (left). Ground truth on the right. In this challenging case, there is extensive oedema/fluid in the right subcutaneous tissue, possibly due to a drain complication, which may have contributed to the failure.

Discussion

In the present study, a fully-automated CNN model was developed for sequential section detection and muscle segmentation that can identify the mid-L3 vertebral body to within one to two CT sections of the expert reader and then accurately segment the abdominal wall musculature at this level with high accuracy when tested against an expert manual segmentation (Dice score of 0.96 ± 0.02). In the cases in which a second human annotator independently selected a section for segmentation, without reference to the section level selected by the FCNN, there was no significant difference between the fully automated process and the manual annotation for muscle area (p=0.9503) and muscle attenuation (p=0.822). This confirmed that despite slight differences in the final section selection, the muscle segmentation output remained highly comparable. The performance of the algorithm in the hold-out test dataset demonstrated a very high correlation for muscle area ($R^2 = 0.95$) and attenuation ($R^2 = 0.98$); however, the pipeline failed in three cases due to algorithmic low confidence in section selection. On the surface, this is disappointing; however, in practice, the ability of the algorithm to identify low confidence cases, with likelihood of segmentation failure, is of great practical value: a very high correlation in the cases "accepted" by the algorithm resulted in a high correlation of the clinically relevant

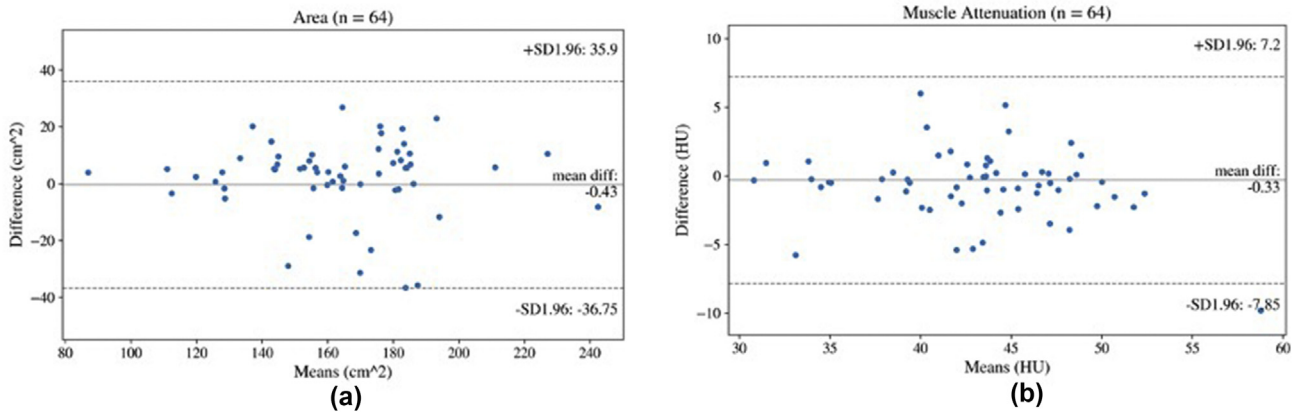

Figure 6 (a) Bland–Altman plot for muscle area for the subset (n=64) of images that the two annotators independently performed section detection and segmentation on. (b) Bland–Altman plot for muscle attenuation for the subset (n=64) of images that the two annotators independently performed section detection and segmentation on.

outputs. It is likely that high confidence cases will not need human checking whereas low confidence case may fail or need to be checked if a segmentation output is provided. Translation of the tool into clinical practice will require further evaluation of the optimal cut-off value of algorithmic confidence that should be recommended.

The pipeline was developed and tested on a range of datasets from multiple sources and two continents, with variable CT section thickness and height. Although cases with an ambiguous L3 level were removed for training, all unselected cases were included in the cross-validation, including cases with scoliosis, degenerative changes, low bone density, and artefacts. This approach should significantly increase the likelihood of generalisability to other datasets. The automated pipeline processing time takes within a second per CT using a standard PC and requires no manual intervention.

Frailty assessments and interventions are the subject of intense research, related to pre- and re-habilitation, in a wide variety of diseases, including cancer, liver disease, as well as in trauma and surgical care.[1–4] The ability to efficiently obtain a muscle area and attenuation by fully automated segmentation will be of value as an opportunistic screening tool. A robust tool that runs efficiently on abdominal CT examinations, with no case selection, was the present aim and the resulting tool is highly promising for clinical translation.

There still remains a small subset of images where the vertebral level prediction is maximally off by one vertebra (outliers are images with an error >10 mm). An inspection of the outlier images revealed that roughly half of them have an apparent reason that might explain why the network made an incorrect prediction. It may be that, despite the best efforts of the annotators, some transitional vertebrae were included in the training set, given that the cases are prevalent in the general population at 15–35%, while only 57 out of 1,070 (5%), were found in the present dataset. With transitional vertebrae cases, it is not possible to determine the correct L3 level without a full view of the spine, which would allow counting from the cervical level. The present network was tested on the excluded set of transitional vertebrae images, the outputted confidence maps always resulted in predictions for one or the other potential L3 candidate, and occasionally both (Electronic Supplementary Material Fig. S1).

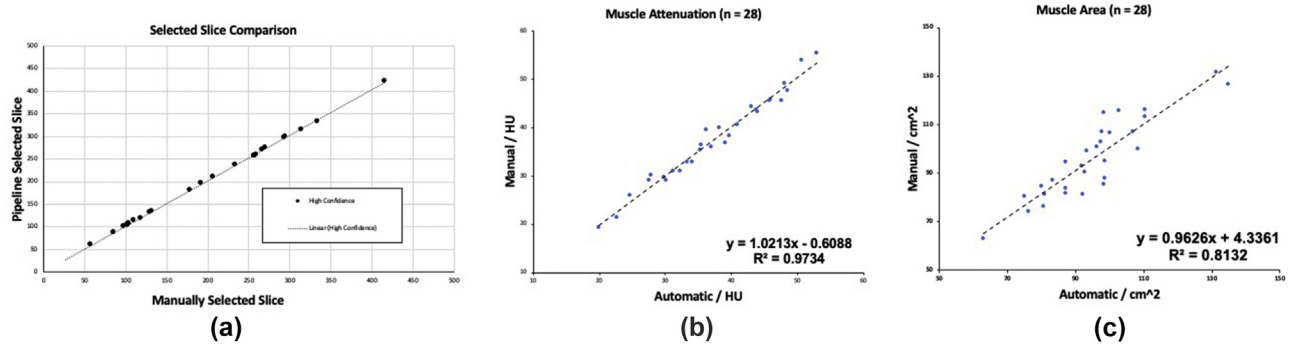

Figure 7 (a) External dataset: comparison of the automated selection of an L3 section with that chosen manually. Three cases with automated selection confidence of <15% are excluded. (b) External dataset: comparison of the mean muscle attenuation measured from automatic segmentation of an automatically selected L3 section with that from manual section selection and segmentation. (c) External dataset: comparison of the mean muscle area measured from automatic segmentation of an automatically selected L3 section with that from manual section selection and segmentation.

Although there was no statistically significant difference observed when comparing sarcopenia measurements obtained with the proposed method to those obtained by a second annotator, a few failure cases remain present. Some causes of segmentation failure related to abdominal wall pathology or due to inclusion of adjacent structures (Fig 5). These identified failures will be the focus of future work.

Previous authors have described efficient automation of vertebral body localisation, with the localisation error being 1.91 ± 2.69 sections (<5 mm).[7] Automated muscle segmentation has also been reported with high accuracy, meeting the performance of expert annotators, with Dice scores above 0.9 in most datasets.[8–10] The present technique utilises a FCNN, which allows inputs of variable sizes. This is especially important when training imaging data of variable matrix size due to differences in scanning resolution, which allows for better generalisability and output performance. The pre-print report and study dataset have now been utilised to successfully transfer the technique to develop an automated sequential pipeline in the paediatric community in a single-centre study using a dataset of 370 patients, benefitting the paediatric community.[11] Automated measurement of more extensive body composition assessment has been reported in a small single-centre study including a total of 50 patients.[12] This is of great interest for the future but the results are based on a small dataset and further model development is awaited.

There are several limitations to the present study, which will form the basis of future work. Sarcopenia assessment is typically done by computing the Skeletal Muscle Index (SMI), which is defined as the skeletal muscle area (square centimetres) divided by patient height (square metre), and muscle attenuation (MA), the average intensity in Hounsfield units within the muscle area. As no patient height data were available for several of the datasets used for model development, only the muscle area as a proxy measure for SMI was reported. A further limitation of this study is the lack of patient outcomes, making it difficult to assess how much of an impact offsets from L3 and differences in segmentation have on the final stratification of patients. Although the results achieved in this study are promising, it remains to be seen how well the method performs on a much larger and more diverse cohort of patients.

Although the L3 section was chosen in the present study, the method is flexible in that it could be generalised to work on different sections at any level. As there is still no universally accepted standard for single-section CT sarcopenia assessment, using a generalised version of the present tool would allow fast computation (within a second) of multiple different measurements in order to evaluate their diagnostic impact for sarcopenia assessment.

In conclusion, an automatic pipeline was developed that segments the muscle at the L3 vertebral level efficiently and accurately in standard of care CT for unselected adults. This tool will be of value as an opportunistic screening tool for frailty in the research setting as well as in patient care.

Conflict of interest

The authors declare no conflict of interest

Acknowledgements

This work was supported by UK National Institute for Health Research Biomedical Research Centre award to Imperial College London and the Cancer Research UK Imperial Centre. The authors acknowledge programmatic support from Imperial College Experimental Cancer Medicines Centres (C37/A7283) and UK Medical Research Council (MR/N020782).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crad.2022.01.036.

References

1. Shachar SS, Williams GR, Muss HB, *et al*. Prognostic value of sarcopenia in adults with solid tumours: a meta-analysis and systematic review. *Eur J Cancer* 2016;**57**:58–67.
2. Vergara-Fernandez O, Trejo-Avila M, Salgado-Nesme N. Sarcopenia in patients with colorectal cancer: a comprehensive review. *World J Clin Cases* 2020;**8**(7):1188–202.
3. Anjanappa M, Corden M, Green A, *et al*. Sarcopenia in cancer: risking more than muscle loss. *Tech Innov Patient Support Radiat Oncol* 2020;**16**:50–7.
4. Mir O, Coriat R, Blanchet B, *et al*. Sarcopenia predicts early dose-limiting toxicities and pharmacokinetics of sorafenib in patients with hepatocellular carcinoma. *PLoS One* 2012;**7**(5):e37563.
5. Kazemi-Bajestani SM, Mazurak VC, Baracos V. Computed tomography-defined muscle and fat wasting are associated with cancer clinical outcomes. *Semin Cell Dev Biol* 2016;**54**:2–10.
6. Kanavati F, Islam S, Aboagye EO, *et al*. *Automatic L3 section detection in 3D CT images using fully-convolutional networks*. arXiv. 2018. p. 181109244.
7. Belharbi S, Chatelain C, Herault R, *et al*. Spotting L3 section in CT scans using deep convolutional network and transfer learning. *Comput Biol Med* 2017;**87**:95–103.
8. Burns JE, Yao J, Chalhoub D, *et al*. A machine learning algorithm to estimate sarcopenia on abdominal CT. *Acad Radiol* 2020;**27**(3):311–20.
9. Lee H, Troschel FM, Tajmir S, *et al*. Pixel-level deep segmentation: artificial intelligence quantifies muscle on computed tomography for body morphometric analysis. *J Digit Imaging* 2017;**30**(4):487–98.
10. Weston AD, Korfiatis P, Kline TL, *et al*. Automated abdominal segmentation of ct scans for body composition analysis using deep learning. *Radiology* 2019;**290**(3):669–79.
11. Castiglione J, Somasundaram E, Gilligan LA, *et al*. Automated segmentation of abdominal skeletal muscle on pediatric CT scans using deep learning. *Radiol Artif Intell* 2021;**3**:e200130.
12. Koitka S, Kroll L, Malamutmann E, *et al*. Fully automated body composition analysis in routine CT imaging using 3D semantic segmentation convolutional neural networks. *Eur Radiol* 2021;**31**(4):1795–804.
13. Kanavati F, Islam S, Zohaib A, *et al*. *Fully-automated deep learning section-based muscle estimation from CT images for sarcopenia assessment*. 2020. arXiv:200606432v1.
14. Tompson J, Goroshin R, Jain AL, *et al*. Efficient object localization using convolutional networks. In: *2015 IEEE conference on computer vision and pattern recognition (CVPR), 7-12 june, Boston, MA*. Piscataway, NJ: IEEE; 2015. p. 648–56.
15. Pfister T, Charles J, Zisserman A. Flowing convnets for human pose estimation in videos. In: *2015 IEEE international conference on computer*

vision (ICCV). Piscataway, NJ: IEEE; 2015. p. 1913–21. https://doi.org/10.1109/ICCV.2015.222.
16. Payer C, Stern D, Bischof H, *et al*. Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselin S, Joskowicz L, Sabuncu M, *et al*., editors. *Medical image computing and computer-assisted intervention – MICCAI 2016. MICCAI 2016. Lecture notes in computer science*vol. 9901. Cham: Springer; 2016, https://doi.org/10.1007/978-3-319-46723-8_27.
17. Cao Z, Simon T, Wei S-E, *et al*. Realtime multi-person 2D pose estimation using part affinity fields. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI*. Piscataway, NJ: IEEE; 2017. p. 1302–10. https://doi.org/10.1109/CVPR.2017.143.
18. Clark K, Vendt B, Smith K, *et al*. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;**26**:1045–57.
19. Grossberg A, Mohamed A, Elhalawani H, *et al*. *Data from head and neck cancer CT atlas*. The Cancer Imaging Archive; 2017, https://doi.org/10.7937/K9/TCIA.2017.umz8dv6s.
20. Smith K, Clark K, Bennett W, *et al*. *Data from CT_COLONOGRAPHY*. The Cancer Imaging Archive; 2015, https://doi.org/10.7937/K9/TCIA.2015.NWTESAY1.
21. Johnson CD, Chen MH, Toledano AY, *et al*. Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med* 2008;**359**(12):1207–17.
22. Bilic P, Christ PF, Eea Vorontsov. *The liver tumor segmentation benchmark (LiTS)*. 2019. arXiv. 1901.04056.
23. Carrino JA, Campbell Jr PD, Lin DC, *et al*. Effect of spinal segment variants on numbering vertebral levels at lumbar MR imaging. *Radiology* 2011;**259**(1):196–202.
24. Ucar D, Ucar BY, Cosar Y, *et al*. Retrospective cohort study of the prevalence of lumbosacral transitional vertebra in a wide and well-represented population. *Arthritis* 2013;**2013**:461425.
25. Konin GP, Walz DM. Lumbosacral transitional vertebrae: classification, imaging findings, and clinical relevance. *AJNR Am J Neuroradiol* 2010;**31**(10):1778–86.
26. Bron JL, van Royen BJ, Wuisman PI. The clinical significance of lumbosacral transitional anomalies. *Acta Orthop Belg* 2007;**73**(6):687–95.
27. Yushkevich PA, Piven J, Hazlett HC, *et al*. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;**31**(3):1116–28.
28. Ronneberger O, Fischer P, Brox T. Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, *et al*., editors. *Medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. Lecture notes in computer science*vol. 9351. Cham: Springer; 2015, https://doi.org/10.1007/978-3-319-24574-4_28.