Development of an automatic muscle atrophy measuring algorithm to calculate the ratio of supraspinatus in supraspinous fossa using deep learning

Joo Young Kim [a,1], Kyunghan Ro [b,1], Sungmin You [a], Bo Rum Nam [a], Sunhyun Yook [a], Hee Seol Park [c], Jae Chul Yoo [c], Eunkyoung Park [e,f], Kyeongwon Cho [d,f], Baek Hwan Cho [d,f,**], In Young Kim [a,*]

[a] Department of Biomedical Engineering, Hanyang University, Seoul, South Korea

[b] Bonbridge hospital, 562, Songpa-daero, Songpa-gu, Seoul, South Korea

[c] Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

[d] Medical AI Research Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

[e] Biomedical Engineering Research Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

[f] Department of Medical Device Management and Research, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, South Korea

ARTICLE INFO

ABSTRACT

Background and objective: Rotator cuff muscle tear is one of the most frequent reason of operations in orthopedic surgery. There are several clinical indicators such as Goutallier grade and occupation ratio in the diagnosis and surgery of these diseases, but subjective intervention of the diagnosis is an obstacle in accurately detecting the correct region.

Methods: Therefore, in this paper, we propose a fully convolutional deep learning algorithm to quantitatively detect the fossa and muscle region by measuring the occupation ratio of supraspinatus in the supraspinous fossa. In the development and performance evaluation of the algorithm, 240 patients MRI dataset with various disease severities were included.

Results: As a result, the pixel-wise accuracy of the developed algorithm is 0.9984 ± 0.073 in the fossa region and 0.9988 ± 0.065 in the muscle region. The dice coefficient is 0.9718 ± 0.012 in the fossa region and 0.9463 ± 0.047 in the muscle region.

Conclusions: We expect that the proposed convolutional neural network can improve the efficiency and objectiveness of diagnosis by quantifying the index used in the orthopedic rotator cuff tear.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Rotator cuff tear caused by progressive and degenerative changes in muscle atrophy and fatty infiltration of supraspinatus, infraspinatus, subscapularis and teres minor is one of the most frequent diseases in orthopedic surgery [1–3]. Rotator cuff repair is an independent and highly specialized area of orthopedic surgery and numerous rotator cuff sutures are performed annually. These procedures not only affect the quality of patients lives, but also affect the economics of surgical departments and the innovation of orthopedic companies [4]. Image analysis using magnetic resonance imaging (MRI) plays a critical role in diagnosing rotator cuff tear, predicting the degree of tear, and planning the surgical resolution. Efforts to accurately diagnose and predict rotator cuff tears have been attempted previously [5–8].

In recent years, MRI analysis has been used to predict the reparability of massive rotator cuff tear [9] or to analyze factors affecting re-tear after rotator cuff suture [10]. MRI can assist in determining the method of surgical treatment and provides significant clinical value for surgeons and patients. Supraspinatus tendon tear is the most common cause of rotator cuff tear and plays an

---

important role in diagnosis. In previous studies, atrophy or supraspinatus tendon retraction of supraspinatus muscle affected re-tear or repairability after rotator cuff repair.

Therefore, it is possible to diagnose rotator cuff tear and determine the treatment method by detecting the supraspinatus fossa and muscle area through MRI as a clinical index, and then using this to calculate the occupation ratio. However, the picture archiving and communication system (PACS), which is used for clinical diagnosis, does not have a function to analyze the area of the image. Therefore, it is difficult to make a clinical diagnosis and a study to determine how to perform quantitative analysis is necessary. Various studies have suggested analytical methods for the diagnosis and treatment of rotator cuff disease using MRI. However, one criticism is that these studies only use external programs and have to be analyzed clinically [11]. Therefore, only a simple and easy-to-classify muscle region can be analyzed [12].

Based on existing research, new clinical diagnostic tools should (1) be able to automatically analyze a large amount of data, (2) detect the muscle and fossa regions that are difficult to detect with image processing algorithms, and (3) be accurate so that they are clinically relevant.

Deep learning methods were chosen in algorithm development to diagnose rotator cuff tears and to determine treatment methods that meet the above criteria. In recent years, the field of deep learning has been growing rapidly due to the development of various types of crowd sourcing, big data, and computing equipment. Compared with the development of deep learning technology using publicly available natural image data, the application of deep learning and advanced technology in the field of medical analysis is somewhat slow and uncommon due to the difficulties in collecting data, such as non-standardization, issues in data analysis, and the burden of employing specialists. However, the rapid development of deep learning technology has facilitated various attempts of its use in medical imaging research, and several studies have met various requirements ranging from the quantitative analysis of lesions to diagnosis. In recent years, numerous studies have reported algorithms that can diagnose more accurately than doctors [13]. This has overcome uncertainties about the lack of explanatory power and outcome, which were identified as limitations of deep learning in the medical field, and the accurate detection of regions became possible.

In this study, we propose a deep learning approach to automate the muscle and fossa region detection and calculation of the quantified occupation ratio, which is an indicator of the diagnosis of rotator cuff tear. To develop an algorithm, a dataset was constructed from annotation of the muscle and fossa area in the Y-view, which is the basis for the clinical analysis of supraspinatus among muscles constituting the rotator cuff.

2. Materials and methods

2.1. Materials

In this study, an MRI-based detection algorithm for the supraspinatus muscle (SSP) and fossa region has been developed using deep learning. The data of 240 patients MRI dataset were used in the learning algorithm. From January 2015 to December 2016, single shoulder pain was randomly extracted from the shoulder MRI of patients at the Samsung Medical Center regardless of diagnosis. Since the present study had designed retrospectively, there were no patients who had took MRI without shoulder pain. However, substantial patients, such as shoulder instability or inflammatory disease, had no pathology in supraspinatus muscle or tendon. Therefore, the whole patients from shoulder clinic in a certain period were chosen and analyzed the MRI image without knowing any clinical information to reduce selection bias. So, a
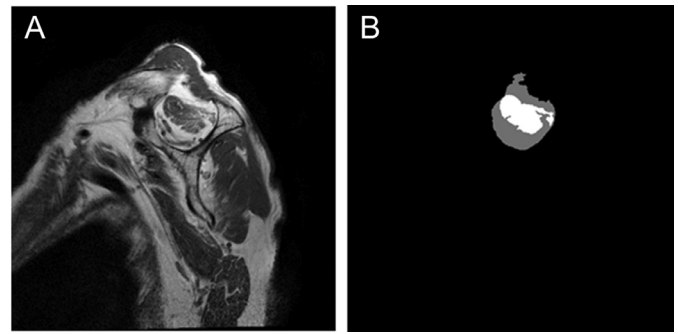
Fig. 1. (A) Sagittal MRI of the Y-view, (B) White: Muscle region, Gray: Fossa region, Black: Background.

considerable data set of nearly normal pathology of supraspinatus muscle and tendon, such as the Goutallier grade 0 or 1, can be secured. The accuracy of the proposed algorithm is improved by using a simple image processing technique to post-process the predicted region and k-fold cross-validation is used to evaluate whether the performance of the algorithm is accurate.

2.2. Data collection and annotation

Data were collected from the MRI slice and analyzed as input and the ground truth was the object to be extracted as the output by analyzing the input. Learning of the data uses the input and ground truth, and the output learns to give the same result as the ground truth. Then, when testing the learning results, only the inputs that are not used for learning are used to analyze how much the output matches the ground truth. The input used in this study was a sagittal oblique plane view with a scapular Y-shaped view image slice of the SSP MRI, and the ground truth was annotated with two regions of supraspinatus fossa and muscle in the image slice.

To annotate the ground truth, we used ITK-snap, a freeware medical image labeling program [14]. ITK-snap supports an adaptive brush tool that can adaptively segment according to the intensity of the periphery of the annotation area. In the annotation region, the fossa region has a clear boundary with the scapular Y-view and the muscle has a high-intensity difference from the surrounding fossa region. Therefore, an adaptive brush is used to minimize the human inspector error.

The supraspinatus fossa and supraspinatus muscle were annotated according to a prior study [15]. First, in the shoulder MRI T1-weighted sagittal oblique view, we chose the most lateral image (called the Y-shaped view) with the scapular spine in contact with the scapular body. When the supraspinatus fossa area was annotated, annotation was performed along the inner bone margin of the Y-shaped scapular, the inferior border of the trapezius, and the inner bone margin of the distal clavicle. When annotating the muscle area, the area drawn along the outer margin of the supraspinatus muscle in the supraspinatus fossa area was annotated as a margin and the neurovascular structure outside the muscle area was excluded. In cases where it is difficult to accurately determine the neurovascular structure with similar signal intensities adjacent to the muscle in a segmentation, serial and anterior and posterior images based on the segmented image are analyzed together to confirm the position of vessels and nerves (Fig. 1).

The data were randomly collected regardless of the diagnosis, therefore, if the severity of the rotator cuff tear is biased to normal or severe in the entire data, a problem of imbalance may cause the learning result to be unsuitable for general use. This is a problem that can cause uncertainty in the results when deep learning is

applied to medical data, therefore, the severity of each of the 240 entries configured into the dataset was determined.

2.3. Severity criteria for patients

Quantitative indicators used to diagnose rotator cuff tear include the occupation ratio stage, which is an index of muscle atrophy, and Goutallier grade, which is the degree of fatty infiltration in muscles.

The muscle atrophy stage is the ratio occupied by muscle in the fossa area and classified as stage $I \geq 0.6$, stage II (0.4–0.6) and stage III < 0.4 according to the occupation ratio [16].

$$Muscle\ Atrophy\ Stage = \begin{cases} 1 & (O.R. \geq 0.6) \\ 2 & (0.4 \leq O.R. < 0.6) \\ 3 & (O.R. < 0.4) \end{cases} \quad (1)$$

The stage according to the occupation ratio changed according to the degree of muscle atrophy, and the size of muscle area changed according to the occupation ratio in the 80 cases of stage I, 68 cases of stage II, and 92 cases of stage III that were collected.

Goutallier grade [17] is an indicator based on fatty infiltration in the muscle area, given as Grade 0 if there is no fat in the muscle, Grade 1 if streaks of fat exist, Grade 2 if there is more muscle than fat, Grade 3 when the muscle and fat area are the same, and Grade 4 when the fat area is larger than the muscle area. The Goutallier grade is not an indicator of muscle atrophy. However, the fossa area does not change in absolute size even when the muscles are atrophied and filled with fat or the neurovascular structure is revealed by the tear of the rotator cuff. The muscle area is not only varied in size due to muscle atrophy, but fatty infiltration can also affect the learning of the algorithm by making the intensity within the muscle region inconsistent. Therefore, we used 55 G0 with normal grade, and 75 G1, 68 G2, 23 G3, and 19 G4 with the highest severity in the algorithm development.

All the annotation using ITK-SNAP and fatty infiltration evaluation by Goutallier grade were conducted by the consensus of two orthopedic surgeon (K.H.R and H.S.P). If there is disagreement between two surgeons, it was discussed with senior professor who were specialized to shoulder & elbow clinic (J.C.Y)

In the collected MRI data, there are 235 right shoulders and 5 left shoulders. Therefore, the data are biased to the right side. A method of solving this biased problem is proposed in the algorithm learning process below.

2.4. Convolutional neural network design

In this study, a convolutional neural network (CNN) was used to develop a deep learning algorithm. CNN is widely used in natural image classification because it is similar in structure and characteristics to those used in conventional 2D image processing fields. For the same reason, studies showing similar or superior performance to those in the field of medical image analysis have been reported [18–20]. Recently, convolution networks suitable for various purposes have been developed and used in various fields. In this study, we use a fully convolutional network (FCN) [21], which can obtain high-resolution output using skip layers.

The FCN consists of a convolution layer, pooling layer, fully convolution layer, and skip layer. Every time an input passes through a convolution layer, a 2D convolution operation of the specified size is performed, and it learns to approach the output whenever this process occurs. The feature map through the convolution layer is reduced in size every half of the max pooling process that picks up a significant value in the pooling layer, thus enabling efficient computation even in a 2D convolution specified at the same size. The feature map that passes all convolution and pooling layers, passes

Table 1
Detailed structure of the fully convolutional network used in algorithm development.

Layer	Filter shape	Output shape
Input	N.A.	$512 \times 512 \times 1$
Conv1_1	$3 \times 3 \times 1 \times 64$	$512 \times 512 \times 64$
Conv1_2	$3 \times 3 \times 64 \times 64$	$512 \times 512 \times 64$
Pool1	N.A.	$256 \times 256 \times 64$
Conv2_1	$3 \times 3 \times 64 \times 128$	$256 \times 256 \times 128$
Conv2_2	$3 \times 3 \times 128 \times 128$	$256 \times 256 \times 128$
Pool2	N.A.	$128 \times 128 \times 128$
Conv3_1	$3 \times 3 \times 128 \times 256$	$128 \times 128 \times 256$
Conv3_2, Conv3_3, Conv3_4	$3 \times 3 \times 256 \times 256$	$128 \times 128 \times 256$
Pool3	N.A.	$64 \times 64 \times 256$
Conv4_1	$3 \times 3 \times 256 \times 512$	$64 \times 64 \times 512$
Conv4_2, Conv4_3, Conv4_4	$3 \times 3 \times 512 \times 512$	$64 \times 64 \times 512$
Pool4	N.A.	$32 \times 32 \times 512$
Conv5_1, Conv5_2, Conv5_3	$3 \times 3 \times 512 \times 512$	$32 \times 32 \times 512$
Pool5	N.A.	$16 \times 16 \times 512$
FC1	$1 \times 1 \times 512 \times 4096$	$16 \times 16 \times 4096$
FC2	$1 \times 1 \times 4096 \times 4096$	$16 \times 16 \times 4096$
FC3	$1 \times 1 \times 4096 \times 2$	$16 \times 16 \times 2$
Deconv1	$4 \times 4 \times 2 \times 512$	$32 \times 32 \times 512$
Deconv2	$4 \times 4 \times 512 \times 256$	$64 \times 64 \times 256$
Deconv3 (Output)	$16 \times 16 \times 256 \times 2$	$512 \times 512 \times 2$

through the convolution layer and is used for the full-size convolution operation size rather than a specific size, thereby minimizing the loss of information and analyzing the entire region. However, the previous process is smaller than the input size due to the pooling layer. In this case, it is difficult to use in the area detection algorithm. To overcome this problem, it is possible to obtain the probability distribution output of the same size as the input image by up-sampling the size of the feature map of the convolution layer and the third and fourth pooling layers by using the skip layer and elementwise sum of these features. Since the supraspinatus fossa and muscle region are selected as the ground truth and used for learning, it is possible to obtain the probability distribution and each of these two regions can be classified in pixel units. The detailed structure is composed of 15 convolution layers and 5 pooling layers based on the VGG19 Network, [22] as shown in Table 1, however three fully convolutional layers have been added. The feature map of the 3rd and 4th pooling layer was up-sampled by a skip layer and obtained as the output. In order to avoid misconvergence or divergence of the network in training, we firstly initialized our network using the ImageNet pre-trained models and then fine-tuned with the added fully convolutional layers.

In the learning process, two images were used per batch, and the Adam optimizer was used for the gradient descent method. The loss function used in the learning process for pixel unit classification is softmax cross entropy, which is widely used in the related field, and calculates the error between the output of the network and the ground truth [23].

2.5. Augmentation

A common problem when developing analysis algorithms for medical images using deep learning is the lack of data. This is because the absolute amount of patient data is small and susceptible to bias, therefore the learning requirements are not met due to the high human cost of the annotation process to create the ground truth for learning. There are various types of augmentation techniques that are widely used. The augmentation used in this study is flip and brightness control [24]. Our original dataset includes more right arm data. Therefore, we used flip data augmentation to resolve this data imbalance problem. In addition, as shown in
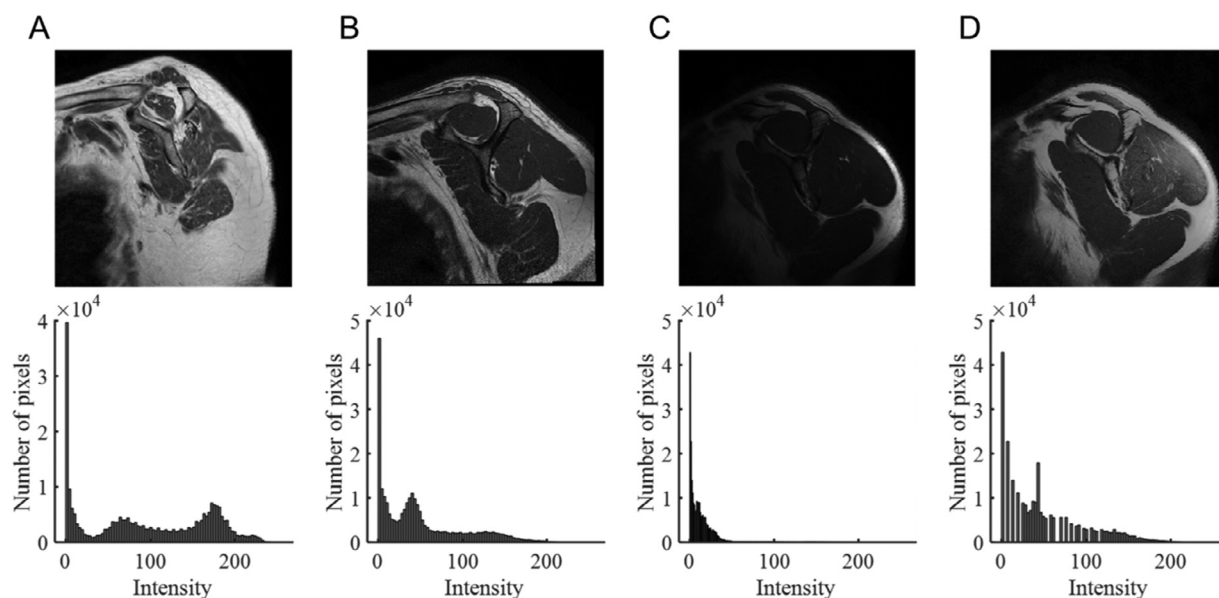
Fig. 2. (A) Brightest, (B) Medium (C) Darkest MRI image and histogram. Histogram changes of images owing to differences in MRI equipment or patient. (D) Histogram adjusted by the medium brightness distribution image and histogram of (C).
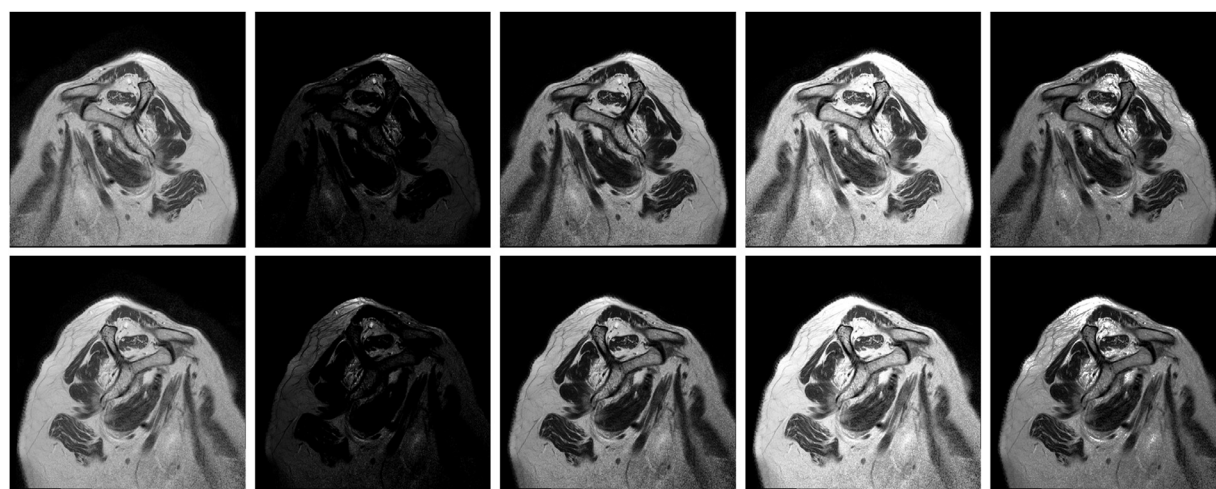

Fig. 3. Examples of applied augmentation results. Each row shows the result of the flip control. Each column represents the original brightness, the histogram deformation results reflecting the darkest image, the median image, the brightest image, and the average brightness distribution of the entire image.

Fig. 2, the MRI image shows a distinct difference in brightness and contrast, which is slightly different in all 240 data. Thus, the augmentation technique was used to adjust the histogram because the intensity of the image was different depending on the patient and the MRI equipment. The intensity range is obtained from the whole image as a histogram. And the augmentation was performed by dividing it into five categories: dark distribution, median brightness distribution, bright distribution, average of all distributions and original image. In this process, three brightest, darkest, and intermediate brightness images were extracted, and the average of three histograms was calculated and expressed as a sum of the normal distributions. This intensity adjustment augmentation is applied randomly. In four adjustment cases except for the original is selected, match the histogram of the input image with the representative histogram distribution. By using the above two techniques, as shown in Fig. 3, the number of images used for learning can be increased by 2 times through flip control and 5 times by brightness control.

2.6. k-fold cross-validation

The total number of shoulder MRI data used in this study is 240, which is insufficient in terms of quantity. In addition, even if augmentation is performed, the results are different depending on how the whole data is divided into the training set and test set since the shapes are similar. Therefore, the k-fold cross-validation technique is used because it is difficult to guarantee reliability by using the hold-out test method, which divides the training set and test set according to a certain ratio. In this method, the fold is repeated *n* times by using the total number of datasets divided by n as a test set and *n*-1 sets as training sets. In this study, 10-fold cross-validation was used to evaluate the performance of the developed algorithm. The split process of the total data for the 10-fold cross validation is performed before the augmentation mentioned above. After the test set and the training set are separated, the augmentation is performed only on the training set and the trained network is evaluated on the blind test set. The raw output
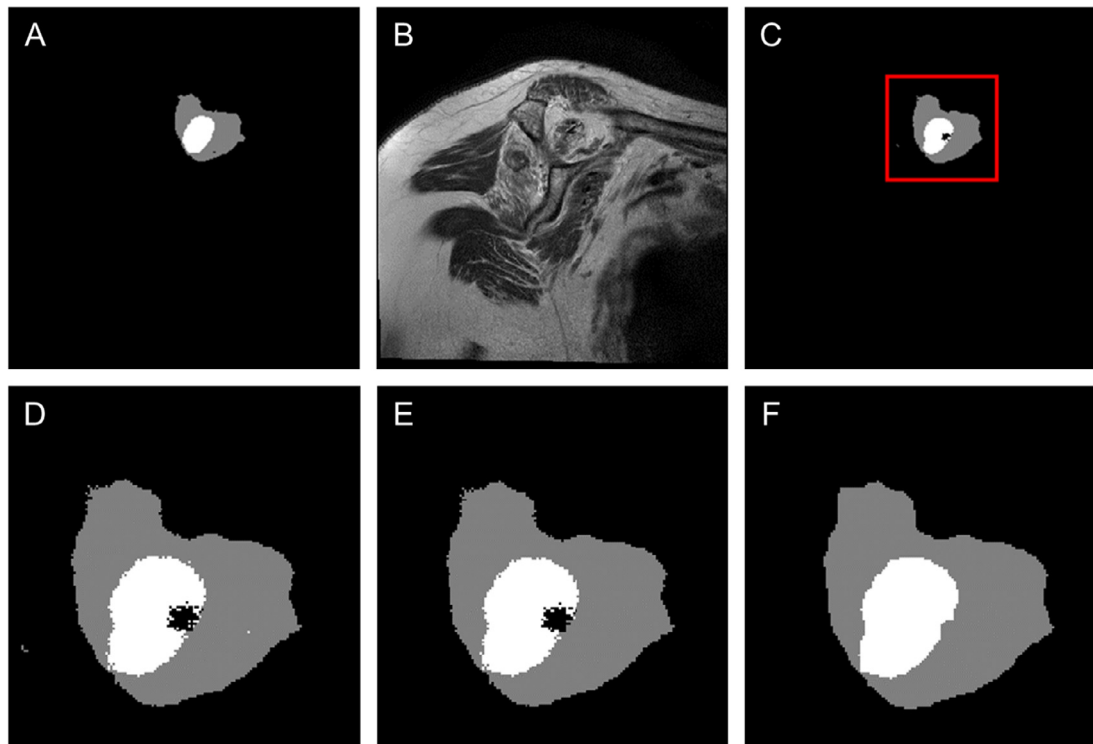
Fig. 4. (A) Ground truth, (B) input image, (C) prediction result, (D) enlarged red box area of (C) with scattered noise and empty space, (E) scatter noise removal through blob filtering, and (F) elimination of empty space by applying the closing and filling algorithms.
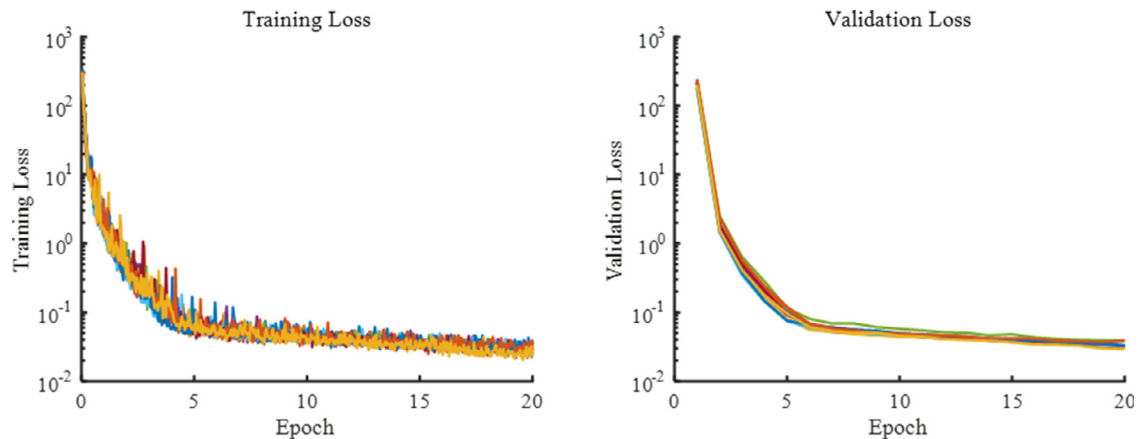
Fig. 5. Decrease of training and validation loss of 10-fold learning according to the learning epoch.

of the developed algorithm causes scatter noise and blank regions, as shown in Fig. 4, due to the receptive field limit of the CNN. This area is a factor that lowers the accuracy of the algorithm, and post-processing was added to overcome this issue.

2.7. Post-processing

The raw output of the developed algorithm causes scatter noise and blank regions, as shown in Fig. 4, due to the receptive field limitation of CNN. Since this area is a factor that lowers the accuracy of the algorithm, post-processing is applied to overcome it. First, all blobs without adjacent pixels are extracted and blob filtering for removing scatter noise is performed by leaving only the largest blob. Then, a closing and filling algorithm is applied to fill the blank space inside the blob. The algorithm applied to post-processing is not complex but has greatly improved the performance of the algorithm.

3. Results

In this study, algorithm learning was performed using 240 shoulder images, and all the data were randomly divided into 10 sets and 10 learnings were performed with 10-fold cross-validation. As shown in Fig. 5, the training loss was 303.6475 and the validation loss was 209.8082 at the beginning of learning, and at the end of learning (after 200 epochs), the training loss is 0.0015 and the validation loss is 0.0122. Based on these results, the loss value converges to zero as the learning progresses in both training and validation. Since the loss value computed by softmax cross entropy is proportional to the pixel-wise error across the entire image, it was decided that the learning of the algorithm was done properly, and the analysis was conducted.

In the result analysis, pixel accuracy was calculated by comparing the 512×512 size prediction output with the ground truth, as shown in Fig. 6. For further analysis, we used the dice coefficient
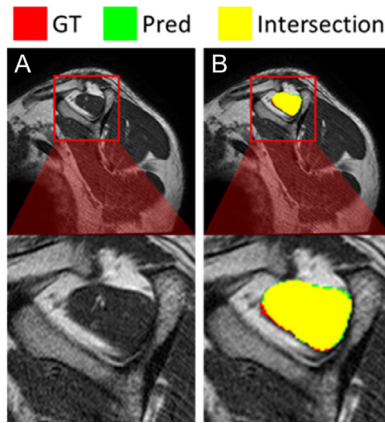
Fig. 6. (A) The input image, (B) comparison image of prediction and ground truth; dice coefficient in the example is 0.9727.

Table 2

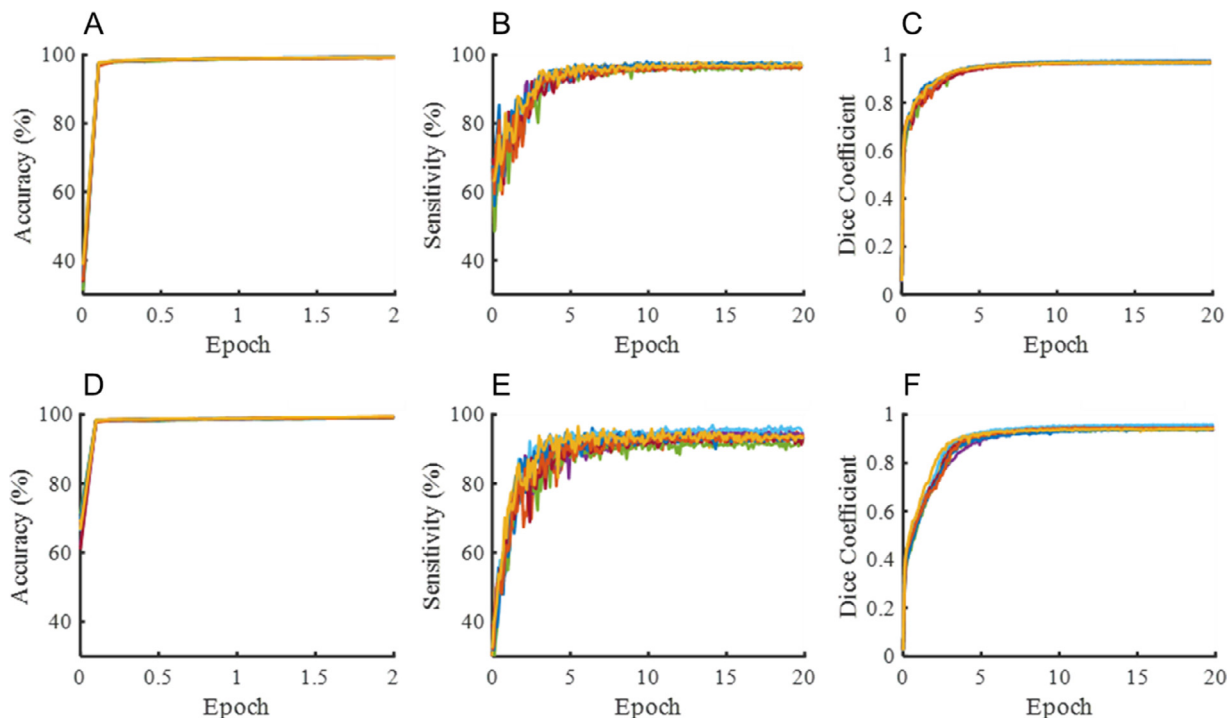Average of 10-fold learning pixel-wise accuracy and dice coefficient in the fossa and muscle regions.

	Fossa region	Muscle region
Pixel-wise accuracy	0.9984 ± 0.073	0.9988 ± 0.065
Dice coefficient	0.9718 ± 0.012	0.9463 ± 0.047

(DC), which is a frequently used index in area classification study, as shown in Eq. (2).

$$DC = \frac{2 * Area_{Intersection}}{Area_{Union} + Area_{Intersection}} \qquad (2)$$

The accuracy, sensitivity, and DC of the epoch in the learning process were as shown in Fig. 7 in the fossa and muscle regions. As shown in Table 2, the final learning results have a pixel-wise accuracy of 0.9984 and a DC of 0.9718 in the fossa region. In the muscle region, the pixel-wise accuracy was 0.9988 and the DC was 0.9463. The results are shown as the average values of the 10-fold learning results.

Because the morphologic change depends on the severity of the patient, the results according to the severity were derived. The severity was based on the atrophy stage and the Goutallier grade. The results determined using the atrophy stage are displayed in Table 3. In the fossa region, the performance of the algorithm did

not change significantly according to the severity, but the performance degraded with the severity in the muscle region. The atrophy stage classification results are shown in Fig. 8.

Table 4 and Fig. 9 show the results for Goutallier grade differences. In this result, there was no correlation between the Goutallier grade and fossa region detection performance, but there was performance degradation according to the grade of the atrophy stage in the muscle region. In addition, image analysis revealed that the size of the muscle region, as well as the fatty infiltration in the muscles, decreased proportionally according to the Goutallier grade.

The result of the algorithm analyzed by the Dice coefficient is for the segmented area. Since the area is the component of the occupation ratio, which is the index used for diagnosis, it is necessary to analyze the consistency of the total patient data by comparing the occupation ratios derived from the clinical diagnosis and the algorithm results. Table 5 shows the results of the comparison using ICC (Intra-class correlation coefficient). ICC tends to decrease according to higher Goutallier grade of atrophy stage, but in all cases showed perfect agreement between clinician and CNN results which was statistically significant.

A scatter plot of the occupation ratio according to the goutallier grade is shown in Fig. 10. The occupation ratio of the graph is obtained through the developed deep learning diagnostic tool. According to this scatter plot and trend line, patients with low occupation ratio are related with the high Goutallier grade, which is severe fatty infiltration. This result is consistent with the clinician's results.

Fig. 7. (A) Graph of 10-fold learning pixel-wise accuracy, (B) sensitivity, and (C) dice coefficient in the fossa and (D) 10-fold learning pixel-wise accuracy, (E) sensitivity, and (F) dice coefficient in the muscle regions.

Table 3

Pixel-wise accuracy and dice coefficient of each fossa and muscle region according to the atrophy stage.

Atrophy stage	Fossa region		Muscle region	
	Pixel-wise accuracy	Dice coefficient	Pixel-wise accuracy	Dice coefficient
Stage 1	0.9983 ± 0.0847	0.9718 ± 0.0135	0.9989 ± 0.0458	0.9749 ± 0.0083
Stage 2	0.9985 ± 0.0593	0.9720 ± 0.0115	0.9989 ± 0.0822	0.9541 ± 0.0454
Stage 3	0.9985 ± 0.0692	0.9714 ± 0.0115	0.9988 ± 0.0660	0.9158 ± 0.0497

Table 4

Pixel-wise accuracy and dice coefficient of each fossa and muscle region according to the Goutallier grade.

Goutallier grade	Fossa region		Muscle region	
	Pixel-wise accuracy	Dice coefficient	Pixel-wise accuracy	Dice coefficient
Grade 0	0.9982 ± 0.0777	0.9714 ± 0.0110	0.9989 ± 0.0498	0.9751 ± 0.0085
Grade 1	0.9984 ± 0.0706	0.9721 ± 0.0131	0.9990 ± 0.0375	0.9661 ± 0.0157
Grade 2	0.9985 ± 0.0637	0.9727 ± 0.0112	0.9989 ± 0.0512	0.9400 ± 0.0287
Grade 3	0.9985 ± 0.0819	0.9692 ± 0.0138	0.9988 ± 0.0848	0.9043 ± 0.0446
Grade 4	0.9985 ± 0.0811	0.9708 ± 0.0137	0.9981 ± 0.1304	0.8591 ± 0.0880

Table 5

Intra-class correlation coefficient of occupation ratio between clinician and CNN.

|	Intraclass	95% confidence interval		P-value
	Correlation	Lower limit	Upper limit	
Total	0.992	0.990	0.994	<0.001
Atrophy stage				
Stage 1	0.973	0.959	0.983	<0.001
Stage 2	0.938	0.900	0.962	<0.001
Stage 3	0.918	0.877	0.946	<0.001
Goutallier Grade				
Grade 0	0.992	0.986	0.995	<0.001
Grade 1	0.989	0.982	0.993	<0.001
Grade 2	0.974	0.957	0.984	<0.001
Grade 3	0.884	0.728	0.951	<0.001
Grade 4	0.820	0.533	0.931	<0.001

4. Discussion

The accuracy according to the epochs showed a steep rise in both fossa and muscle areas, as shown in Fig. 7. This is because the ratio of the area to be detected in the entire image is relatively small and the position is concentrated in the first and second quadrants of the image. In the same graph, sensitivity and DC converge relatively quickly in the fossa region, while the muscle region is stabilized after a slight fluctuation. This is because the fossa area is in the bone structure to form the Y-view, whereas the muscle area is highly variable in size due to the severity of the patient's supraspinatus tendon integrity, the size of the tear, and the effect on blood vessels and fatty infiltration.

In this study, the fossa region and the muscle region of the supraspinatus Y-view of the rotator cuff muscle were detected, therefore it was confirmed that it is possible to detect the
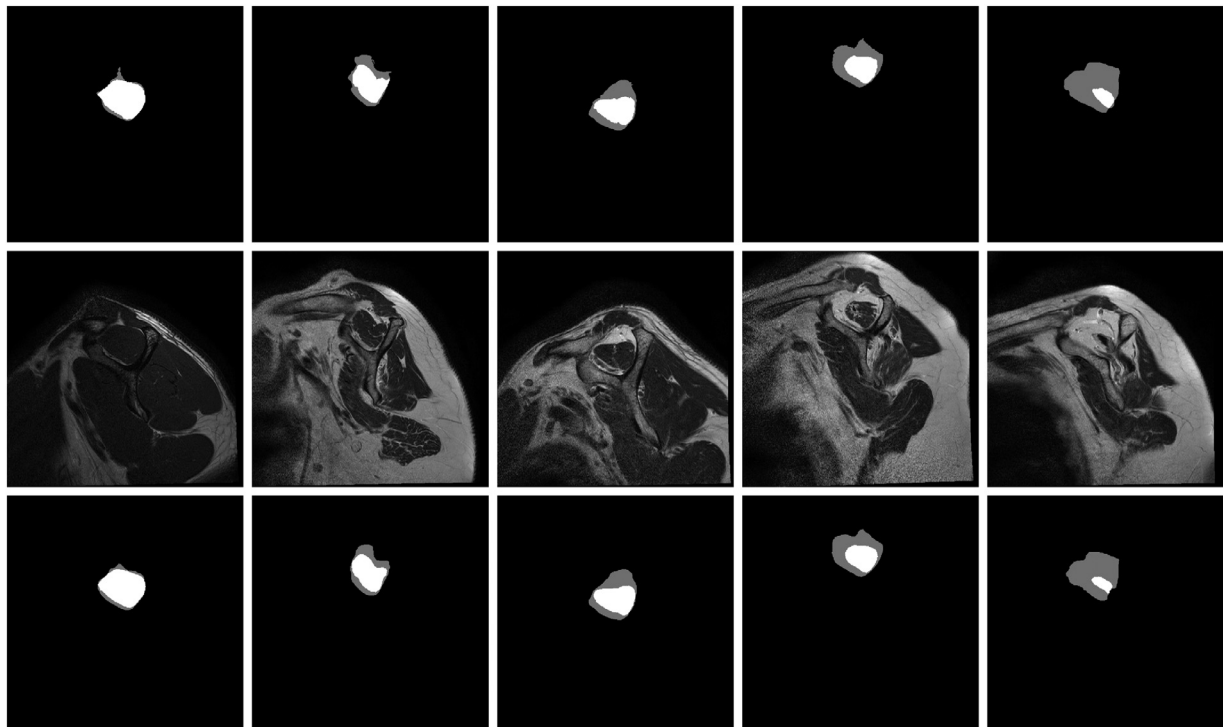
Fig. 8. Detection of the fossa and muscle region according to the atrophy stage. Each row is the ground truth, input image, and prediction result. The occupation ratio in the ground truth of each column are 0.8369 (stage I), 0.6008 (≈0.6), 0.5256 (stage II), 0.4009 (≈0.4), and 0.1339 (stage III) from left to right.
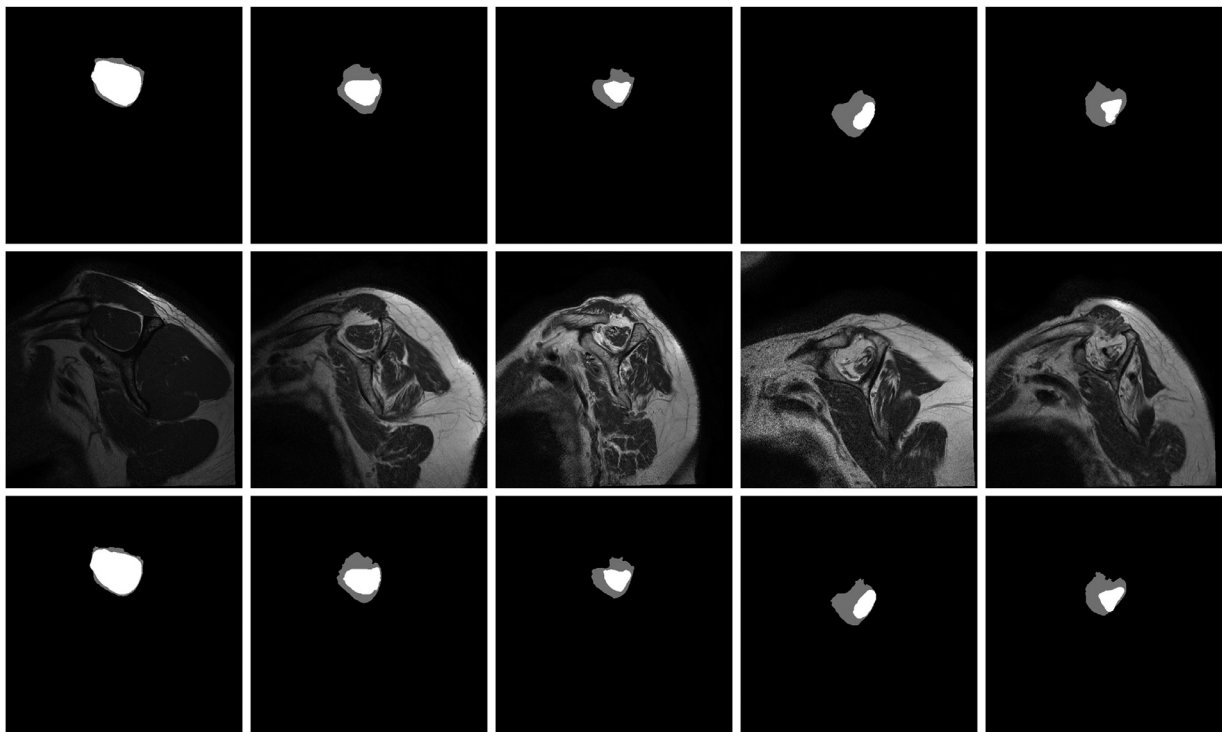
Fig. 9. Detection of fossa and muscle region according to the Goutallier grade. Each row is the ground truth, input image, and prediction result and each column are grades 0–4 from left to right.
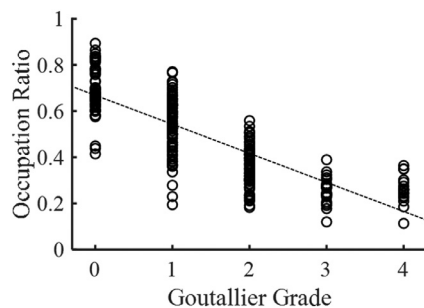
Fig. 10. The scatter plot of the occupation ratio according to the Goutallier grade. The dotted line is the trend line of the whole element.

occupation ratio by deep learning. The RMSE(Root mean square error) value with the occupation ratio based on the ground truth area by the clinician's cross-check was 0.033, which is effective for clinical diagnosis. The detection accuracy of the fossa region according to severity was not significantly different. However, in the case of the muscle area, the error increases according to the severity. To confirm the reason, three results with the lowest detection accuracy of the muscle region are shown in Fig. 11. Fig. 11(a) is the third from the worst case. In this case, the DC is 0.7934 and according to the judgment of the clinician, the prediction result can substitute the ground truth.

On the other hand, in the cases shown in Fig. 11(b) and (c), fat degeneration occurred in muscle fiber according to severity and the classification was less accurate. Especially in Fig. 11(c), when the fat degeneration of the muscle is severe and only one slice is used to detect the area, it is classified as a case where even a clinician found it difficult to distinguish the muscle contour. However, it can be detected by analyzing the adjacent slice in a clin-

ical situation. Therefore, a 3D CNN network capable of analyzing the neighboring slice may be used in the future, and the development of a more accurate algorithm will be possible. The severity of the data collection, such as in the example, will also need to be done in parallel. The CNN of this study were used to annotate the outline of the supraspinatus fossa and muscle. So, it can help quantitative analysis, such as occupation ratio. The network used in the proposed algorithm is not based on the state-of-the-art network in the latest image segmentation reports. Nonetheless, it was able to obtain a high-resolution output with high-resolution input using the simplicity of the structure while obtaining sufficient accuracy for clinical use with only one slice of MRI. Despite the small amount of our dataset, the use of augmentation and the pre-trained model may have led to good results. One reason for the successful results might be that the shapes and locations of the fossa area were not much varied among the Y-view MRI slice dataset. This could be another reason of the success even though we adopted relatively simple VGG based network architectures. We also tried inception based network [25], but the results were not better than our VGG based model, which is consistent with the results in a previous study that compared both architectures for semantic segmentation [21]. With relatively few data and simple network models, the results have shown sufficient segmentation performance for clinical use, so we did not explore the model architectures more extensively. However, a further research will be needed on different network architectures and training methods for better and faster clinical use.

The purpose of this study was to classify the muscle and fossa areas from the MRI of rotator cuff tear patients. Various indexes such as the occupation ratio [26] and Goutallier grade were used in predicting the diagnosis and treatment prognosis of rotator cuff tears, but most of the diagnoses depend on the intuitive judgment of the clinician. Fatty infiltration by Goutallier grade is qualitative analysis which was not treated by the CNN in the present study,
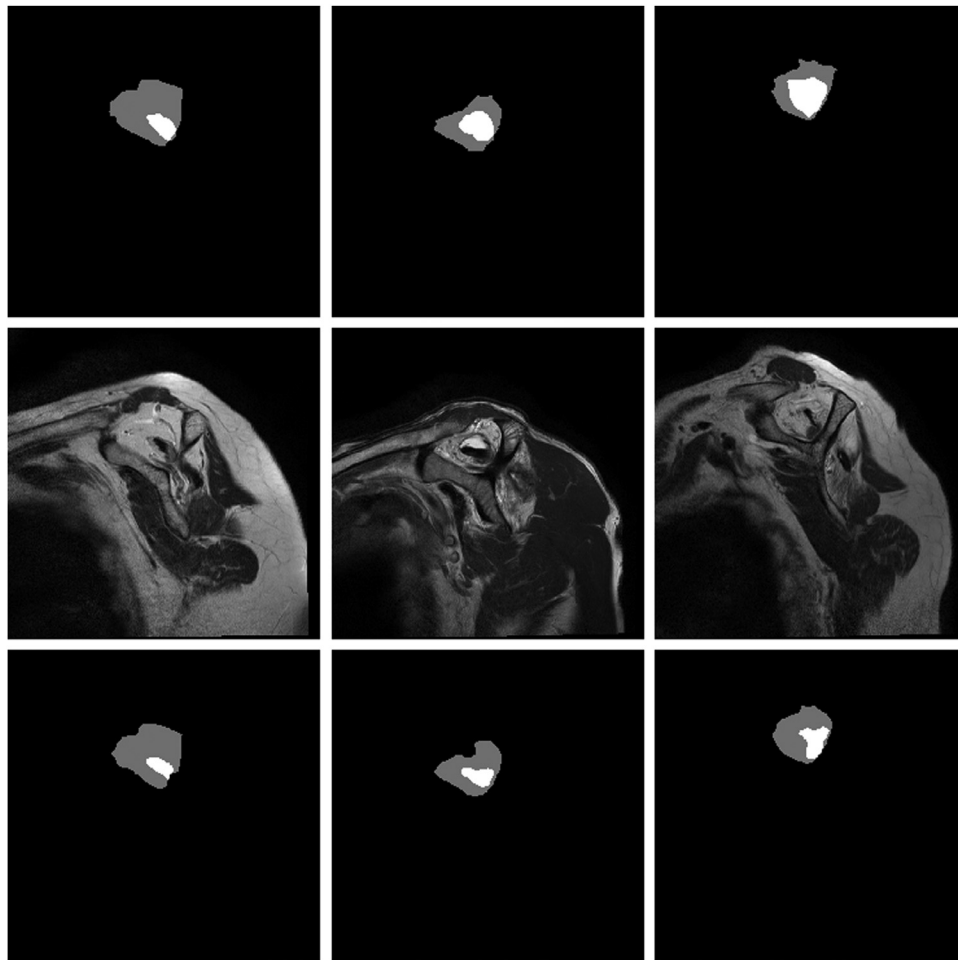
Fig. 11. Three bad results of muscle detection. Each row is the ground truth, input image, and prediction result. The worst result is on the right. Each dice coefficient is 0.7934, 0.6616, and 0.6475 from left to right.

so it was evaluated by clinician. In the future, the analysis for the fatty infiltration of the muscle should be studied based on this CNN tool. The process of annotation by clinicians which contains the efforts to achieve elaborate and accurate data without human error was complicated and highly time-consuming. Meanwhile, the analysis from CNN in the present study shows high accuracy compared to clinician. By comparing the dice coefficient values and intra-class correlation coefficient of occupation ratio in Table 5, the proposed system shows the desired learning results with the given data. Furthermore, the detection of the supraspinatus muscle and fossa took only 0.1483 s (148.3369 ms) with a high-performance computer (Intel® Core™ i7-7700 CPU @ 3.60 Ghz, 32.0GB RAM, NVIDIA GeForce GTX 1080 Ti 11 Gbps) and 0.2472 s with a mid-budget computer (Intel Core i5-9600 K @ 3.70 GHz, 8.0GB RAM, NVIDIA GeForce GTX 1060). It took 6.7904 s without the GPU card. Therefore, even when we do not use graphical processing functionality, our approach can save a substantial amount of time compared to a direct clinician's annotation, which takes over five minutes per image. This automated system for detecting region of interest from the MRI not only has the advantage of saving time but also has possibility of increased accuracy, consistent interpretation, efficiency, reproducibility, and lower barriers to approach. Although the FCN is computationally intensive architecture, the computational time in inference is near real-time. However, considering deploying our models into a high-throughput PACS system, reducing computational cost by model compression or knowledge distillation could be further developed in the future study. This result not

only helps clinicians to plan the treatment of patients, but also allows them to accurately track the before and after surgery changes in supraspinatus muscle volume of the supraspinatus fossa using follow-up MRI. In addition, it is possible to provide more objective and reliable data for various studies using these indicators, and it is expected to save time and reduce the measured bias in studies that require a large amount of MRI analysis. By using the results described above, it is expected that various statistical analysis and follow-up studies will be possible through the data analysis of actual rotator cuff tendon tear patients.

Declaration of Competing Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2019.105063.

References

[1] H. Thomazeau, E. Boukobza, N. Morcet, J. Chaperon, F. Langlais, Prediction of rotator cuff repair results by magnetic resonance imaging, Clin. Orthop. Rel. Res. 344 (1997) 275–283.

[2] K. Nakagaki, J. Ozaki, Y. Tomita, S. Tamai, Function of supraspinatus muscle with torn cuff evaluated by magnetic resonance imaging, Clin. Orthop. Rel. Res. (1995) 144–151.

[3] D. Goutallier, J.M. Postel, J. Bernageau, L. Lavau, M.C. Voisin, Fatty infiltration of disrupted rotator cuff muscles, Rev. Rhum. Engl. Ed. 62 (1995) 415–422.

[4] Randelli et al., history of rotator cuff surgery, 10.1007/s00167-014-3445-z

[5] H. Ellman, Diagnosis and treatment of incomplete rotator cuff tears, Clin. Orthop. Rel. Res. 254 (1990) 64–74.

[6] D. Patte, Classification of rotator cuff lesions, Clin. Orthop. Rel. Res. 254 (1990) 81–86.

[7] H.L. McLaughlin, (1994) lesions of the musculotendinous cuff of the shoulder. The exposure and treatment of tears with retraction, Clin. Orthop. Rel. Res. 304 (1944) 3–9.

[8] J. Davidson, S.S. Burkhart, The geometric classification of rotator cuff tears: a system linking tear pattern to treatment and prognosis, Arthroscopy 26 (3) (2010) 417–424.

[9] Kim et al., 2017 AJSM, can preoperative magnetic resonance imaging predict the reparability of massive rotator cuff tears.

[10] Jeong et al., 2018 AJSM, actors predictive of healing in large rotator cuff tears; is it possible to predict retear preoperatively.

[11] S.-K. Tae, et al., Evaluation of fatty degeneration of the supraspinatus muscle using a new measuring tool and its correlation between multidetector computed tomography and magnetic resonance imaging, Am. J. Sports Med. 39 (3) (2011) 599–606.

[12] S. Kim, et al., Automatic segmentation of supraspinatus from MRI by internal shape fitting and autocorrection, Comput. Methods Programs Biomed. 140 (2017) 165–174.

[13] V. Gulshan, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (22) (2016) 2402–2410.

[14] http://www.itksnap.org, accessed on Nov. 13, 2017.

[15] J.Y. JEONG, et al., Supraspinatus muscle occupation ratio predicts rotator cuff reparability, J. Shoulder Elb.Surg. 26 (6) (2017) 960–966.

[16] V. Khoury, É. Cardinal, P. Brassard, Atrophy and fatty infiltration of the supraspinatus muscle: sonography versus MRI, Am. J. Roentgenol. 190 (4) (2008) 1105–1111.

[17] D. Goutallier, et al., Fatty muscle degeneration in cuff ruptures: pre-and postoperative evaluation by CT scan, Clin. Orthop. Rel. Res. 304 (1994) 78–83.

[18] G. Litjens, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[19] Y. Liu, et al., Artificial intelligence-based breast cancer nodal metastasis detection, Arch. Pathol. Lab. Med. (2018).

[20] D.S. Kermany, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122–1131.

[21] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[22] Simonyan, KarenK, and Andrew A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014).

[23] W. Liu, et al., Large-margin softmax loss for convolutional neural networks, ICML 2 (3) (2016).

[24] Perez, LuisL, and Jason J. Wang. The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621 (2017).

[25] C. Szegedy, et al., Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[26] H. Thomazeau, et al., Atrophy of the supraspinatus belly assessment by MRI in 55 patients with rotator cuff pathology, Acta Orthop. Scand. 67 (3) (1996) 264–268.