

An externally validated deep learning model for the accurate segmentation of the lumbar paravertebral muscles

Frank Niemeyer¹, Annika Zanker¹, René Jonas¹, Youping Tao¹,
Fabio Galbusera^{2,*}, Hans-Joachim Wilke¹

¹ Institute of Orthopaedic Research and Biomechanics, Center for Trauma Research Ulm, Ulm University, Ulm, Germany; ² IRCCS Istituto Ortopedico Galeazzi, Milan, Italy

Abstract

Purpose. Imaging studies about the relevance of muscles in spinal disorders, and sarcopenia in general, require the segmentation of the muscles in the images which is very labour-intensive if performed manually and poses a practical limit to the number of investigated subjects. This study aimed at developing a deep learning-based tool able to fully automatically perform an accurate segmentation of the lumbar muscles in axial MRI scans, and at validating the new tool on an external dataset.

Methods. A set of 60 axial MRI images of the lumbar spine was retrospectively collected from a clinical database. Psoas major, quadratus lumborum, erector spinae, and multifidus were manually segmented in all available slices. The dataset was used to train and validate a deep neural network able to segment muscles automatically. Subsequently, the network was externally validated on images purposely acquired from 22 healthy volunteers.

Results. The Jaccard index for the individual muscles calculated for the 22 subjects of the external validation set ranged between 0.862 and 0.935, demonstrating a generally excellent performance of the network. Cross-sectional area and fat fraction of the muscles were in agreement with published data.

Conclusions. The externally validated deep neural network was able to perform the segmentation of the paravertebral muscles in axial MRI scans in an accurate and fully automated manner, and is therefore a suitable tool to perform large-scale studies in the field of spinal disorders and sarcopenia, overcoming the limitations of non-automated methods.

Keywords

Deep learning; segmentation; sarcopenia; paravertebral muscles; fat fraction; lumbar spine

* Corresponding author: Fabio Galbusera, PhD; IRCCS Istituto Ortopedico Galeazzi, via Galeazzi 4, 20161 Milan, Italy; email: fabio.galbusera@grupposandonato.it; phone: +39.0266214939.

Introduction

The scientific study of sarcopenia gained high attention in recent years and is nowadays one of the most discussed issues in the field of the health of ageing subjects. According to the European Working Group on Sarcopenia in Older People, sarcopenia is defined as a “syndrome characterized by progressive and generalized loss of skeletal muscle mass and strength” [1]. Sarcopenia is a multifactorial condition associated with age-related reduction of physical activity, protein intake, anabolic hormonal activity, and vitamin D levels, as well as to a pro-inflammatory status due to increased intracellular oxidative stress [2]. The decrease of hormonal levels, especially testosterone, estrogen, and growth hormone, has been found to be associated with muscle wasting [3].

Several studies associated sarcopenia with spinal disorders. Toyoda and colleagues demonstrated that sarcopenia is correlated with back muscle strength and in turn with the presence of spinal degeneration [4]. The research highlighted that the association may be seen in two distinct ways: sarcopenia could induce spinal disorders, pain and spine-related disability, but also spinal disorders may cause muscle waste and increased fatty infiltration, probably due to the reduced physical activity associated with pain and fear-avoidance behavior. Indeed, sarcopenic patients were reported to show anxiety and catastrophizing signs more commonly than the age-matched population [5], initiating a vicious circle in which reduced physical activity due to psychological issues further promotes muscle waste. A correlation between sarcopenia and degenerative scoliosis, possibly involving reduced bone mineral density, has also been shown [6].

In general, the study of sarcopenia as a possible risk factor for spinal disorders, a consequence of it or its role as a risk factor enhancing symptoms and disability in patients suffering from spinal disorders is at its infancy, since most studies are very recent and have been conducted on relatively small patient cohorts. While the importance of sarcopenia has been undoubtedly recognized, large-scale studies quantifying it and investigating it in detail are still lacking.

Several imaging modalities, namely dual-energy X-ray absorptiometry (DXA), computed tomography (CT), magnetic resonance (MR), and ultrasound (US), have been used to investigate sarcopenia as well as for diagnostic purposes with the aim of quantifying muscle mass. MRI is recently gaining a wide interest since it allows measuring the muscle mass and the amount of fat infiltration inside muscle tissue, as well as additional information regarding muscular oedema, fibrous infiltration, fiber contractility, and elasticity [7], without any exposure to ionizing radiation. Fat and water discrimination can be done using multiecho gradient-echo sequences like the Dixon technique, which allows obtaining a quantitative

assessment of intramuscular fat [8]. However, it should be noted that the use of MRI to study and diagnose sarcopenia requires the segmentation of the images, i.e. the identification of the contours of the muscle mass or of the individual muscles in each slice. Although semi-automated and fully automated methods to segment MRI scans have been presented [9–12], in most clinical papers about sarcopenia this step has been conducted by human operators since such methods are not publicly available and their performance is generally lower than that of expert operators, not generalizing well to pathological cases [11]. Besides, manual segmentation requires a substantial amount of work and practically poses a limit to the sample size, since processing high numbers of subjects such as several thousands requires a time not compatible with the length of most research projects.

The aim of this study is to develop and validate an accurate automated tool to perform the segmentation of axial MRI scans of the paravertebral muscles in the lumbar region, to be used for the assessment and diagnosis of sarcopenia as well as of the condition of the lumbar muscles in general. The automated method shall allow processing large amounts of data, thus allowing the imaging investigation of sarcopenia and its association with spinal disorders in large-scale populations. The secondary aim of the work is to externally validate the new algorithm in a population of 22 healthy volunteers, in order to quantitatively check its ability to deal with a wide scenario of ages and body sizes.

Materials and methods

Training data

A set of 60 axial images of the lumbar spine was retrospectively collected from the imaging database of University Hospital Ulm (Ulm, Germany). The retrospective data collection has been approved by the ethics committee of Ulm University (approval nr. 50/20). All slices relative to the L1-L5 region were manually segmented by means of ITK-SNAP free software (<http://www.itksnap.org>) by a single operator, using a published method as a reference [13]. The following muscles were segmented, depicting the left and right sides as separate entities: psoas major, quadratus lumborum, erector spinae, multifidus. The segmented images were randomly split into training and internal validation sets based on an 80-20% ratio.

Network architecture

The architecture of the deep artificial neural network used for the segmentation of the muscles was derived from the two-dimensional U-Net [14], with several modifications: (1) Leaky ReLU ($\alpha = 0.01$) and softmax as activation functions instead of ReLU and sigmoid; (2) transposed

convolutions for upsampling; (3) dropout ($p = 0.03$) added to the encoder; (4) one additional layer in both encoder and decoder (Fig. 1). The network was trained using the Adam optimizer, with a learning rate of 0.001 and a minibatch size of 16. The architecture and the hyperparameters were extensively optimized by checking the performance of the network on the internal validation set.

Metrics

The “Intersection over Union”, in short IoU and also known as Jaccard similarity coefficient [15], was used as a metric to evaluate the quality of the segmentations generated by it. For a specific muscle C , the relative IoU can be calculated as follows:

$$I_C = \{i \in P \mid y_{true}(i) = y_{pred}(i) = C\} \quad (1)$$

$$U_C = \{i \in P \mid y_{true}(i) = C \vee y_{pred}(i) = C\} \quad (2)$$

$$IoU_C = \frac{|I_C|}{|U_C|} \quad (3)$$

where i is a pixel belonging to the image P (intended as a set of pixels), C is the muscle of interest, $y_{true}(i)$ is the ground truth for the specific pixel i , $y_{pred}(i)$ is the value predicted by the neural network, I_C is the set of pixels corresponding to the muscle C in both the ground truth and the network output, U_C is the set of pixels corresponding to the muscle C in either the ground truth or the network output. The value of IoU is therefore 1 if the network output corresponds perfectly to the ground truth, and 0 if there is no overlap between the ground truth and the predictions.

For the sake of training the neural network, the sum of the IoU calculated among all segmented muscles was used as loss function:

$$loss = \sum_{C=1}^n IoU_C \quad (4)$$

where n is the number of segmented muscles (8 in the present study).

External validation

In order to test the performance of the segmentation tool on an external dataset, axial MRI scans of the whole lumbar spine were prospectively collected for 22 healthy volunteers, 10 males (mean age: 38.6 years (range 25-61), mean weight: 83.3 kg (68-120), mean height: 179.4

cm (170-185)) and 12 females (mean age: 33.5 years (range 22-57); mean weight: 57.5 kg (48-68); mean height: 166.3 cm (153-178)). The prospective image collection was approved by the ethics committee of Ulm University (approval nr. 255/20). The whole lumbar spine of all subjects was imaged with a 3T MRI system (MAGNETOM Skyra, Siemens Healthineers AG, Erlangen, Germany) by using standard imaging protocols. All images were manually segmented by the same operator who performed the segmentation of the training data using the same software tools and methods, as well as by the deep neural network. The manual segmentations were used as the reference for the quantitative assessment of the performance of the automated tool, which was conducted by calculating the *IoU* of the individual paravertebral muscles for all slices.

The segmented images generated by the neural network were used to calculate the cross-sectional area of the individual paravertebral muscles for the 22 volunteers, exploiting the information about pixel size and slice spacing included in the DICOM files (attributes (0028, 0030) and (0018, 0088) respectively). Furthermore, the fat fraction inside each muscle was determined by attributing each voxel to either fat or lean tissue, by using a simple Otsu binary thresholding implemented in the scikit-image Python library for image processing (<https://scikit-image.org/>).

Results

The neural network provided excellent outputs from a qualitative point of view (Fig. 2). A three-dimensional reconstruction of the shape of the paravertebral muscles showed realistic anatomies from a perceptual point of view, with smooth surfaces and limited artefacts (Fig. 3). The median values of the *IoU* for the individual muscles calculated during the external validation ranged between 0.862 and 0.935 (Table 1), demonstrating a generally excellent agreement between the ground truth and the outputs of the neural network. The psoas major and the erector spinae showed a generally higher performance with respect to the quadratus lumborum and the multifidus (Fig. 4), suggesting that a higher cross-sectional area was associated with a higher *IoU*. Nevertheless, the lower quartile of the statistical distribution of the *IoU* was in all cases higher than 0.75, indicating a high quality of the output in the vast majority of the processed slices.

Among the 22 subjects included in the external validation, the male subjects showed generally higher cross-sectional areas of all muscles with respect to females (Fig. 5). The erector spinae was the larger muscle in all volunteers, followed by the psoas major in males while in females

the size of the latter muscle generally overlapped those of the quadratus lumborum and multifidus. The fat fraction was larger in females than in males (Fig. 6) and generally increased with age. A non-negligible variability of the fat fraction was found among the individual muscles, with the multifidus and the erector spinae generally showing higher fatty infiltration with respect to psoas major and quadratus lumborum.

Discussion

In this paper, we presented a novel tool based on artificial intelligence to perform the automatic segmentation of the paravertebral muscles in axial MRI scans. The tool demonstrated excellent performance, potentially enabling its use for processing large amounts of data in retrospective population studies about sarcopenia and its relation with spinal disorders. Since manual segmentation has always constituted a bottleneck for imaging studies about muscles due to its high requirements in terms of manual labour, existing works have been generally limited to dozens or hundreds of cases (e.g. [16, 17], with large populations consisting of several thousand subjects being practically not manageable. An accurate segmentation tool that does not require any human intervention can therefore act as the key enabling technology for large-scale studies. Sarcopenia, in particular in terms of reduced mass and quality of the back muscles, is associated with higher disability and lower quality of life in patients suffering from spinal disorders with respect to non-sarcopenic patients [18]. Sarcopenic patients subjected to spine surgery have a slower and sometimes less complete recovery, more severe postoperative symptoms and lower satisfaction, whereas the correlation between sarcopenia and increased risk of postoperative complications could not be proved [19]. In general, despite the literature indicating that sarcopenia is a highly relevant risk factor for patients suffering from spinal disorders in terms of worse outcomes, relatively little research has been performed in the field, and the availability of fully automated imaging segmentation tools can definitely play a role in future studies addressing this need.

Other automated solutions based on neural networks aimed at segmenting the lumbar paravertebral muscles have been previously presented. Li et al. [12] developed a U-Net based network to perform the segmentation of multifidus and erector spinae, obtaining average Dice similarity coefficients of 0.949 and 0.913 respectively, generally in line with the performance of the current model while taking into account that the metrics used in the present and in the literature study are not directly comparable. Zhang et al. developed another U-Net based model and achieved similar performances [20]. However, it should be noted that in both studies psoas major and quadratus lumborum were not segmented, and that an internal validation on a test

set randomly selected from the original database rather than an external validation of a distinct dataset was performed. Xia and colleagues [11] tested several network architectures, including the U-Net as well as novel solutions, to segment psoas major, erector spinae, and multifidus, obtaining excellent Dice similarity scores between 0.913 and 0.95 for the best-performing neural network. Other studies achieved relatively lower performances [21], or focused on different imaging modalities or fields of view [22, 23]. In general, it can be concluded that our model achieved state-of-the-art performance and is the only one that has been externally validated on a purposely created dataset so far.

The results extracted from the population recruited for the external validation, i.e. the cross-sectional areas of the individual paravertebral muscles and their fat fraction, were in good agreement with existing research, further consolidating the validity of the novel model. Cooper et al. measured cross-sectional areas very similar to those in the present study for the paraspinal muscles (erector spinae and multifidus) as well as for the psoas major as well as the same gender-based differences, including the same disproportion between the size of the psoas in males and females with respect to the other muscles [24]. Fat fractions were also in line with previous observations [16, 25].

The study suffers from some limitations. First, the manual segmentations of the training set and for the external validation were performed only once by a single operator, and a quantitative assessment of the reliability and quality of the ground truth data could therefore not be performed. Besides, a single network architecture was employed as a basis for the development and optimization of the model; while U-Net is largely considered as a state-of-the-art solution for the segmentation of medical images [26], other solutions such as for example generative adversarial networks may be worthy of investigation. Finally, the subjects recruited for the external validation were in a relatively low number and, while they covered a rather wide range of age and body sizes, were all asymptomatic. Therefore, the performance of the tool was not validated for severe pathological cases, which can be nevertheless highly relevant from a clinical point of view.

In conclusion, the tool here presented was able to perform the automated segmentation of the paravertebral muscles in axial MRI scans in an accurate manner in an external validation set of patients, and is therefore a suitable tool to perform future large-scale studies aimed at analyzing the size and quality of the muscles of the lumbar region at the population level.

Figure captions

Figure 1. The architecture of the neural network derived from the two-dimensional U-Net employed in this study.

Figure 2. Segmentations of the paravertebral muscles for randomly selected slices of eight representative subjects recruited for the external validation. The average *IoU* for the eight muscles is also shown. Green/blue: psoas major; orange/purple: quadratus lumborum; brown/yellow: erector spinae; gray/pink: multifidus.

Figure 3. Three-dimensional reconstructions of the segmented paravertebral muscle for four representative volunteers. R: right; L: left; S: superior; I: inferior.

Figure 4. Boxplots and jittered scatter plots of the *IoU* calculated for the individual muscles in all slices acquired for the 22 subjects recruited for the external validation.

Figure 5. Cross-sectional areas (mean value and standard deviation, in mm²) of the eight paravertebral muscles for the 22 subjects (top: females, bottom: males) in ascending order.

Figure 6. Fat fraction in the paravertebral muscles depending on age and stratified by sex (top) and for the individual muscles (bottom) for the 22 subjects.

References

1. Cruz-Jentoft AJ, Landi F, Schneider SM (2014) Prevalence of and interventions for sarcopenia in ageing adults: a systematic review. Report of the International Sarcopenia Initiative (EWGSOP and IWGS).
2. Brioché T, Pagano AF, Py G, Chopard A (2016) Muscle wasting and aging: experimental models, fatty infiltrations, and prevention. *Mol Aspects Med* 50:56-87. doi: 10.1016/j.mam.2016.04.006.
3. Marty E, Liu Y, Samuel A, et al (2017) A review of sarcopenia: Enhancing awareness of an increasingly prevalent disease. *Bone* 105:276–286. <https://doi.org/10.1016/j.bone.2017.09.008>
4. Toyoda H, Hoshino M, Ohyama S, et al (2019) The association of back muscle strength and sarcopenia-related parameters in the patients with spinal disorders. *Eur Spine J* 28:241–249. <https://doi.org/10.1007/s00586-018-5858-8>
5. Wada T, Tanishima S, Osaki M, et al (2019) Relationship between sarcopenia and pain catastrophizing in patients with lumbar spinal stenosis: A cross-sectional study. *Osteoporos Sarcopenia* 5:132–136. <https://doi.org/10.1016/j.afos.2019.12.001>

6. Eguchi Y, Suzuki M, Yamanaka H, et al (2017) Associations between sarcopenia and degenerative lumbar scoliosis in older women. *Scoliosis Spinal Disord* 12:9. <https://doi.org/10.1186/s13013-017-0116-0>
7. Fischer MA, Pfirrmann CWA, Espinosa N, et al (2014) Dixon-based MRI for assessment of muscle-fat content in phantoms, healthy volunteers and patients with achillodynia: comparison to visual assessment of calf muscle quality. *Eur Radiol* 24:1366–1375
8. Li K, Dortch RD, Welch EB, et al (2014) Multi-parametric MRI characterization of healthy human thigh muscles at 3.0 T - relaxation, magnetization transfer, fat/water, and diffusion tensor imaging. *NMR Biomed* 27:1070–1084. <https://doi.org/10.1002/nbm.3159>
9. Mhuiris ÁN, Volken T, Elliott JM, et al (2016) Reliability of quantifying the spatial distribution of fatty infiltration in lumbar paravertebral muscles using a new segmentation method for T1-weighted MRI. *BMC Musculoskelet Disord* 17:234. <https://doi.org/10.1186/s12891-016-1090-z>
10. Antony J, McGuinness K, Welch N, et al (2016) An interactive segmentation tool for quantifying fat in lumbar muscles using axial lumbar-spine MRI. *IRBM* 37:11–22. <https://doi.org/10.1016/j.irbm.2015.10.004>
11. Xia W, Fortin M, Ahn J, et al (2019) Automatic Paraspinal Muscle Segmentation in Patients with Lumbar Pathology Using Deep Convolutional Neural Network. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Springer International Publishing, pp 318–325
12. Li H, Luo H, Liu Y (2019) Paraspinal Muscle Segmentation Based on Deep Neural Network. *Sensors* 19.: <https://doi.org/10.3390/s19122650>
13. Crawford RJ, Cornwall J, Abbott R, Elliott JM (2017) Manually defining regions of interest when quantifying paravertebral muscles fatty infiltration from axial magnetic resonance imaging: a proposed method for the lumbar spine with anatomical cross-reference. *BMC Musculoskelet Disord* 18. <https://doi.org/10.1186/s12891-016-1378-z>
14. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, pp 234–241
15. Jaccard P (1912) The distribution of the flora in the alpine zone.1. *New Phytol* 11:37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
16. Khil EK, Choi J-A, Hwang E, et al (2020) Paraspinal back muscles in asymptomatic volunteers: quantitative and qualitative analysis using computed tomography (CT) and magnetic resonance imaging (MRI). *BMC Musculoskelet Disord* 21:403. <https://doi.org/10.1186/s12891-020-03432-w>
17. Fortin M, Gibbons LE, Videman T, Battié MC (2015) Do variations in paraspinal muscle morphology and composition predict low back pain in men? *Scand J Med Sci Sports* 25:880–887. <https://doi.org/10.1111/sms.12301>
18. Virk S, Wright-Chisem J, Sandhu M, et al (2021) A Novel Magnetic Resonance Imaging-based Lumbar Muscle Grade to Predict Health-related Quality of Life Scores Among Patients Requiring Surgery. *Spine* 46:259–267. <https://doi.org/10.1097/BRS.0000000000003833>
19. Wu W-T, Lee T-M, Han D-S, Chang K-V (2021) The Prevalence of Sarcopenia and Its Impact on Clinical Outcomes in Lumbar Degenerative Spine Disease—A Systematic Review and Meta-Analysis. *J Clin Med Res* 10:773. <https://doi.org/10.3390/jcm10040773>

20. Zhang Y, Shi Z, Wang H, et al (2019) LumNet: A Deep Neural Network for Lumbar Paraspinal Muscles Segmentation. In: AI 2019: Advances in Artificial Intelligence. Springer International Publishing, pp 574–585
21. Shen H, Huang J, Zheng Q, et al (2021) A Deep-Learning–Based, Fully Automated Program to Segment and Quantify Major Spinal Components on Axial Lumbar Spine Magnetic Resonance Images. *Phys Ther* 101. <https://doi.org/10.1093/ptj/pzab041>
22. Kamiya N, Li J, Kume M, et al (2018) Fully automatic segmentation of paraspinal muscles from 3D torso CT images via multi-scale iterative random forest classifications. *Int J Comput Assist Radiol Surg* 13:1697–1706. <https://doi.org/10.1007/s11548-018-1852-1>
23. Castiglione J, Somasundaram E, Gilligan LA, et al (2021) Automated Segmentation of Abdominal Skeletal Muscle on Pediatric CT Scans Using Deep Learning. *Radiology: Artificial Intelligence* 3:e200130. <https://doi.org/10.1148/ryai.2021200130>
24. Cooper RG, Holli S, Jayson MI (1992) Gender variation of human spinal and paraspinal structures. *Clin Biomech* 7:120–124. [https://doi.org/10.1016/0268-0033\(92\)90025-Y](https://doi.org/10.1016/0268-0033(92)90025-Y)
25. Zhao Y, Huang M, Serrano Sosa M, et al (2019) Fatty infiltration of paraspinal muscles is associated with bone mineral density of the lumbar spine. *Arch Osteoporos* 14:99. <https://doi.org/10.1007/s11657-019-0639-5>
26. Mittal M, Arora M, Pandey T, Goyal LM (2020) Image Segmentation Using Deep Learning Techniques in Medical Images. In: Verma OP, Roy S, Pandey SC, Mittal M (eds) *Advancement of Machine Intelligence in Interactive Medical Image Analysis*. Springer Singapore, Singapore, pp 41–63

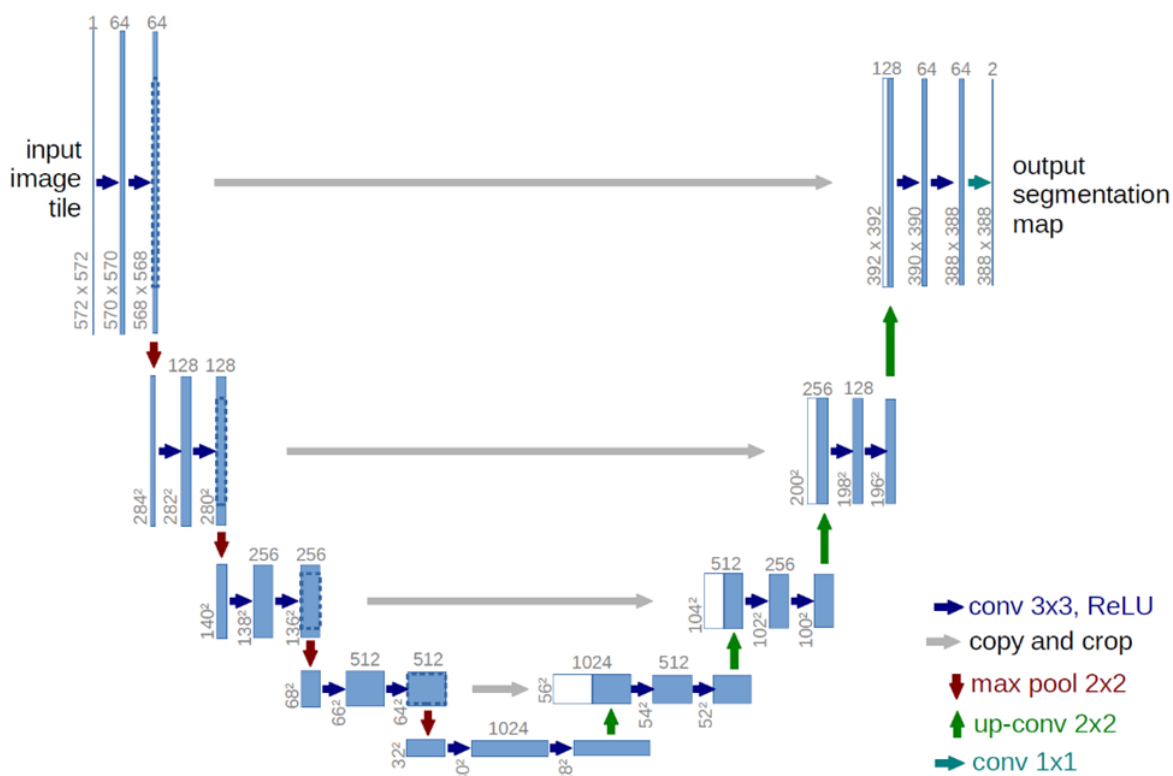


Figure 1. The architecture of the neural network derived from the two-dimensional U-Net employed in this study.

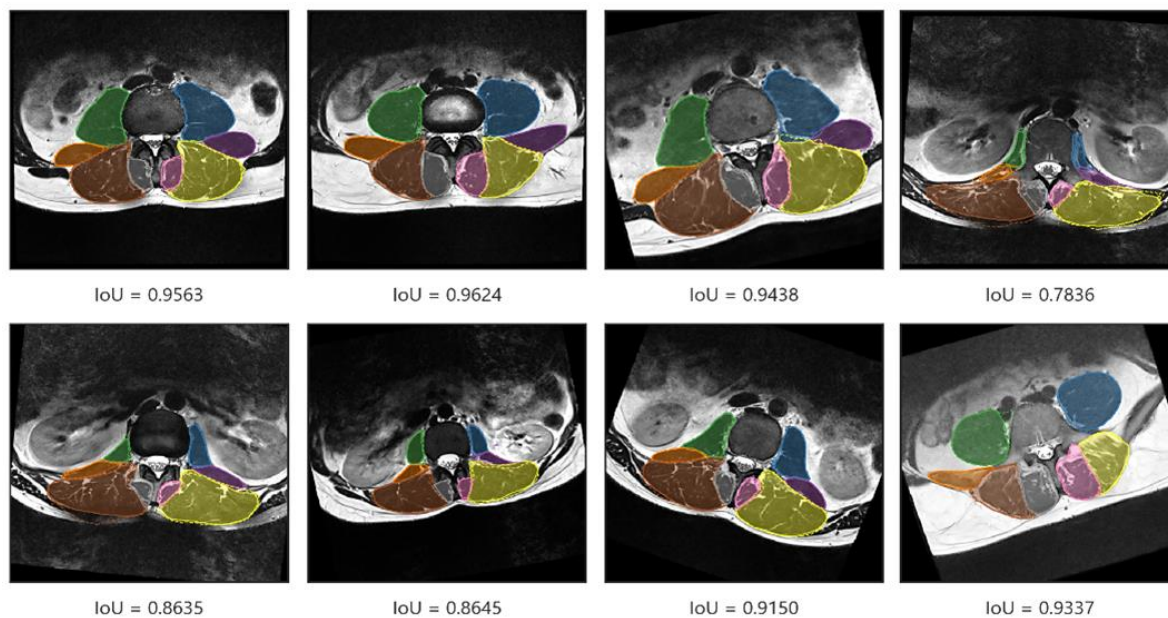


Figure 2. Segmentations of the paravertebral muscles for randomly selected slices of eight representative subjects recruited for the external validation. The average *IoU* for the eight muscles is also shown. Green/blue: psoas major; orange/purple: quadratus lumborum; brown/yellow: erector spinae; gray/pink: multifidus.

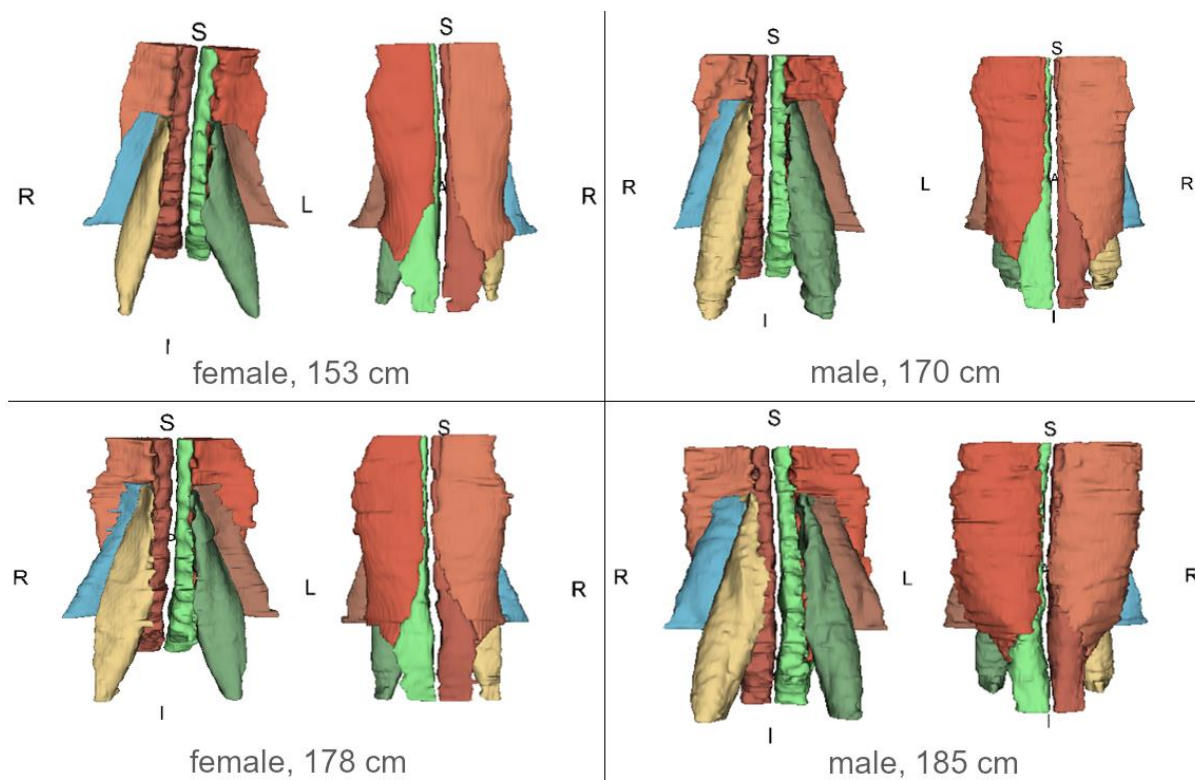


Figure 3. Three-dimensional reconstructions of the segmented paravertebral muscle for four representative volunteers. R: right; L: left; S: superior; I: inferior.

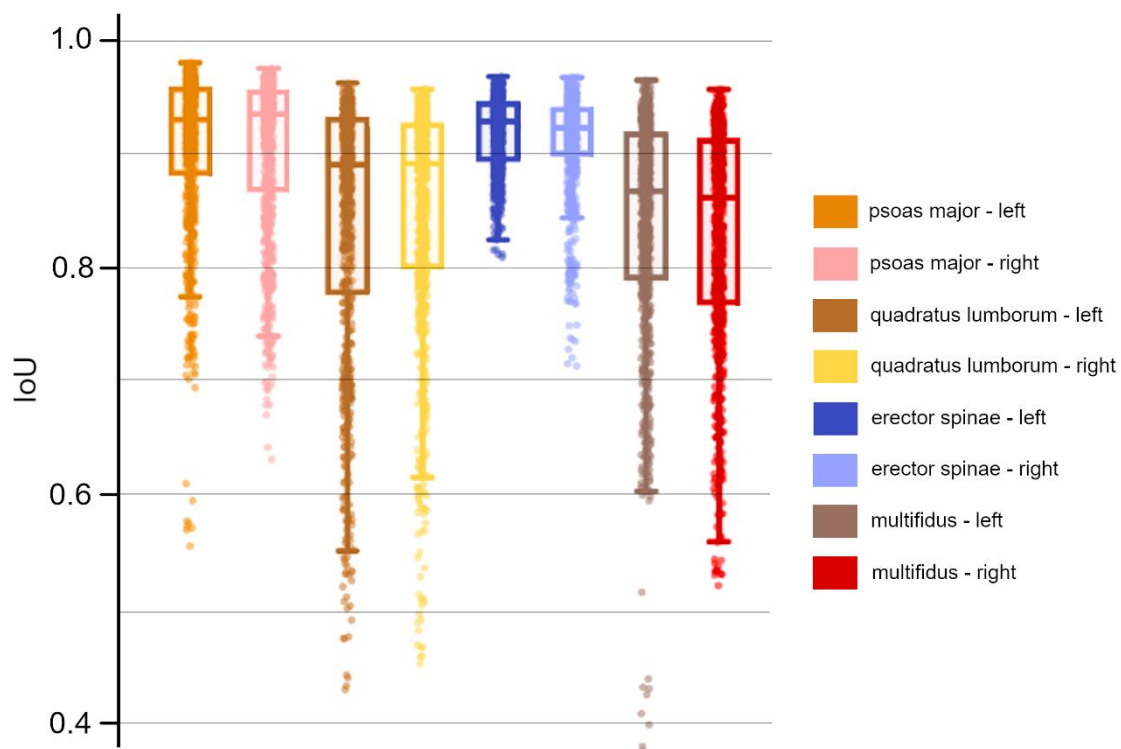


Figure 4. Boxplots and jittered scatter plots of the *IoU* calculated for the individual muscles in all slices acquired for the 22 subjects recruited for the external validation.

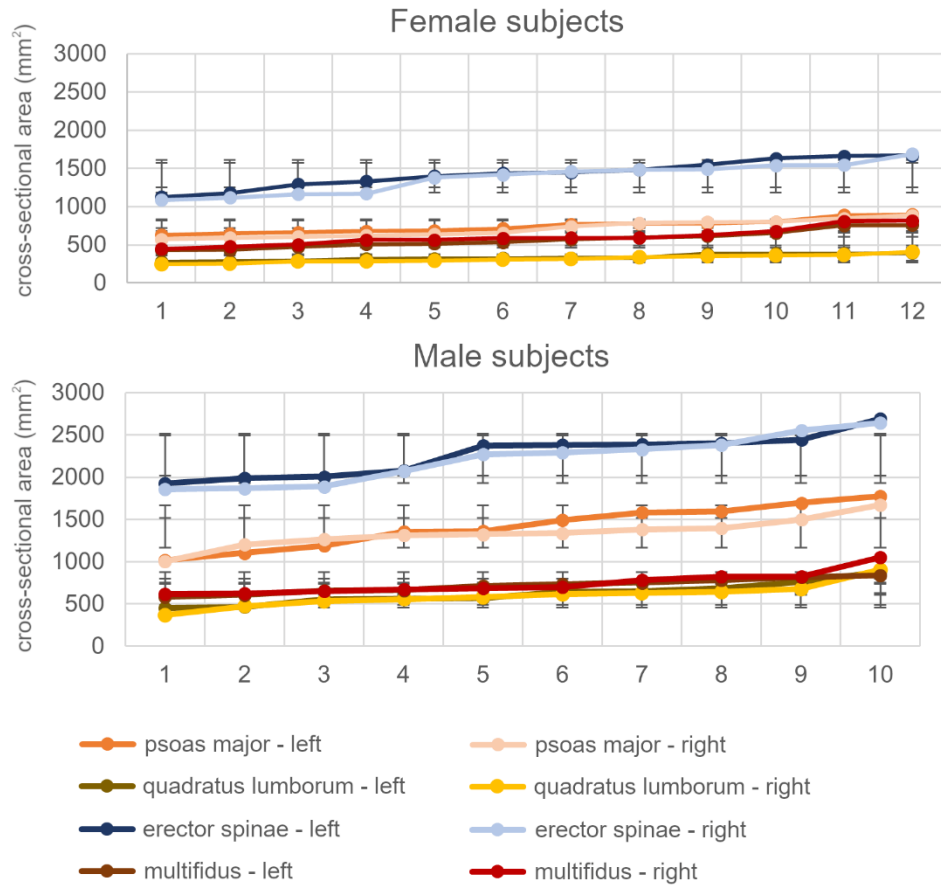


Figure 5. Cross-sectional areas (mean value and standard deviation, in mm²) of the eight paravertebral muscles for the 22 subjects (top: females, bottom: males) in ascending order.

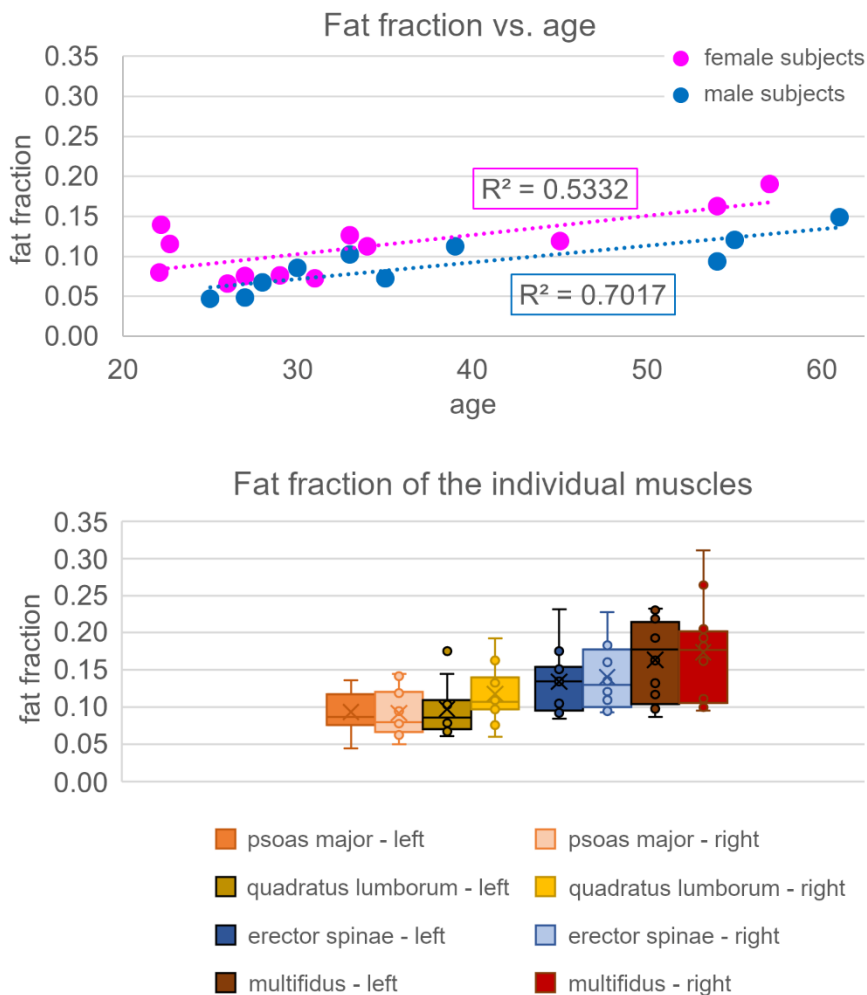


Figure 6. Fat fraction in the paravertebral muscles depending on age and stratified by sex (top) and for the individual muscles (bottom) for the 22 subjects.

Table 1. Median values of the *IoU* for the individual paravertebral muscles calculated for the 22 volunteers recruited for the external validation.

	psoas major		quadratus lumborum		erector spinae		multifidus	
	<i>left</i>	<i>right</i>	<i>left</i>	<i>right</i>	<i>left</i>	<i>right</i>	<i>left</i>	<i>right</i>
<i>IoU</i>	0.930	0.935	0.891	0.892	0.929	0.923	0.868	0.862