



# Mask Region-Based Convolutional Neural Network segmentation of the humerus and scapula from proton density-weighted axial shoulder magnetic resonance images

Aysun Sezer 

Biruni University, Computer Engineering, Istanbul, Türkiye

Humerus and scapula are the two main bony structures of the shoulder joint. Different from other ball and socket articulations of the human body, glenohumeral articulation has a limited bony congruency but a strong soft tissue coverage, permitting a wide range of movement. However, this architecture of shoulder joint creates a vulnerability to luxation with eventual instability. Shoulder instability is a growing problem of sportive population due to increasing high-demand participation.<sup>[1]</sup>

In case of instability, analysis of the glenohumeral articulation often requires advanced imaging techniques as there is a risk of soft tissue injury and bony deformation resulting from initial trauma or instability. Concurrent soft tissue and bony lesions of humerus and scapular glenoid are not rare after

Received: June 30, 2023

Accepted: July 29, 2023

Published online: September 20, 2023

**Correspondence:** Aysun Sezer, Biruni Üniversitesi, Bilgisayar Mühendisliği, 34010 Zeytinburnu, İstanbul, Türkiye.

E-mail: asezer@biruni.edu.tr

Doi: 10.52312/jdrs.2023.1291

**Citation:** Sezer A. Mask Region-Based Convolutional Neural Network segmentation of the humerus and scapula from proton density-weighted axial shoulder magnetic resonance images. Jf Dis Relat Surg 2023;34(3):583-589. doi: 10.52312/jdrs.2023.1291

©2023 All right reserved by the Turkish Joint Diseases Foundation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (<http://creativecommons.org/licenses/by-nc/4.0/>).

## ABSTRACT

**Objectives:** This study aimed to evaluate the effectiveness of Mask Region-Based Convolutional Neural Network (R-CNN) in humerus and scapula segmentation.

**Patients and methods:** The study included 665 axial proton density (PD)-weighted magnetic resonance images of 665 consecutive shoulder instability patients (412 males, 253 females; mean age: 27±5.2 years; range, 18 to 42 years) between January 2011 and December 2014. Mask R-CNN was used to automatically segment humerus and scapula regions simultaneously. Segmentation success of Mask R-CNN was compared to the manual segmentation results of an orthopedic surgeon. Statistical evaluation was done with the Dice coefficient and the mean average precision) score. According to the humeral head structure three groups were generated: the healthy humeral head group, the edematous humeral head group, and the Hill-Sachs group (humeral heads with Hill-Sachs lesions).

**Results:** In the test images, 81 humeral heads were healthy, 100 were edematous, and 38 had a Hill-Sachs lesion. According to the Dice metric, the overall success rate of Mask R-CNN configuration was 96.47 and 93.87% for the segmentation of the humeral head and scapula, respectively, and 95.86 and 92.35% for an intersection over union of 0.5 according to the mean average precision. According to the Dice metric, the segmentation success of the humerus and scapula of the healthy group was 94.58 and 97.42%, the segmentation success of the edematous humerus group was 93.56 and 96.53%, and the segmentation success of the Hill-Sachs group was 93.47 to 95.48%. The segmentation success of scapula in the case of discontinuity was 92.86% according to Dice metric.

**Conclusion:** Mask R-CNN-based humerus and scapula segmentation provided promising results compared to manual segmentation of an expert. Mask R-CNN overcomes the problem of discontinuous edges and Rician noise in axial PD-weighted shoulder magnetic resonance imaging.

**Keywords:** Deep learning, humerus and scapula segmentation, Mask R-CNN, PD-weighted magnetic resonance imaging.

a luxation episode.<sup>[2]</sup> Therefore, axial proton density (PD)-weighted imaging, which is a powerful magnetic resonance imaging (MRI) modality in the soft tissue and bone at the same time, is demanded very often to localize soft tissue lesions like labral tears.

Object detection, classification, and instance segmentation are actively evolving fields in computer vision research. In the literature, classical computer vision methods have been employed to segment bony regions, including active contour, watershed-based region growing, and multiatlas-based methods on different imaging modalities.<sup>[3,4]</sup> However, none of the image segmentation algorithms are foolproof and may be jeopardized by some of the aspects of magnetic resonance (MR) images, such as noise and anatomical complexity.<sup>[5,6]</sup>

Although clinically useful in the detection of bone edema and soft tissue injury, PD-weighted MRI is prone to a considerable blur, so called image noise. Certain bony borders, particularly those of thin cortices and in the vicinity of traumatized regions may not be clearly demonstrated by PD-weighted MR sequence.<sup>[5]</sup> Classical approaches often face challenges when dealing with noisy objects and images with intensity inhomogeneity, resulting in limited performance. Deep learning-based techniques, such as semantic and instance-based segmentation methods, have emerged as effective solutions to address these problems and have shown promising results in the field of bone segmentation.<sup>[7-10]</sup>

Instance-level segmentation, on the other hand, is based on target detection, where different objects in an image are identified and classified. The original model for target detection is called Region Convolutional Neural Network (R-CNN). It selects candidate regions likely to contain target objects through selective search, then employs a convolutional neural network to classify each region and determine if it contains the target. Furthermore, Mask R-CNN was introduced to extend R-CNN to pixel-level detection, enabling instance segmentation by providing both class labels and pixel-wise masks. Mask R-CNN is robust against the class imbalance problem and capable of the segmentation of many instances simultaneously in an image in different scales and orientations.<sup>[11]</sup>

The study hypotheses of this study were that Mask R-CNN could achieve a promising result in shoulder MRI with respect to its robustness in segmentation of discontinuous and noisy images and that the success rate of the segmentation is dependent on the tissue clarity; thus, with increasing edema and deformity, segmentation success is presumed to decrease. To our knowledge, this is the first study to evaluate the performance of Mask R-CNN on segmentation of the humeral head and scapula, concomitantly, from axial PD-weighted MRI slices.

### Related research

Li et al.<sup>[7]</sup> have developed an automatic needle tracking algorithm based on the Mask R-CNN for MRI-guided interventions. Boutillon et al.<sup>[8]</sup> introduced a framework that utilized an autoencoder and conditional adversarial network as regularizers. This framework was employed to train a convolutional encoder-decoder model using a dataset comprising 15 pediatric shoulder examinations. The predictions generated by the proposed framework exhibited agreement with the shape prior, while the adversarial score promoted realistic shape generation. The authors reported that their framework outperformed UNet and other recent derivatives in terms of performance.<sup>[8]</sup>

He et al.<sup>[10]</sup> introduced a recursive learning framework combined with a deep end-to-end network for the segmentation of 50 shoulder joint bones. This approach effectively handles challenges posed by small datasets with significant parameter variations. It not only enhanced the segmentation accuracy but also reduced errors and human artifacts in annotations.

Wu et al.<sup>[12]</sup> employed a framework consisting of two components: Mask R-CNN for segmenting hand bones from X-ray images and a residual attention subnet for generating the final prediction and visual supports. In another study, a convolutional neural network was developed for the segmentation and quantification of bone mineral density in the humerus using calibration phantoms and soft tissue correction.<sup>[13]</sup> The evaluation of the proposed system yielded an average Dice similarity coefficient value of 97.81% for the humeral bone density estimation based on a dataset of 210 X-ray images.<sup>[13]</sup>

Ryba and Krnoul<sup>[9]</sup> proposed a semiautomatic segmentation of MR images of the shoulder joint utilizing watershed-based region growing for coarse segmentation and active contour model for refining produced label. Although their method showed promising results for segmenting humerus on a two-dimensional MR image, the applicability to a three-dimensional problem was questionable.

### PATIENTS AND METHODS

In the study, 665 shoulder MR images of 665 shoulder instability patients (412 males, 253 females; mean age: 27±5.2 years; range, 18 to 42 years) who were admitted to the Şişli Etfal Training and Research Hospital between January 2011 and December 2014 were included. The MR images utilized in this study were acquired using a 1.5 Tesla scanner and were

of the PD-weighted sequence with a 4 mm slice thickness. The original dimensions of the DICOM images were 256×256 pixels.

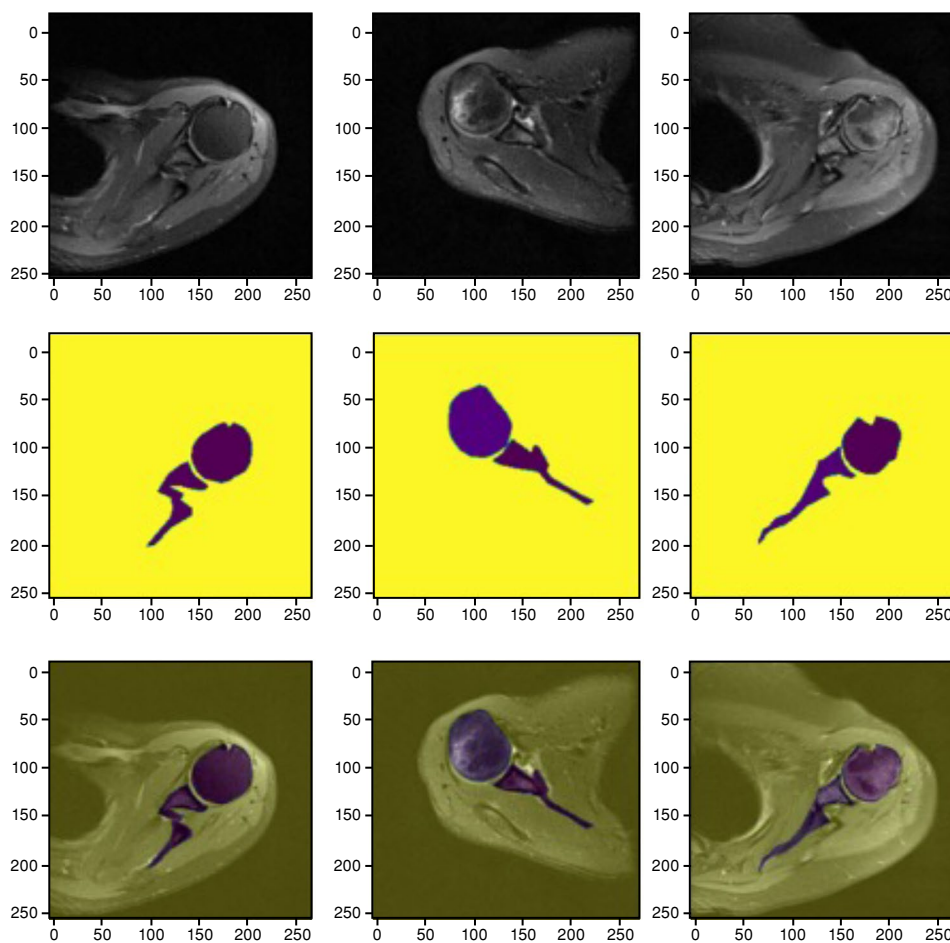
The dataset was divided into two: 446 were training data, and the remaining 219 were testing data. According to the humeral head structure three groups were generated: the healthy humeral head group, the edematous humeral head group, and the Hill-Sachs group (humeral heads with Hill-Sachs lesions). None of the scapular images showed glenoid bony defects, but 187 scapulae had discontinuity in the alar portion. In the scapular images, no glenoid bony defects were observed, as the cuts were positioned above the mid-glenoid axis.

Automatic segmentation was done based on Mask R-CNN. Simultaneous segmentation of multiple

anatomical regions (humerus and scapula) was done. For the sake of comparison, the manual segmentation using the LabelMe annotation tool<sup>[14]</sup> was carried out by an orthopedic surgeon with 20 years of experience, as shown in Figure 1. Statistical evaluation of the segmentation results was conducted using the Dice coefficient and mean average precision (mAP) score.

### Mask R- CNN

Mask R-CNN, belonging to the R-CNN series, is an innovative instance-based model developed by He et al.<sup>[15]</sup> in 2017. It combines instance segmentation and target detection, building upon the foundation of Faster R-CNN. In Mask R-CNN, the object detection component remains the same as in Faster R-CNN. However, by integrating the Feature Pyramid Network and Residual Network for feature extraction, it



**FIGURE 1.** Demonstration of original and manual segmentation of three MR images. Each column corresponds to a single image. The first row demonstrates the original PD-weighted axial MR image of each shoulder. The second row shows manual segmentation results by an expert. The third row demonstrates the visualization of manual segmentation result on the original MR images.

MR: Magnetic resonance; PD: Proton density.

TABLE I

Average segmentation results of humeral head and scapula based on Mask R-CNN by mAP with an intersection over union of 0.5 and the Dice metric

	Average success rate (with IoU of 0.5)	Average success rate (Dice metric)
	%	%
Mask R-CNN (Humeral head)	95.86	96.47
Mask R-CNN (Scapula)	92.35	93.87

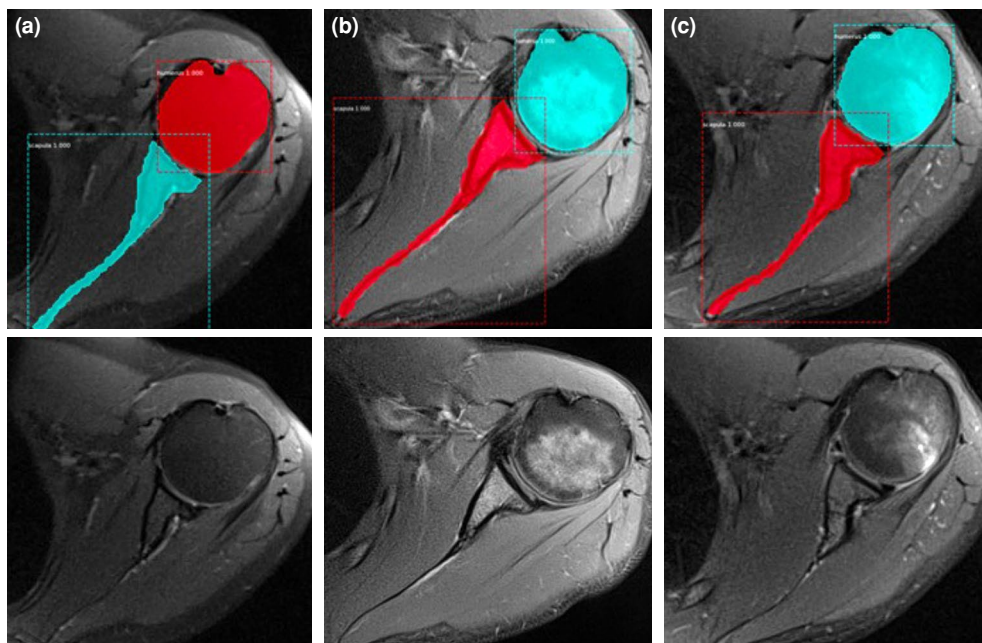
R-CNN: Region-Based Convolutional Neural Network; mAP: Mean average precision; IoU: Intersection over union.

effectively utilizes multiscale information. The image segmentation architecture is constructed as a Fully Convolutional Network, replacing fully connected layers to minimize additional computational overhead. Notably, the RoIAlign layer ensures precise instance localization through pixel-based alignment.

The main steps of Mask R-CNN are as follows. First, the input image is processed by Residual Network to extract features, generating multi-scale feature maps. Next, the side connections involve upsampling the feature maps of each stage and merging them with adjacent feature layers. The Region Proposal Network then generates proposal boxes on the feature maps of varying sizes. These proposal boxes, along with the feature maps, are fed into RoIAlign. Each proposal box corresponds to specific feature layers, and the

pooled results are utilized for classification and regression, obtaining adjustment parameters for the proposal boxes. Finally, the proposal boxes are refined to obtain prediction boxes, and segmentation masks for the detected objects are generated.

Unlike Faster R-CNN, which focuses solely on classification and bounding box recognition, Mask R-CNN adds an additional mask head for predicting an object mask for each region of interest. Consequently, Mask R-CNN provides an additional output as a binary mask in addition to class labels and bounding box offsets. The Mask R-CNN framework exhibits high efficiency and robustness against occlusion. It effectively addresses the class imbalance problem and enables the segmentation of multiple anatomical regions simultaneously.



**FIGURE 2.** The first row demonstrates the segmentation results of the humeral head and scapula based on Mask R-CNN. The second row displays the original images of (a) a healthy humeral head, (b) a humeral head with edema, and (c) a humeral head with Hill-Sachs lesions.

R-CNN: Region-based Convolutional Neural Networks.

TABLE II

Comparison of overall segmentation results of the humeral head and scapula based on Mask R-CNN with the Dice metric

Methods and datasets	Average success rates of segmentation (Dice metric)			
	Normal humeral head	Humeral head with edema	Humeral head with Hill-Sachs lesion	Average success rate
	%	%	%	%
Number of data	81	100	38	
Mask R-CNN (Humeral head)	97.42	96.53	95.48	96.47
Mask R-CNN (Scapula)	94.58	93.56	93.47	93.87

R-CNN: Region-Based Convolutional Neural Network.

In this study, Mask R-CNN was implemented using Keras and Tensorflow. The Mask R-CNN was trained on an NVIDIA V100 graphics processing unit (Nvidia Corp., Santa Clara, CA, USA) for 1,000 epochs using Adam optimizer at a learning rate of 0.001.

To assess the similarity between the predicted segmentation mask and the ground truth mask, Dice coefficient and mAP score were employed. The average precision score is defined as the area under the precision–recall curve, averaged for each object in an image at discrete recall levels. The precision–recall curves and average precision scores for all 219 images in the test data set were averaged. For PD-weighted MR images with successful instance detection, the intersection over union for 0.5 of the segmentation masks was calculated with respect to the manual segmentation of an expert for each case.

## RESULTS

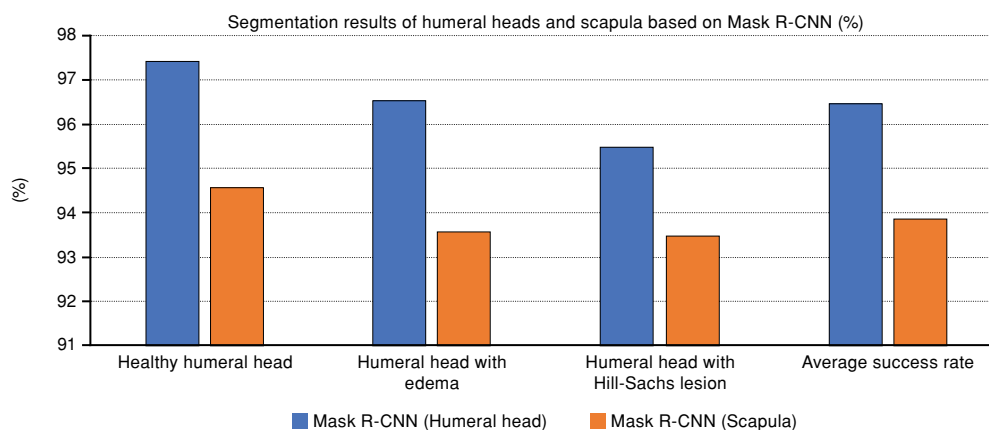
In the test images, 81 humeral heads were healthy, 100 were edematous, and 38 had a Hill-Sachs lesion. According to the Dice metric, the overall success rate

of Mask R-CNN on the humeral bone segmentation was 96.47 and 93.87% for the humeral head and scapula, respectively, and 95.86 and 92.35% for intersection over union of 0.5 according to mAP (Table I, Figure 2 and 3).

According to the Dice metric, the segmentation success of humerus and scapula in the healthy group was 97.42 and 94.58%, 96.53 and 93.56% in the edematous group, and 95.48 and 93.47% in the Hill-Sachs group, respectively (Table II). Scapular wing defect in PD-weighted MRI was detected by the expert in 187 cases. The segmentation success of scapula in segmental defects was 92.86% according to the Dice coefficient score (Figure 4).

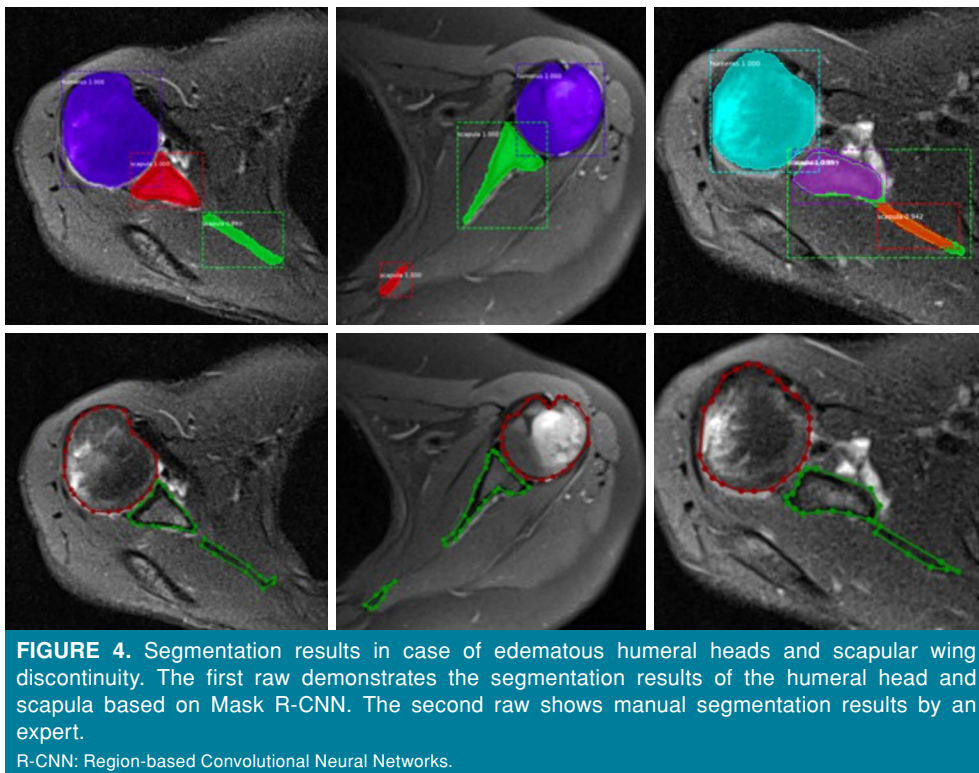
## DISCUSSION

This study confirms our first hypothesis. Mask R-CNN-based segmentation of the humeral head and scapula from PD-weighted shoulder MRI had high success. The overall segmentation success was very high (96.47% for the humeral head and 93.87% for the scapula according to the Dice metric). The best instances for segmentation were achieved in



**FIGURE 3.** Graphical representation of our segmentation results of humeral head and scapula based on Mask R-CNN with dice coefficient scores.

R-CNN: Region-based Convolutional Neural Networks.



the healthy group, with 97.42 and 94.58% for the humerus and scapula, respectively.

This study revealed a tendency of the segmentation success to decrease in an inverse proportion to the level of trauma to the bone tissue in shoulder MRI, as a proof of the second hypothesis. The minimal success rate was observed in the Hill-Sachs group, with 95.48 and 93.47% for the humerus and scapula, respectively (Figure 2). Humeral head segmentation had superior outcome values compared to scapula independent of the humeral head group, which might be a result of the discontinuity in the scapular wing.

Proton density-weighted MRI often suffer from challenges such as high noise, density inhomogeneity, anatomical structure variations, and pathological alterations.<sup>[1,2]</sup> These factors significantly complicate the segmentation and recognition of regions of interest. This might be the reason of the scarcity of the bone segmentation studies from shoulder MRI. To overcome these difficulties, various segmentation methods, noise reduction techniques, and image enhancement approaches were employed specifically for segmenting shoulder bones.<sup>[3,5,10]</sup> In this study, despite inherent complexities, successful segmentation of both the humerus and scapula was achieved using the Mask R-CNN method. Although

the scapula is anatomically more complex and represents a discontinuity problem and irregular borders, the segmentation results of scapula by the proposed system were highly satisfactory compared to the humerus.

Mask R-CNN is robust against multiple object segmentation, defect recognition, and irregularities in the contours.<sup>[7,14]</sup> Regarding the segmentation results of humeral heads based on classical methods, such as active contours without edges and signed pressure force, Mask R-CNN revealed higher results compared to the literature.<sup>[5]</sup>

Scapula segmentation studies are rarer in the literature.<sup>[10]</sup> Segmentation of scapula regarding its anatomical complexity is more challenging. Medial scapular irregularities and alar defects in PD-weighted images create discontinuities, thus decreasing the overall scapular segmentation success. In an attempt to reconstruct three-dimensional shoulder bone images from PD-weighted MRI of 50 shoulders, He et al.<sup>[10]</sup> tried to mitigate errors and minimize manual segmentation artifacts. Nevertheless, the best segmentation results they reported were 92.00 and 75.00% for the humerus and scapula, respectively. In our study, the best results were 97.42 and 94.58% for the humerus and scapula regions, respectively. The Mask R-CNN (93.87%)

results of our study surpassed the autoencoder (82.19%), UNet (79.21%), CAEUNet (80.52%), and cGAN-UNet (80.69%) results of another study for scapular segmentation in all groups.<sup>[8]</sup>

This study has some limitations. This study only included instability patients, which superimposes edema and bony defects in the images, diminishing automatic segmentation success. In addition, the amount of training and test data, although much higher than in previous studies, was still lower than the desired levels. A larger amount of training data would increase the segmentation results. Lastly, even more successful results would be achieved with increasing resolution. This study presented successful two-dimensional bone segmentation in the presence of shape and texture inhomogeneity. Furthermore, the results are promising for a three-dimensional reconstruction task that was not the subject of this study.

In conclusion, the application of Mask R-CNN for humerus and scapula segmentation yielded promising results. However, the humeral heads of some patients showed intensity inhomogeneity as edema and shape distortions, such as Hill-Sachs lesions, and some cases showed irregularities in the medial border and defects in the wing region of the scapula. Scapular segmentation with Mask R-CNN yielded high success, although the success rate was slightly lower than that of humeral head segmentation. The success rate of bone segmentation with Mask R-CNN is inversely proportional to the magnitude of the tissue trauma in PD-weighted shoulder MRI.

**Acknowledgements:** Special thanks to Hasan Basri Sezer, who provided annotation of MR images for the research.

**Ethics Committee Approval:** The study protocol was approved by the University of Health Sciences Hamidiye Etfal Training and Research Hospital Clinical Research Ethics Committee (date: 13.12.2016, no: 1343). The study was conducted in accordance with the principles of the Declaration of Helsinki.

**Data Sharing Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflict of Interest:** The author declared no conflicts of interest with respect to the authorship and/or publication of this article.

**Funding:** The author received no financial support for the research and/or authorship of this article.

## REFERENCES

1. Uluyardımcı E, Öçgüder DA, Bozkurt İ, Korkmazer S, Uğurlu M. All-suture anchors versus metal suture anchors

- in the arthroscopic treatment of traumatic anterior shoulder instability: A comparison of mid-term outcomes. *Jt Dis Relat Surg* 2021;32:101-7. doi: 10.5606/ehc.2021.75027.
2. Orhan Ö, Sezgin EA, Güngör İ, Çetinkaya M, Ataoğlu MB, Kanatlı U. Interscalene block applied by an experienced anesthesiologist has a good anesthetic effect, a long duration of action, and less postoperative pain after arthroscopic shoulder procedures independent of surgery type and operation duration. *Jt Dis Relat Surg* 2023;34:445-50. doi: 10.52312/jdrs.2023.1064.
3. Ryba T, Krnoul Z. Segmentation of Shoulder MRI Data for Musculoskeletal Model Adaptation. *Bioimaging* 2019; 33:155-60. doi: 10.5220/0007580701550160.
4. Wang H, Suh JW, Das SR, Pluta JB, Craige C, Yushkevich PA. Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Anal Mach Intell* 2013;35:611-23. doi: 10.1109/TPAMI.2012.143.
5. Sezer A, Sezer HB, Albayrak S. Segmentation of bone with region based active contour model in PD weighted MR images of shoulder. *Comput Math Methods Med* 2015;2015:754894. doi: 10.1155/2015/754894.
6. Taghizadeh E, Terrier A, Becce F, Farron A, Büchler P. Automated CT bone segmentation using statistical shape modelling and local template matching. *Comput Methods Biomech Biomed Engin* 2019;22:1303-10. doi: 10.1080/10255842.2019.1661391.
7. Li X, Young AS, Raman SS, Lu DS, Lee YH, Tsao TC, et al. Automatic needle tracking using Mask R-CNN for MRI-guided percutaneous interventions. *Int J Comput Assist Radiol Surg* 2020;15:1673-84. doi: 10.1007/s11548-020-02226-8.
8. Boutillon A, Borotikar BS, Burdin V, Conze PH. Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images. *ISBI* 2019;1164-7.
9. Siddique N, Sidike P, Elkin C, Devabhaktuni V. U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* 2021;1. doi: 10.1109/ACCESS.2021.3086020.
10. He X, Tan C, Tan V, Li K. Recursive 3D segmentation of shoulder joint with coarse-scanned MR Image. 2022.
11. Felfeliyan B, Hareendranathan A, Kuntze G, Jaremko JL, Ronsky JL. Improved-Mask R-CNN: Towards an accurate generic MSK MRI instance segmentation platform (data from the Osteoarthritis Initiative). *Comput Med Imaging Graph* 2022;97:102056. doi: 10.1016/j.compmedimag.2022.102056.
12. Wu E, Kong B, Wang X, Bai J, Lu Y, Gao F, et al. Residual attention-based network for hand bone age assessment. 2019 IEEE 16<sup>th</sup> International Symposium on Biomedical Imaging. 11 July 2019, Venice, Italy: IEEE; 2019.
13. Liu YC, Lin YC, Tsai PY, Iwata O, Chuang CC, Huang YH, Tsai YS, Sun YN. Convolutional neural network-based humerus segmentation and application to bone mineral density estimation from chest x-ray images of critical infants. *Diagnostics (Basel)* 2020;10:1028. doi: 10.3390/diagnostics10121028.
14. Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: A database and web-based tool for image annotation. *Int J Comput Vis* 2008;77:157-73. doi: 10.1007/s11263-007-0090-8.
15. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy; 2017. p. 2961-9. doi: 10.1109/ICCV.2017.322.