

Development of a deep learning-based fully automated segmentation of rotator cuff muscles from clinical MR scans

Acta Radiologica
2024, Vol. 65(9) 1126–1132
© The Foundation Acta Radiologica
2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02841851241262325
journals.sagepub.com/home/acr



Sae Hoon Kim¹ , Hye Jin Yoo^{2,3} , Soon Ho Yoon^{2,3,4},
Yong Tae Kim⁵, Sang Joon Park^{2,3,4}, Jee Won Chai⁶, Jiseon Oh²
and Hee Dong Chae^{2,3}

Abstract

Background: The fatty infiltration and atrophy in the muscle after a rotator cuff (RC) tear are important in surgical decision-making and are linked to poor clinical outcomes after rotator cuff repair. An accurate and reliable quantitative method should be developed to assess the entire RC muscles.

Purpose: To develop a fully automated approach based on a deep neural network to segment RC muscles from clinical magnetic resonance imaging (MRI) scans.

Material and Methods: In total, 94 shoulder MRI scans (mean age = 62.3 years) were utilized for the training and internal validation datasets, while an additional 20 MRI scans (mean age = 62.6 years) were collected from another institution for external validation. An orthopedic surgeon and a radiologist manually segmented muscles and bones as reference masks. Segmentation performance was evaluated using the Dice score, sensitivities, precision, and percent difference in muscle volume (%). In addition, the segmentation performance was assessed based on sex, age, and the presence of a RC tendon tear.

Results: The average Dice score, sensitivities, precision, and percentage difference in muscle volume of the developed algorithm were 0.920, 0.933, 0.912, and 4.58%, respectively, in external validation. There was no difference in the prediction of shoulder muscles, with the exception of teres minor, where significant prediction errors were observed (0.831, 0.854, 0.835, and 10.88%, respectively). The segmentation performance of the algorithm was generally unaffected by age, sex, and the presence of RC tears.

Conclusion: We developed a fully automated deep neural network for RC muscle and bone segmentation with excellent performance from clinical MRI scans.

Keywords

Shoulder, rotator cuff, segmentation, artificial intelligence, magnetic resonance imaging

Date received: 15 June 2023; accepted: 21 May 2024

Introduction

Rotator cuff tendon tear is a common degenerative disease and increases with age. It is known that 20% of the general population experiences rotator cuff tears, especially supraspinatus tendon tears (1). The fatty infiltration and atrophy in the muscle after a rotator cuff tear are important in surgical decision-making and are linked to poor clinical outcomes after rotator cuff repair (2–5). Therefore, preoperative evaluation of muscle atrophy and fatty infiltration is essential.

In routine practice, the rotator cuff muscles are commonly evaluated using Thomazeau's occupation ratio (6), or Zanetti's tangent sign (7) for supraspinatus muscle atrophy, and Goutallier classification for fatty infiltration of rotator cuff muscles (8). The Goutallier classification

¹Department of Orthopaedic Surgery, Seoul National University Hospital, Seoul, Republic of Korea

²Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea

³Department of Radiology, Seoul National University College of Medicine, Seoul, Republic of Korea

⁴MEDICALIP Co. Ltd., Seoul, Republic of Korea

⁵Department of Orthopaedic Surgery, Hallym University Dongtan Sacred Heart Hospital, Gyeonggi, Republic of Korea

⁶Department of Radiology, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

Corresponding author:

Hye Jin Yoo, Department of Radiology, Seoul National University Hospital, 101 Daehangno, Jongno-gu, Seoul, 03080, Republic of Korea.

Email: dalnara3@snu.ac.kr

has the advantage of being intuitive and easy to apply, but there is high variation among studies reporting inter-observer reliability of this scale because it could be influenced by subjective factors (9–11). In addition, the Goutallier classification has a limitation of being determined in a single slice image, which cannot reflect overall muscle quality sufficiently. A more accurate and reliable quantitative method should be developed to assess the entire rotator cuff muscles. Image segmentation is a process where specific pixels in an image are allocated into regions, highlighting pixels of interest. Manual or semi-automated segmentation strategies have been reported but their clinical usage was limited because they can be time-consuming and labor-intensive.

Recently, several studies have been conducted to apply fully automatic segmentation based on artificial intelligence techniques in medical image segmentation to reduce the analysis time (12–17). Medina et al. developed a convolutional neural network (CNN) to segment rotator cuff muscles at Y-view on T1-weighted shoulder magnetic resonance imaging (MRI) (13). Taghizadeh et al. and Riem et al. also developed an automated artificial intelligence (AI) algorithm for the quantification of muscle volume and fatty infiltration of rotator cuff muscles using shoulder computed tomography (CT) or MR images (14,17).

The aim of the present study was to develop a fully automated approach based on a deep neural network to segment rotator cuff muscles from clinical MRI scans, which could be utilized in future studies to quantify muscle volume and intramuscular fat content.

Material and Methods

The institutional review board of the participating institutions approved this retrospective study for model development, and the requirement for informed patient consent was waived (IRB no. H-2209-160-1364).

Study population

For the original training dataset, 100 shoulder MRI scans were retrospectively collected from 100 patients at a single tertiary center for the purpose of workup of shoulder pain. Among them, six patients were excluded due to poor image quality. Finally, 94 exams from 94 patients were included (27 men, 67 women; 66 right shoulders, 28 left shoulders; mean age = 62.3 years; age range = 46–76 years). We separately collected 20 shoulder MRI scans from another institution for external validation to evaluate segmentation performance between the reference masks and the network-derived masks. In the external validation set, eight men and 12 women were included (13 right shoulders, 7 left shoulders; mean age = 62.6 years; age range = 45–80 years).

Imaging protocol

All shoulder MRI examinations were performed on 3.0-T MRI scanners (Skyra; Siemens Medical Solutions, Erlangen, Germany) with dedicated shoulder coils. The patients were placed in the supine position with their forearms in the neutral position. The MRI protocol included the following sequences: axial fast spin echo proton-density weighted image with fat saturation (FSE-PDWI; TR/TE = 3500–4000/30–40 ms, slice thickness = 3 mm, slice gap = 0 mm, field of view [FOV] = 140 × 140 mm, matrix = 384 × 307 mm, and echo train length = 7); oblique coronal and sagittal fast spin echo (FSE) T2-weighted (T2W) images with fat saturation (TR/TE = 3300–4600/40–60 ms, slice thickness = 3 mm, slice gap = 0 mm, FOV = 140 × 140 mm, matrix = 384 × 307, and echo train length = 7); and oblique coronal FSE T1-weighted (T1W) images (TR/TE = 400–800/10–15 ms, slice thickness = 3 mm, slice gap = 0 mm, FOV = 140 × 140 mm, matrix = 384 × 346, and echo train length = 3). Extended sagittal T1W images were taken with a larger FOV (160 × 160 mm) covering the medial border of the scapula body to obtain a farther medial section of rotator cuff muscles.

All shoulder MRI scans from another institution for external validation included oblique coronal and sagittal FSE T2W images (TR/TE = 3690–4700/70–80 ms, slice thickness = 3 mm, slice gap = 0 mm) and extended sagittal T1W images (TR/TE = 500–652/12–19 ms, slice thickness = 3 mm, slice gap = 0 mm). Extended sagittal T1W imaging was used to evaluate muscle segmentation performance and FSE, while FSE T2W images were applied to evaluate the status of rotator cuff tendons.

Manual segmentation

Extended sagittal T1W images were manually segmented by a single author with 10 years of clinical experience into five classes as follows: (i) supraspinatus; (ii) subscapularis; (iii) infraspinatus; (iv) teres minor; (v) deltoid muscles; (vi) humerus; and (vii) scapula. Muscle was segmented by tracing the muscle border with a freehand region of interest (ROI) from the medial scapular border to the most lateral aspect of the tendon that could still be differentiated from the other rotator cuff tendons (16). In cases where a muscle exhibited significant fatty replacement, the fascial boundaries were outlined instead of delineating each preserved fiber. All segmented images were independently reviewed by a second author with 17 years of experience in musculoskeletal radiology. These served as reference ground truth labels as the benchmark against which our model's predictions were compared during model training and evaluation. All MR data in development and validation sets were processed using a commercial software program (MEDIP PRO v2.0.0; MEDICAL IP Co., Ltd., Seoul, Republic of Korea).

Training and validation datasets

Our training dataset consisted of a collection of input MRI scans, each paired with corresponding specific ROIs of the shoulder muscles. We also utilized a separate validation dataset, which was not used during the training phase, to ensure an unbiased evaluation of the model's performance. The 2820 MR slices with corresponding ground truth labels from 94 patients were split into training (2520 slices from 84 exams) and internal validation (300 slices from 10 exams) sets. For external validation, 560 slices from 20 exams were used to output predictions in five classes that were compared with manual segmentations using Dice score.

Model development

To segment the shoulder muscle area from the MRI scans, a three-dimensional (3D) nnU-Net architecture was selected. The U-Net network is a widely recognized fundamental model for medical image segmentation, and the nnU-Net has showed an excellent performance in various segmentation tasks by automatically configuring preprocessing methods and network hyper-parameters (normalization, input patch size, target pixel spacing) instead of complicated manual tuning according to the dataset. This automation alleviates the need for manual adjustments based on the dataset specifics, thereby enhancing model adaptability and performance.

Given the MRI nature of our training dataset, we applied the z-score normalization to preprocess input images. This technique involves subtracting the mean and dividing by standard deviation, thus standardizing the input data and facilitating the learning process by ensuring consistent data distribution.

The input patch size of the network is $16 \times 320 \times 320$. Target pixel spacing is calculated by calculating the median spacing for each axis from training images, resizing all images to conform to this median spacing. The target spacing for nnU-Net in this dataset is 5 mm, 0.3125 mm, and 0.3125 mm, respectively.

The optimization algorithm for the network is SGD (Stochastic Gradient Descent) with a nesterov momentum value of 0.99. The combination of dice loss and cross-entropy loss was chosen as the loss function, leveraging the strengths of both to guide the training process effectively.

The polynomial learning rate was set at an initial value of 0.01 and a decay scheduler with a power of 0.9 was applied during the 1000 epochs of model training. The learning rate corresponds to a coefficient to determine the multiplier of each weight correction (speed of training). The polynomial learning rate indicates a learning rate that is not constant throughout the training (the learning rate changes according to certain formula). The decay scheduler

means scheduling the reduction (decay) of learning rate throughout the training. The epoch indicates the number of training iterations. Our approach adjusts the learning rate dynamically across training iterations, unlike a constant rate. More specifically, our model's learning rate decreases following a predetermined formula. This strategy, known as a decay scheduler, mitigates the risk of overshooting the optimal model parameters by gradually reducing the learning rate, thereby enhancing the precision of the model's convergence.

The concept of learning rate is fundamental, acting as the coefficient that determines the magnitude of weight adjustments during training. Setting this rate too high risks overshooting the optimal solution, while too low a rate can lead to prolonged training or convergence to suboptimal solutions. Therefore, the polynomial learning rate and decay scheduler are crucial for balancing training speed and accuracy.

Ground truth refers to the manually segmented images that serve as a benchmark for accuracy, representing the ideal segmentation outcome. Automatically segmented masks, on the other hand, are the result of the model's segmentation attempts. Comparing these against the ground truth enables us to evaluate the model's performance and make necessary adjustments to achieve the highest possible accuracy.

Statistical analysis

To evaluate the performance of the automated segmentation, masks generated by the developed model were compared against reference segmentation masks on the external validation set. We calculated the Dice score, sensitivity, precision, and percent difference in muscle volume by comparing the ground truth (G) and automatically segmented (S) masks. The Dice score is defined as the degree of overlap between two different masks and is calculated using the following formula:

Dice score = $2 \times \text{the number of pixels overlapped between both the ground truth and the model prediction} / \text{the total number of pixels in the ground truth and the model prediction}$

A higher Dice score indicates a greater degree of overlap and, consequently, higher accuracy.

Percent differences in muscle volume (V) were calculated by subtracting the volume of the ground truth mask (V_G) from the volume of the automatically segmented mask (V_S) as follows:

$$V(\%) = (V_G - V_S) / V_G \times 100$$

Precision was defined as the number of correct pixels over the total number of pixels in the body compartment. In addition, we evaluated segmentation performance according to sex, age, and the presence of rotator cuff tendon tear. We

used the Mann–Whitney U test to compare segmentation performance depending on sex, age (<62 years vs. ≥ 62 years) and the presence of rotator cuff tendon tear. Bland–Altman analysis was used to test for proportional bias. $P < 0.05$ was considered indicative of a statistically significant difference. Statistical analyses were performed with SPSS (version 22; IBM Corp., Armonk, NY, USA).

Results

The Dice scores, sensitivity, and precision of the developed algorithm are summarized in Table 1. Overall, mean muscle segmentation Dice scores for internal and external test sets were >0.92 , sensitivities were >0.91 , and precisions were >0.91 . Segmentation yielded low volume errors (mean = 4.58%). There was no difference in the prediction of shoulder muscles (range = 1.15%–3.52%) except for teres minor (10.88%) (Table 2). Significant prediction errors were seen in teres minor muscle segmentation ($10.88\% \pm 4.86\%$), especially along the superior border of the teres minor, where it was challenging to identify the interface between infraspinatus and teres minor muscles (Fig. 1).

The segmentation performance of the algorithm was unaffected by age, and sex (Table 3). Despite the presence of tendon tears, there was no difference in predicting the shoulder muscles, except for precision and volume error (%). The precision and volume error (%) for the prediction of muscles with full-thickness tears were significantly different compared to those for muscles without tears: precision = 0.933 ± 0.016 vs. 0.910 ± 0.072 , $P = 0.036$; volume error = $1.25\% \pm 1.38\%$ vs. $5.21 \pm 16.6\%$, $P = 0.005$. However, the difference disappeared when the supraspinatus and infraspinatus muscles were included in the analysis: precision = 0.933 ± 0.016 vs. 0.936 ± 0.020 , $P = 0.649$; volume error = $1.25\% \pm 1.38\%$ vs. $2.42\% \pm 5.54\%$, $P = 0.298$. Since there were no cases of full-thickness tears in the teres minor muscles, the inclusion of teres minor, which had significantly lower segmentation performance, could have adversely affected the overall performance of the cases without tears. Upon analyzing the prediction of two muscles, i.e. supraspinatus, and infraspinatus muscles, we observed a slight increase in overall performance as expected (from 0.912 ± 0.065 to 0.935 ± 0.018). Furthermore, the presence of the rotator cuff tendon tears did not appear to

have any influence on the segmentation performance of these muscles. Bland–Altman analysis demonstrated no proportional bias for all muscles ($P < 0.05$) (Fig. 2.)

Discussion

Traditional methods for evaluating rotator cuff muscle degeneration, such as Thomazeau's muscle occupation ratio and Goutallier's semi-quantitative grading system, have limited applicability due to their restricted use on the supraspinatus muscle and inconsistent reliability with high inter- and intra-rater variability (9,10). In contrast, deep learning and CNN techniques have emerged as more recent approaches for automatically quantifying rotator cuff muscles.

In this study, we developed and validated a fully automated segmentation algorithm based on a deep neural network to segment rotator cuff muscles. Our technique enables fast automatic volumetric segmentation of entire rotator cuff muscles with good performance from clinical MRI scans.

Our technique generally provides similar or improved segmentation results compared to previous deep-learning methods for the same muscle (13,14,17). For instance, Medina et al. developed CNN to segment rotator cuff muscles at Y-view on T1W shoulder MRI (13). They reported Dice similarity coefficients of 0.964, 0.951, and 0.933 for the supraspinatus, infraspinatus, and subscapularis muscles, respectively. However, they did not individually segment the teres minor muscle due to poorly defined boundaries with the infraspinatus muscle. Taghizadeh et al. developed a CNN to automatically quantify and assess muscle atrophy and fatty infiltration of rotator cuff

Table 1. Segmentation performance in internal and external validation.

	Dice	Sensitivity	Precision
Internal validation	0.927 ± 0.014	0.911 ± 0.026	0.945 ± 0.030
External validation	0.920 ± 0.018	0.933 ± 0.019	0.912 ± 0.017

Values are given as mean \pm SD.

Table 2. Segmentation performance in external validation datasets.

	Dice	Sensitivity	Precision	Volume error (%)
Supraspinatus	0.941 ± 0.015	0.944 ± 0.019	0.938 ± 0.020	1.15 ± 2.52
Infraspinatus	0.928 ± 0.017	0.923 ± 0.036	0.933 ± 0.019	3.35 ± 6.74
Subscapularis	0.993 ± 0.011	0.938 ± 0.015	0.929 ± 0.017	1.47 ± 2.25
Teres minor	$0.831^* \pm 0.108$	$0.854^* \pm 0.124$	$0.835^* \pm 0.131$	$10.88^* \pm 4.86$
Deltoid	0.929 ± 0.019	0.940 ± 0.028	0.919 ± 0.019	3.52 ± 2.17
Humerus	0.960 ± 0.018	0.992 ± 0.008	0.931 ± 0.034	6.65 ± 4.15
Scapula	0.962 ± 0.019	0.985 ± 0.012	0.940 ± 0.030	5.04 ± 2.98

Values are given as mean \pm SD.

*Significant difference ($P < 0.05$) compared to other muscles.

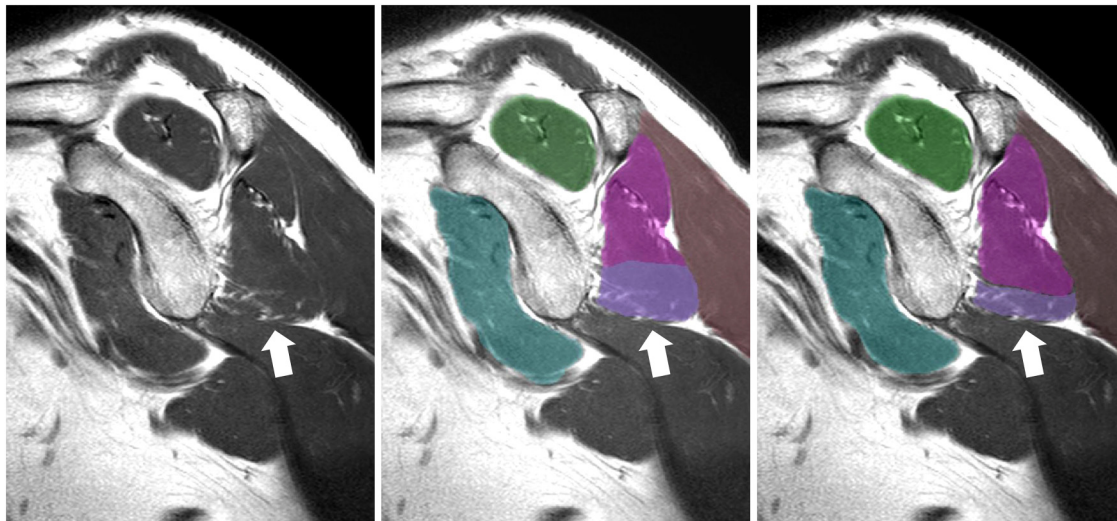


Fig. 1. A 54-year-old man with right shoulder pain. (a) A sagittal T1-weighted image showed an isolated teres minor muscle (arrow). (b) A developed deep neural network algorithm performed auto-segmentation of rotator cuff muscles. The largest error in the segmentation occurred in the teres minor muscle (arrow). (c) After correction, the segmentation of the teres minor muscle was refined (arrow), resulting in an accurate match to its actual size.

Table 3. Subgroup analysis of segmentation performance according to sex, age, and tendon tears.

		Dice	Sensitivity	Precision	Volume error (%)
Sex	Female (n = 12)	0.928 ± 0.041	0.938 ± 0.087	0.918 ± 0.054	2.48 ± 10.2
	Male (n = 8)	0.925 ± 0.078	0.941 ± 0.044	0.919 ± 0.074	3.82 ± 22.6
	P value	0.192	0.351	0.336	0.172
Age	<62 (n = 10)	0.928 ± 0.070	0.941 ± 0.078	0.922 ± 0.067	3.51 ± 20.2
	≥62 (n = 10)	0.926 ± 0.044	0.939 ± 0.047	0.915 ± 0.057	2.52 ± 11.1
	P value	0.780	0.864	0.522	0.464
Tendon tear (all muscles)*	Absent (n = 94)	0.912 ± 0.065	0.920 ± 0.068	0.910 ± 0.072	5.21 ± 16.6
	Presence (n = 6)	0.929 ± 0.013	0.927 ± 0.024	0.933 ± 0.016	1.25 ± 1.38
	P value	0.056	0.595	0.036 [‡]	0.005 [‡]
Tendon tear (SS, IS) [†]	Absent (n = 34)	0.935 ± 0.018	0.935 ± 0.031	0.936 ± 0.020	2.42 ± 5.54
	Presence (n = 6)	0.929 ± 0.013	0.927 ± 0.024	0.933 ± 0.016	1.25 ± 1.38
	P value	0.289	0.325	0.649	0.298

Values are given as mean ± SD.

*Analysis was performed for all rotator cuff muscles individually.

[†]Analysis was performed for only supraspinatus and infraspinatus muscles.

[‡]Shows a statistically significant difference ($P < 0.05$) between two groups.

muscles using shoulder CT images (14). Their automatic muscle segmentation yielded a mean Dice similarity coefficient of 0.88 ± 0.09 . Within the segmented muscles, the teres minor muscle exhibited the lowest mean Dice coefficient (0.86 ± 0.10), which significantly influenced the overall moderate accuracy of their CNN model. Similarly, Riem et al. also developed and validated an automated AI algorithm for 3D segmentation of all four rotator cuff muscles to quantify intramuscular fat infiltration and individual muscle volume (17). In their study, the AI model demonstrated excellent performance with high Dice scores (mean = 0.92 ± 0.14), and minimal errors in muscle volume and fat infiltration. However, they

encountered challenges in accurately distinguishing the border between the infraspinatus and teres minor muscles, resulting in the largest errors in those areas. In the present study, we further improved the segmentation performance with a fully automated segmentation algorithm, even in patients with rotator cuff tendon tears. However, similar to previous studies, we faced difficulties in precisely identifying the border between the infraspinatus and teres minor muscles, leading to the largest errors in the segmentation of teres minor muscles. In particular, there were large errors in the case of isolated teres minor muscle atrophy because it was so challenging to delineate the teres minor muscle when there was severe atrophy or complete fatty

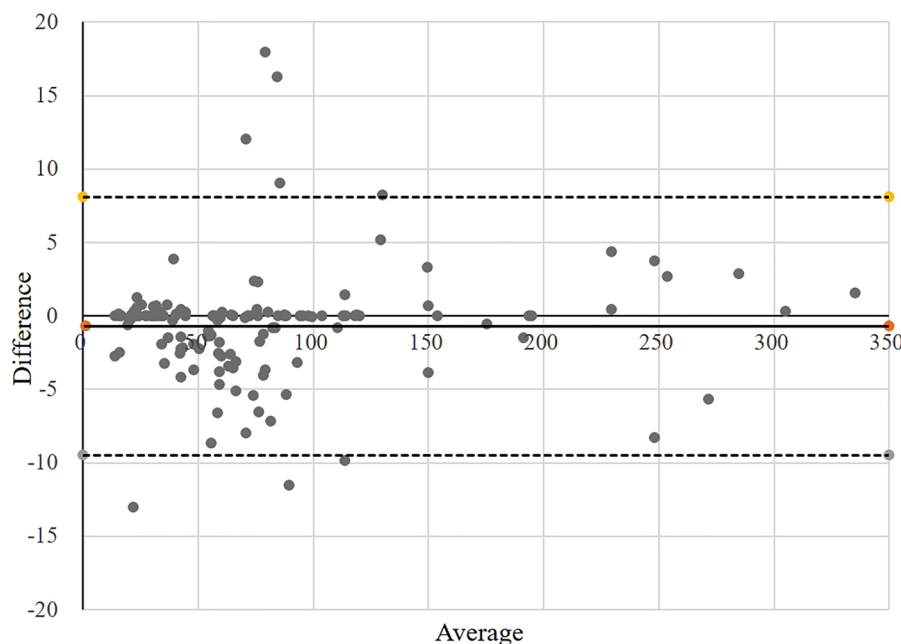


Fig. 2. Bland–Altman plots showed no proportional bias in all muscles.

replacement of teres minor muscle. This study also employed automated segmentation and automatic volume calculation techniques for both the humerus and scapula. In future studies, the volumes of the humerus and scapula can serve as an internal reference, allowing for the adjustment of rotator cuff muscle volumes based on variations in body size (18).

The present study has some limitations. First, the external validation dataset was limited in size, comprising a small number of pathologic shoulders with rotator cuff tendon tears, cases with severe muscle atrophy, and fatty infiltration. Furthermore, the dataset had an insufficient number of cases with glenohumeral joint effusion or subacromial-subdeltoid bursitis. On T1W images, fluid is hard to differentiate from muscle, potentially leading to segmentation failure of the CNN model. In the future, a more extensive dataset will be required to comprehensively evaluate the overall applicability of the developed algorithm. Second, we did not conduct inter- and intra-observer reliability analysis for manual segmentation as a result of the large number of cases and the significant time required to manually segment the entire rotator cuff muscles. Third, due to the ambiguous interface between the teres minor and infraspinatus muscles, it is inevitable that substantial errors occur during the manual segmentation process. As a result, it is not possible to consider it as an absolute ground truth. Finally, we initially considered combining the infraspinatus and teres minor as one muscle group due to the aforementioned ambiguous interface between the two muscles. Segmentation of this double muscle could significantly improve the overall performance of our developed CNN model. However, the teres minor muscle often exhibits isolated muscle atrophy without tendon

tear. In addition, compensatory hypertrophy of the teres minor muscle could occur in cases with severe infraspinatus muscle atrophy due to a complete tear of the infraspinatus tendon. If these two muscles are considered as one group, the accurate evaluation of infraspinatus muscle atrophy could be compromised in such situations. Therefore, we decided to perform the segmentation of these muscles separately. Nevertheless, it is possible to develop an additional algorithm for the segmentation of those two muscles as one muscle group with two tendons.

In conclusion, we developed a fully automated deep neural network for rotator cuff muscle segmentation with excellent performance from clinical MRI scans. The developed algorithm has the potential to facilitate various quantitative studies assessing the impact of rotator cuff muscles in the treatment of patients with rotator cuff tears.



Declaration of conflicting interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: SHY is the Chief Medical Officer of MEDICAL IP, Co. Ltd.; SJP is the CEO of MEDICAL IP, Co. Ltd.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sae Hoon Kim  <https://orcid.org/0000-0002-6848-350X>
Hye Jin Yoo  <https://orcid.org/0000-0002-9704-7870>

References

1. Minagawa H, Yamamoto N, Abe H, et al. Prevalence of symptomatic and asymptomatic rotator cuff tears in the general population: from mass-screening in one village. *J Orthop* 2013;10:8–12.
2. Jost B, Pfirrmann CW, Gerber C, et al. Clinical outcome after structural failure of rotator cuff repairs. *J Bone Joint Surg Am* 2000;82:304–314.
3. Kijowski R, Thurlow P, Blankenbaker D, et al. Preoperative MRI shoulder findings associated with clinical outcome 1 year after rotator cuff repair. *Radiology* 2019;291:722–729.
4. Mellado JM, Calmet J, Olona M, et al. Surgically repaired massive rotator cuff tears: MRI of tendon integrity, muscle fatty degeneration, and muscle atrophy correlated with intraoperative and clinical findings. *AJR Am J Roentgenol* 2005;184:1456–1463.
5. Thomazeau H, Boukobza E, Morcet N, et al. Prediction of rotator cuff repair results by magnetic resonance imaging. *Clin Orthop Relat Res* 1997;344:275–283.
6. Thomazeau H, Rolland Y, Lucas C, et al. Atrophy of the supraspinatus belly. Assessment by MRI in 55 patients with rotator cuff pathology. *Acta Orthop Scand* 1996;67:264–268.
7. Zanetti M, Gerber C, Hodler J. Quantitative assessment of the muscles of the rotator cuff with magnetic resonance imaging. *Invest Radiol* 1998;33:163–170.
8. Goutallier D, Postel JM, Bernageau J, et al. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. *Clin Orthop Relat Res* 1994;304:78–83.
9. Lippe J, Spang JT, Leger RR, et al. Inter-rater agreement of the Goutallier, Patte, and Warner classification scores using preoperative magnetic resonance imaging in patients with rotator cuff tears. *Arthroscopy* 2012;28:154–159.
10. Schiefer M, Mendonca R, Magnanini MM, et al. Intraobserver and interobserver agreement of Goutallier classification applied to magnetic resonance images. *J Shoulder Elbow Surg* 2015;24:1314–1321.
11. Spencer EE Jr, Dunn WR, Wright RW, et al. Interobserver agreement in the classification of rotator cuff tears using magnetic resonance imaging. *Am J Sports Med* 2008;36:99–103.
12. Kim JY, Ro K, You S, et al. Development of an automatic muscle atrophy measuring algorithm to calculate the ratio of supraspinatus in supraspinous fossa using deep learning. *Comput Methods Programs Biomed* 2019;182:105063.
13. Medina G, Buckless CG, Thomasson E, et al. Deep learning method for segmentation of rotator cuff muscles on MR images. *Skeletal Radiol* 2021;50:683–692.
14. Taghizadeh E, Truffer O, Becce F, et al. Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder CT datasets. *Eur Radiol* 2021;31:181–190.
15. Yoo HJ, Kim YJ, Hong H, Hong SH, Chae HD, Choi JY. Deep learning-based fully automated body composition analysis of thigh CT: comparison with DXA measurement. *Eur Radiol* 2022.
16. Wallenberg RB, Belzer ML, Ramsey DC, et al. MRI-based 3-dimensional volumetric assessment of fatty infiltration and muscle atrophy in rotator cuff tears. *J Shoulder Elbow Surg* 2022;31:1272–1281.
17. Riem L, Feng X, Cousins M, et al. A Deep learning algorithm for automatic 3D segmentation of rotator cuff muscle and fat from clinical MRI scans. *Radiol Artif Intell* 2023;5:e220132.
18. Werthel JD, Boux de Casson F, et al. Three-dimensional muscle loss assessment: a novel computed tomography-based quantitative method to evaluate rotator cuff muscle fatty infiltration. *J Shoulder Elbow Surg* 2022;31:165–174.