Automatic Segmentation Technique for Lumbar Spine Muscle Evaluation from MRI Images

Germán Balerdi[1(B)], Johann Henckel[2], Anna Di Laura[2], Alister Hart[2,3], Martín Belzunce[1,2,4,5]

[1] Centro Universitario de Imágenes Médicas (CEUNIM), Universidad Nacional de Gral, San Martín, Campus Miguelete, 1650 San Martín, Buenos Aires, Argentina gbalerdi22@gmail.com,mbelzunce@unsam.edu.ar

[2] Royal National Orthopaedic Hospital, Stanmore, UK

[3] Institute of Orthopaedics and Musculoskeletal Science, University College London, Stanmore, UK

[4] Center for Complex Systems and Brain Sciences (CEMSC3), Instituto de Ciencias Físicas (ICIFI-CONICET), Escuela de Ciencia y Tecnología, Universidad Nacional de Gral, San Martín, Campus Miguelete, San Martín, Buenos Aires, Argentina

[5] Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina

Abstract

New quantitative muscle quality assessment tools are needed to improve the diagnosis, treatment and study of the lumbar spine muscles. Quantitative magnetic resonance imaging (MRI) has become an important and increasingly relevant technique for diagnosing muscular diseases, tracking their progression, and measuring muscle composition. The Dixon sequence provides fat-only and water-only images, which allows the evaluation of muscle composition and size. Nevertheless, to discriminate a single muscle, a health professional has to manually segment the muscle from the MRI image, which is a slow and impractical process. In this study, we introduce a deep learning-based solution to automatically segment the lower spine muscles from Dixon MRI scans. To achieve that, we trained and validated a U-Net model using 26 manually segmented MRI images of the lower back muscles that was capable of automatically segmenting this muscle group, achieving a mean Dice score of 0.88 in the validation set. This high level of accuracy could allow the

execution of new research looking at the size and composition of this muscle group and may also serve as a valuable tool for enhancing the diagnosis and treatment of lower back issues.

Keywords: MRI · Lumbar Spine · Muscle Health · Deep Learning · U-Net

Introduction

Fat infiltration and atrophy of the lumbar spine muscles is linked to spinal degenerative conditions and can lead to functional impairments in the affected muscles. This infiltration negatively affects muscle strength as muscle fibers are replaced by noncontractile tissue [1, 2]. Since the function of the spinal and lumbar muscles is to provide standing stability and strength, these conditions can derive into everyday illnesses or even disabilities. Consequently, new quantitative muscle quality assessment tools are needed to improve the diagnosis, treatment, and study of the lumbar spine muscles.

Magnetic resonance imaging (MRI) is the best imaging modality to study muscle health thanks to its great soft tissue contrast. To be more specific, the Dixon sequence is the standard for measuring muscle fat fraction, a quantitative metric of muscle composition. This technique involves capturing two distinct images, the first is a conventional spin echo image where the water and fat signals are in-phase, the other is acquired with the water and fat signals out-of-phase [3]. This acquisition allows the creation of water and fat images, and finally a fat fraction image.

Although the fat fraction image provides powerful visual and quantitative information at a voxel level, achieving quantitative assessment of an individual muscle requires the segmentation of the region of interest that delimits the muscle in a 3D image. While this task would be manageable with a single slice (i.e., a 2D image), assessing the entire muscle volume and composition presents a daunting challenge, consuming an impractical amount of time from a skilled professional and, thus, becomes infeasible

[4].

For this reason, the development of new automated tools for the segmentation of the lumbar spine muscles is an unmet need in the study of this muscle group. Semantic segmentation convolutional neural networks (CNN) are a branch of deep learning algorithms that have shown good results in segmenting objects in any kind of image. More specifically, the U-Netnetwork has become the new standard for biomedical image segmentation [5, 6].

The aim of this work is to develop a tool that automatically segments the lumbar spine muscles, namely the psoas (P), iliacus (I), quadratus lumborum (QL), erector spinae (ES) and multifidus (M) from Dixon MR images to obtain quantitative muscle health metrics. The latter two muscles will be considered as a group (ES + M) since both provide motion and stability of the spine, are surrounded by the thoracolumbar fascia and are prone to intramuscular fat infiltration [7–9]. To achieve this, we implemented, trained, and evaluated a U-Net neural network with manually segmented MRI data from a group of amateur cyclists and sedentary subjects.

Materials and Methods

A 3D U-Net neural network was trained to perform automated segmentations of the lumbar spine muscles using MRI images of recreational cyclists and sedentary subjects. The training and validation sets were created by manually segmenting the left and right P, I, QL and ES + M muscles of 26 Dixon MRI images, and then performing data augmentation to obtain a total of 116 images. The performance of the model was measured using the Dice-Sorensen score. In the following sections, the data and methods are fully described.

Data

The Dixon MRI images used in this work belong to a cross-sectional study looking at the muscle health of middle-aged individuals that included sedentary and recreational cyclist groups. The images were acquired in a 3T scanner (Siemens Magneton Vida, Erlangen, Germany) using a body coil. The scanning protocol consisted of an axial TSE Dixon sequence (slice thickness 1.5 mm, spacing between slices 1.95 mm, repetition time (TR) 4570 ms, echo time (TE) 45 ms, number of excitations 1, number of echoes 14, flip angle 120°), with a field of view (FOV) that covered axially from 1 cm below the lesser trochanter to the origin of the psoas muscle at the level of the L1 vertebra. The voxel size was $0.47 \times 0.47 \times 1.95$ mm^3. The 26 MRI scans correspond to 14 women and 12 men with mean(std) age of 45.8(13.6) years and mean(std) BMI of 26.8(6.0).

DataSet

The 26 MRI scans were manually segmented by a trained operator using Simpleware ScanIP software (Version 2021.3; Synopsys, Inc., Mountain View, USA). Figure1shows an example of a manually segmented image.

To homogenize the dataset, every Dixon image and its labels were registered to a reference image using Simple ITK. In this process, a rigid registration with the normalized cross correlation as the similarity metric was used. As a result, all the images had a size of 800 × 640 × 145 voxels and had the same geometrical space.

Furthermore, all images were down sampled to half the size in the trans-axial plane to reduce the memory requirements of the model. The resulting image had 400 × 320 × 145 voxels and a voxel size of 0.94 × 0.94 × 1.95 mm^3, which translated into a four-fold reduction in its memory size with respect to the original images, without a significant loss in image quality.

The registered and downsampled data set of 26 segmented images was split into training and validation sets using a 70/30 ratio. For the 18 images of the training set, data augmentation was implemented using linear transformations that consisted of a two step process. The first consisted of applying a reflection transform in the x axis to each intensity and label images. Then, a clockwise and anticlockwise 5-degree rotation was performed to each of them, making up a dataset with a total of 108 images. In Fig. 2, a flow diagram of the preprocessing and data augmentation process is shown. The validation set consisted of only 8 images.

U-Net

A 3D U-Net that receives the in-phase Dixon image as input and delivers a multi-label mask in its output was implemented. Different from the original architecture, the number of initial filters was reduced from 64 to 16, to reduce the memory requirements of the model. Regarding the loss function and optimizer, we chose the Binary cross entropy, widely used in semantic segmentation applications, and the ADAM optimizer with a learning rate of 10^{-4} respectively. Finally, the model's performance was evaluated using the Sorensen-Dice coefficient.

Training

The 3D U-Net was implemented and trained in Pytorch on a workstation with a 24 GB Nvidia 3090 GPU. Due to video RAM constraints, the training batch size was of a single image. The training took 20 h and completed 200 epochs. The best model was extracted at epoch 178 where the validation loss values started to diverge from training ones.

Results

The proposed method successfully segmented the lumbar spine muscles for most of the subjects. The Dice scores for the automated segmentations of the validation set can be seen in Fig. 3.

The boxplot shows the scores for each muscle, where the median value was 0.92, 0.88, 0.84 and 0.93 for the left P, I, QL and Es + M respectively, and 0.92, 0.89, 0.86 and 0.92 for the right counterpart. Figure 4 shows an accurate segmentation, while Fig. 5 depicts a suboptimal segmentation that corresponds to image number 4 in Table 1.

Table1. Validation Set dice scores for 3D Unet Model

Image	LP	LI	LQL	LES + M	RP	RI	RQL	RES + M
1	0.942	0.892	0.802	0.915	0.919	0.869	0.856	0.919
2	0.921	0.859	0.825	0.926	0.916	0.869	0.855	0.922
3	0.923	0.92	0.844	0.93	0.937	0.913	0.88	0.936
4	0.92	0.84	0.806	0.832	0.881	0.904	0.377	0.766
5	0.905	0.88	0.849	0.927	0.891	0.898	0.83	0.938
6	0.941	0.866	0.836	0.927	0.934	0.868	0.896	0.925
7	0.947	0.903	0.876	0.926	0.937	0.898	0.891	0.911
8	0.92	0.878	0.865	0.927	0.902	0.882	0.822	0.931
Median (iqr)	0.921 (.021)	0.878 (.031)	0.844 (.033)	0.927 (.004)	0.916 (.036)	0.898 (.031)	0.855 (.055)	0.925 (.015)
Mean (std)	0.925 (.014)	0.878 (.025)	0.843 (.026)	0.914 (.033)	0.914 (.022)	0.890 (.018)	0.793 (.173)	0.904 (.057)

1 Discussion

An automated segmentation tool for left and right lumbar spine muscles that uses a U-Net architecture neural network model is presented in this work. The segmentation performance was measured with the Dice-Sorensen coefficient and averaged 0.88 in the validation set for the four muscles. This value is substantially affected by one of the 8 validation images that has shown poor results, particularly for the QL muscle (Image number4ofTable1). This is a small muscle, and therefore its Dice score is more sensitive

Fig. 5. 3D model (right) and axial slice (left) at the height of the iliac crest of an inaccurate segmentation. The shading represents the manual segmentation while the outline is the model's prediction. It can be seen in the axial slice that the model could not segment the right QL.

to segmentation inaccuracies. To illustrate this issue, the mean Dice scores were 0.84 and 0.80 for QL (the smallest muscle in this group), while the best segmentation performance was 0.93 for the ES + M group.

Every deep learning model performance is tied to the size of the training set, which in this case is considerably low because there are no public datasets of manually segmented images of the lumbar spine muscles. In addition, the GPU memory limitations restricted the number of filters used in the U-Net architecture, consequently reducing the number of parameters. These two factors could have affected the performance of our model.

Despite the reduced amount of manually segmented data available, the model has shown excellent results in larger muscles like the Psoas and the Erector Spinae and Multifidus group. As expected, the performance was not as good when the muscles were small.

2 Conclusion

In this work, we trained a volumetric semantic segmentation model based on the U-Net architecture capable of automatically segmenting the psoas, iliacus, quadratus lumborum, and the joined erector spinae and multifidus muscles. The good accuracy of the proposed method could allow the execution of new research studying the size and composition of this muscle group and may also serve as a valuable tool for enhancing the diagnosis and treatment of lower back issues.

References

1. Hamrick, M.W., McGee-Lawrence, M.E., Frechette, D.M.: Fatty infiltration of skeletal muscle: mechanisms and comparisons with bone marrow adiposity. Front Endocrinol (Lausanne). **7**, 69 (2016). Epub 2016/06/20. https://doi.org/10.3389/fendo.2016.00069 PMID: 27379021; PubMed Central PMCID:PMCPMC4913107
2. Menezes-Reis, R., Bonugli, G.P., Salmon, C.E.G., Mazoroski, D., Herrero, C., et al.: Relationship of spinal alignment with muscular volume and fat infiltration of lumbar trunk muscles. PLoS ONE **13**(7), e0200198 (2018). https://doi.org/10.1371/journal.pone.0200198
3. Dixon, W.T.: Simple proton spectroscopic imaging. Radiology **153**, 189–194 (1984). https://doi.org/10.1148/radiology.153.1.6089263
4. Belzunce, M.A., Henckel, J., Fotiadou, A., Di Laura, A., Hart, A.: Automated multi-atlassegmentation of gluteus maximus from Dixon and T1-weighted magnetic resonance images.
Magma(NewYork,N.Y.)**33**(5),677–688(2020).https://doi.org/10.1007/s10334-020-00839-3
5. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical imagesegmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image ComputingandComputer-AssistedIntervention–MICCAI2015.MICCAI2015.LNCS,vol.9351. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
6. Zhihui, G., et al.: Fully automated 3D segmentation of MR-imaged calf muscle compartments: Neighborhood

relationship enhanced fully convolutional network. Comput. Med. Imaging Graph. **87**, 101835 (2021). ISSN 0895-6111, https://doi.org/10.1016/j.compmedimag.2020.101835

7. Lee, H.J., et al.: The relationship between cross sectional area and strength of back muscles inpatients with chronic low back pain. Ann. Rehabil. Med. **36**, 173–181 (2012). https://doi.org/ 10.5535/arm.2012.36.2.173
8. Pezolato, A., de Vasconcelos, E.E., Defino, H.L.A., et al.: Fat infiltration in the lumbar multifidus and erector spinae muscles in subjects with sway-back posture. Eur. Spine J. **21**, 2158–2164 (2012). https://doi.org/10.1007/s00586-012-2286-z
9. Willard, F.H., Vleeming, A., Schuenke, M.D., Danneels, L., Schleip, R.: The thoracolumbarfascia: anatomy, function and clinical considerations. J. Anat. **221**, 507–536 (2012). https:// doi.org/10.1111/j.1469-7580.2012.01511.x

Figure legends

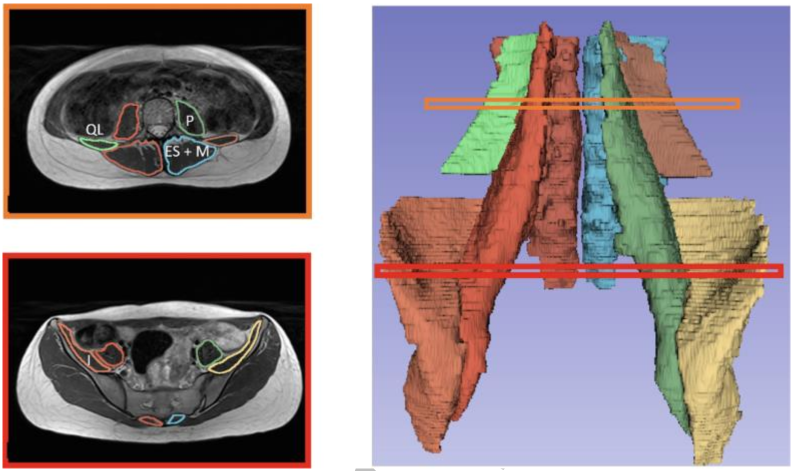
Fig. 1. On the left, segmented slices at the level of the L2 vertebra (top) and the iliac crest (bottom) are shown. On the right, a 3D render of the manual segmentation is presented.

Fig. 2. Schematic of the data augmentation process. First, the 26 manually segmented data are registered to a reference image. The registered images are down sampled four-fold in the transaxial images and then flipped creating 52 different images. Finally, anti-andclockwise5degreerotations were applied resulting in a total of 156 augmented images and labels.

Fig. 3. Validation Dice score for the segmentation of every muscle.

Fig. 4. 3D model(right) and axial slice (left) at the height of the iliac crest of an accurate segmentation. The shading represents the manual segmentation while the outline is the model's prediction.

Figures


QL
P
ES + M
I

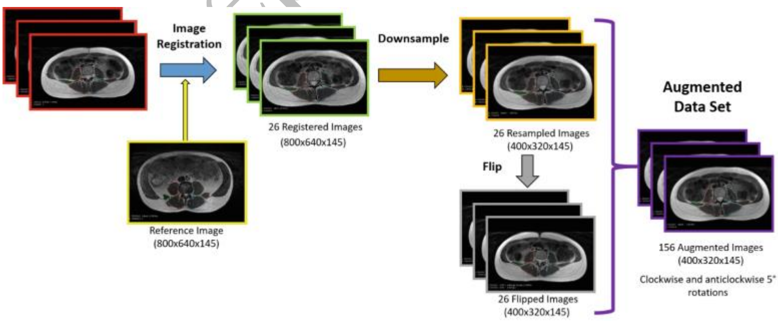
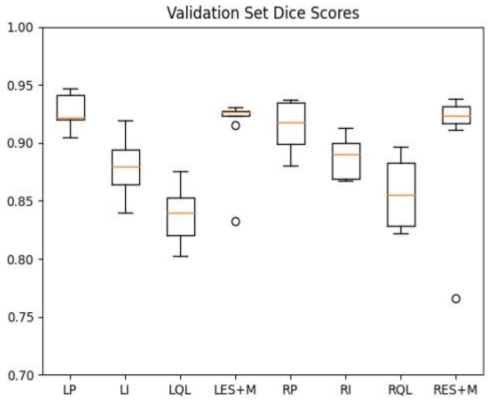


Image
Registration
Downsample
26 Registered Images
(800x640x145)
Reference Image
(800x640x145)
26 Resampled Images
(400x320x145)
Flip
26 Flipped Images
(400x320x145)
Augmented
Data Set
156 Augmented Images
(400x320x145)
Clockwise and anticlockwise 5°
rotations


Validation Set Dice Scores
1.00
0.95
0.90
0.85
0.80
0.75
0.70
LP
LI
LQL
LES+M
RP
RI
RQL
RES+M