Capsule network-based classification of rotator cuff pathologies from MRI

Aysun Sezer, Hasan Basri Sezer

▶ **To cite this version:**

Aysun Sezer, Hasan Basri Sezer. Capsule network-based classification of rotator cuff pathologies from MRI. Computers and Electrical Engineering, 2019, 80, pp.106480 -. 10.1016/j.compeleceng.2019.106480 . hal-03488548

HAL Id: hal-03488548
https://hal.science/hal-03488548

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0045790618331240
Manuscript_3be783acc90371cba2ded1c0a669cfd4

Capsule Network-based Classification Of Rotator Cuff Pathologies From MRI ☆

Aysun Sezer[a,*], Hasan Basri Sezer[b]

[a] *Unité d'Informatique et d'Ingénierie des Systèmes, ENSTA-ParisTech, Université de Paris-Saclay, France*
[b] *Clinique du Sport Paris V, Paris, France*
University of Medical Sciences, Şişli Hamidiye Etfal Training and Research Hospital, Orthopaedics and Traumatology Clinic, Istanbul, Turkey

Abstract

Rotator cuff lesions are very frequent events. The diagnosis of these lesions is challenging and requires experience. The goal of this study is to develop a computer aided diagnosis (CAD) system based on Capsule Network (CapsNet) to classify rotator cuff lesions as normal, degenerated or torn in a new dataset of 1006 shoulder proton density (PD) weighted MRIs. Increasing the number of primary capsules and adding two cascaded convolution layers before capsule layer provided the CapsNet model to extract discriminative features for the better recognition of rotator cuff pathologies. The overall success rate of proposed Capsnet model was 94.75%, compared to custom designed CNN, AlexNet, GoogLenet and the gray level co-occurrence matrix (GLCM) which provided overall success rates of 93.21%, 88.45%, 87.63% and 85.20%, respectively. CapsNet performs better than CNNs on the augmented dataset as well, and robustly handles classification difficulties of rotator cuff pathologies from MRI.

Keywords: Capsule network, convolutional neural network, rotator cuff pathologies, PD weighted MRI, image classification

1. Introduction

1.1. Overview of Rotator Cuff Pathologies and Imaging Modalities

Shoulder pain is highly frequent. Although certain rotator cuff injuries may be clinically silent, they are the leading source of shoulder pain worldwide [1]. The incidence of rotator cuff wear or tear increases with age. The spectrum of these lesions ranges from degeneration to complete tears. Rotator cuff injuries are usually associated with both loss of function and pain [2].

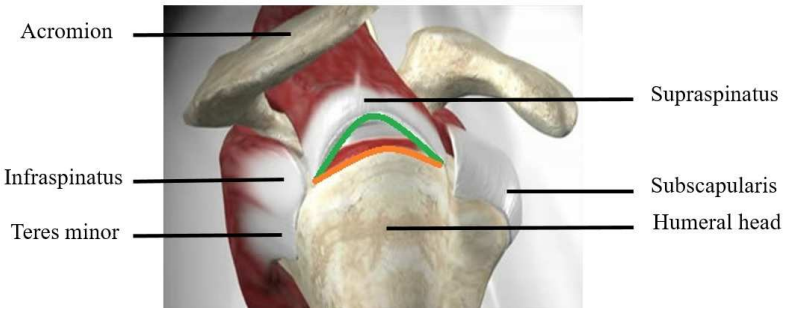
*Corresponding Author: Aysun Sezer; Unité d'Informatique et d'Ingénierie des Systèmes, ENSTA-ParisTech, Université de Paris-Saclay, France; aysun.sezer@ensta-paristech.fr

© 2019 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

The rotator cuff is a musculotendinous structure that originates from the scapula and attaches to the humeral head as a broad band. It is a highly important functional unit of the shoulder including 4 muscles (supraspinatus, infraspinatus, teres minor and subscapularis) that occupy the subacromial space (Fig.1). A well-functioning rotator cuff serves as the primary abductor and rotator of the shoulder joint as well as the stabilizer of the humerus against proximal static and dynamic migration.

Rotator cuff integrity is crucial for the balance of all the muscular components of the rotator cuff due to shared anchoring to the humeral head. The tendon of the supraspinatus muscle is the most frequently injured (Fig.1). Cuff insufficiency and pain are the main indications for surgical treatment. Clinicians have recently reported that the rotator cuff tends to retract following tears creating a wider gap between torn ends requiring more frequent surgical interventions. Moreover, loss of rotator cuff function creates impingement of the cuff itself between the humeral head and acromion resulting in a further risk of loss of integrity and cuff tear arthropathy [3].

There are numerous imaging techniques to assess rotator cuff integrity. X-ray is the first line imaging technique in clinical settings and provides valuable information on the skeletal anatomy of the shoulder joint. However, even the most suspicious X-ray findings are only an indirect sign of rotator cuff malfunction [4]. Ultrasound is a very useful technique for the evaluation of soft tissues such as the rotator cuff [5]. This inexpensive and extensively used approach has innate limitations such as operator dependency, anisotropy and unreliability in patients with highly abnormal echogenicity [1]. Computerized tomography (CT) is widely used for many musculoskeletal problems of shoulder. It provides precise visualization of bony structures and is potentially convertible to three dimensional images. CT is routinely used for fractures of shoulder bones, tumoral lesions of shoulder, evaluation of bony instability and sometimes for preopera-



(a)

Figure 1: a) Demonstration of subacromial space on an anatomical diagram. Rotator cuff muscles are located between the humerus and acromion. The orange line delineates the original attachment site of supraspinatus tendon. The green line stands for torn and retracted supraspinatus tendon.

tive templating before shoulder arthroplasty [6]. Although CT has been shown to be successful in evaluating bone, results in soft tissues like rotator cuff are less good. The radiation exposure and mis-diagnosis risk of computerized tomography decreased its popularity for the rotator cuff pathologies. Contrast enhanced CT and MR arthrography provide better visualization of rotator cuff tears, however special expertise is required for these interventional protocols, which are not possible in all facilities. There here are also some concerns about patient acceptance due to pain during interventional techniques.

MRI is a reliable technique that can visualize all the components and dimensions of the shoulder. High resolution MR imaging is useful for the preoperative assessment of the rotator cuff [3]. An intact and well-functioning cuff is almost smooth on MR imaging while tears are seen as high intensity areas on proton-density (PD) weighted sequences. MR imaging is a sensitive and specific technique that defines the size of defects, retraction, tear shape, and muscle status [7]. It also provides very important information when deciding on rotator cuff tear repair in relation to the location and size of the tear as well as the quality of injured tissue [3].

There are more patients with asymptomatic partial tears and degeneration in the population than expected. In symptomatic patients with partial tears or degeneration a clinical examination and imaging may not be enough for discrimination. Although MR imaging is the technique of choice for rotator cuff injuries, it is not always easy differentiate partial tears from degeneration in MRI, as well [8]. Degeneration and partial thickness tears are seen as inhomogeneous gray areas on MR imaging, which are very similar visually. In controversial cases CAD systems may be a useful, effective, consistent and quantitative decision-making system for less experienced evaluators [1].

In this study an automatic recognition system was designed for rotator cuff injuries using MRI to decrease evaluator workload and observer dependency. This tool can help radiologists and orthopedists obtain a more precise diagnosis of rotator cuff tears, which can influence patient treatment options. This CAD system can also be used in epidemiological research for rotator cuff pathologies and help public health experts more thoroughly define the social and economic impact of these lesions.

1.2. Related Work

Computer based classification of medical images is mainly based on the recognition of distinctive imaging features. Various shape and texture features have been extracted by various feature extraction methods such as histograms oriented gradient, local binary pattern and the gray level co-occurrence matrix (GLCM). Although combinations of these features have been used to increase the efficacy of medical image classification in the literature, this has not reliably improved performance because it may increase the so-called "curse of dimensionality". Consolidating too many scalar and source-generated features increases the combinatorial complexity of classification and may even cause suboptimal results. Zheng, H. et al combined multiple features by multiple kernel learning.

This approach increased the performance of handcrafted classification methods but increased the complexity of the system.

An alternative approach to artificial intelligence based image processing is an imitation of natural inherited methods of perception and analysis. Deep CNN, a popular machine-learning approach [9], was designed to mimic the mechanism of the visual cortex of vertebrates. A typical CNN recognizes the texture and shape characteristics of an image by processing them through several layers of virtual neurons. By changing the hierarchical organization or the number of layers of CNNs, the system can be modified to reach the desired goal. Shallow CNNs are used to recognize basic structures such as lines and edges. More complex structures can be learned by increasing the depth of a CNN. Several studies have investigated the accuracy of convolutional network depth for general image classification and obtained significant improvement by increasing network depth [10]. However, as network depth increases, accuracy saturates and then rapidly degrades.

CNNs are high capacity networks with complex layers of working mechanisms used to extract unchanging features such as translation, rotation or scale. The primary requirement of any CNN, whatever the organization, is the need for massive training data that is not easily obtained in the field of medical image analysis. Another limitation is the loss of information due to local details such as location and pose during CNN image processing. Therefore, CNNs are not optimal for extracting diverse features such as semantic class, orientation, location or scale, which are needed for accurate image reconstruction.

A study by Hinton et al. attempted to overcome the limitations of CNNs. They depicted human visual perception as the deconstruction of an image in the brain to simulate its hierarchical representation [11]. They assumed that the representation of an image in the human brain depends upon hierarchical pose (translation and rotation) relationships rather than angular information that must be preserved during image processing. Therefore, the capsules of this CapsNet method were designed to represent output information in terms of vectors rather than single scalar values. In this way the model could be forced to learn the feature variants of an image. Thus, potential variants could be more effectively extrapolated by the model with less training data [11, 12].

Deep learning requires relatively big datasets for training which, in fact, is not an easy task to overcome in medical field. Data augmentation is one of the alternative techniques to reduce over fitting, in case of inadequate dataset. There are numerous data augmentation approaches defined for different image classification problems, such as translation, rotation, mirroring or scaling of target and adding noise [13]. There are also alternative data augmentation techniques, published in the literature. DeVries et al. [14] proposed to use an autoencoder for the purpose of data augmentation. They concluded that their proposed technique based on transformation was more successful in a learned feature space than in input space. Ding et al. proposed data augmentation in synthetic aperture radar (SAR) images by increasing speckle noise [13]. Fawaz et al. utilized dynamic time warping distance for data augmentation using synthetic data with deep residual networks [15]. Kooi et al. applied scale and

translation transformation to mammography images using the center information of the lesions [16]. Wang et al. studied the performance of CycleGAN for data augmentation and concluded that generative adversarial networks were less energy efficient and less effective than traditional techniques [17].

At present there are a few studies analyzing medical images for the classification and segmentation of rotator cuff lesions. Shoulder ultrasound is popular due to its extensive use in the medical field. Horng et al. classified ultrasound images of cuff lesions using multiclass fuzzy support vector machines and firefly algorithms with texture analysis and GLCM and texture feature coding in a population of 75 patients divided into 4 classes with a success rate of 92.5% [18]. Chao et al. extracted texture features based on GLCM, texture spectrum, fractal dimension and texture feature coding to classify 80 rotator cuff ultrasound images. Firefly radial basis function was used to classify rotator cuffs into four classes with a success rate of 93.75% [19]. Chang et al classified rotator cuff lesions as tears, inflammation and calcific tendinitis based on GLCM with a success rate of 87%. Park et al. have proposed a computer aided diagnosis system based on texture analysis with a histogram, GLCM and grey level run length matrix and classified rotator cuff lesions from ultrasound images of 40 patients [20]. Gupta et al. reported a segmentation study of ultrasound images of supraspinatus tendon based on curvelet transform, and they ignored the classification of cuff lesions [21].

Although ultrasound is extensively used and less expensive for the visualization of the rotator cuff, MRI provides clearer and anatomically exact images. Kang et al. studied texture analysis of rotator cuff tears by T2-weighted magnetic resonance arthrography in 50 patients based on GLCM [22]. The limited data with arthrography may be due to the invasiveness of this approach. Our study used PD weighted coronal shoulder MRI which is a routine and easily accessible sequence.

Although CNNs are powerful tools for the representation of medical images, and there are numerous studies in the medical field using deep CNNs, there are very few studies in the literature on musculoskeletal imaging. Deep CNNs have been used for the segmentation and localization of human vertebrae on CT and MRI. Roth et al. used deep CNNs in the detection of sclerotic spinal metastases in spinal CT images of 59 patients with a success rate of 83% [23]. Geng et al. studied bone tumors on scintigraphy with a success rate of 88% based on a dice metric [24]. CNNs were also used to assess age from hand X-rays [25] imaging. CNNs were applied to diagnose brain related diseases from MR images [26]. Our search of the literature did not identify any studies of deep CNNs for the classification of rotator cuff lesions on MR imaging.

At present there are only a few studies using CapsNet in medical image analysis, probably because this is a very recent technique in the computer vision field. Qiao et al. reported highly positive results with CapsNet for the reconstruction of image stimuli in human fMRI. In a study by Afshar et al. CapsNet was found to be highly effective in the recognition of brain tumors from human brain MRI [27]. Mobiny et al. used CapsNet for the detection of lung nodules on CT. CapsNet was found to be significantly better than CNNs

when the number of training samples is small [28]. Rodney et al. proposed SegCaps to segment pathological lungs from low dose CT scans and reported a dice metric of 98.47% [29].

In this study a CapsNet model was proposed to classify rotator cuff lesions from coronal PD-weighted MRI images of the shoulder.

This study is novel for the following reasons:

- To the best of our knowledge there are no studies in the literature evaluating deep learning for the diagnosis of rotator cuff lesions on MRI. This study describes a CapsNet based CAD system to classify rotator cuff lesions and compares the results with a custom-designed CNN, pre-trained CNNs and GLCM.
- A new dataset including coronal PD weighted MR imaging of the shoulder in 1006 patients were collected for this study. This is a significantly larger dataset than those found in other studies in the literature.
- The region of interest was automatically determined to feed our proposed CapsNet to classify rotator cuff images as normal, degenerated or torn. The results of our system are highly promising compared to hand-labeling by an experienced orthopedist.

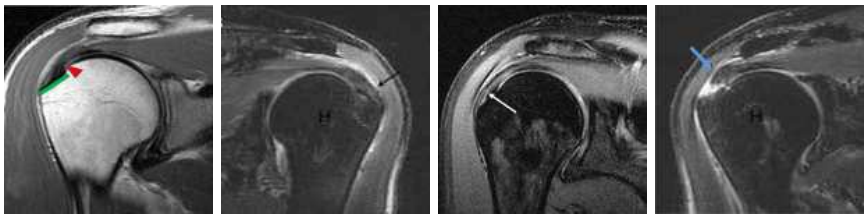
The remainder of paper is organized as follows: Section II describes the dataset and proposed method in detail. Experimental result and discussion are given in Section III and IV.

2. Materials and Methods

MRI from 1006 patients with shoulder pain admitted to an outpatient clinic were included in this study. All shoulder images were obtained from the Sisli Hamidiye Etfal Training and Research Hospital on a 1,5 Tesla Picker MR Unit (Philips Medical Systems Eindhoven, Netherlands) and a shoulder coil was used during investigations. Coronal PD weighted MR images of the shoulder were selected to analyze the condition of the rotator cuffs and labeled by an experienced orthopedic surgeon as 316 normal and 311 degenerated and 379 torn rotator cuffs.

Coronal PD weighted MR images of the shoulder were selected to detect rotator cuff lesions as this view is preferred by clinicians. Coronal images are useful to clinicians because they provide sections of the longitudinal axis of the cuff muscles. Mid-humeral coronal sections were selected to include the supraspinatus, which is the most frequently affected muscle in cuff injuries.

Although T1-weighted MRI provides better contrast between the humeral attachment of the rotator cuff and the humerus, PD weighted sequences provide more anatomical detail. Identification of cuff injuries is also better on PD weighted images showing clear changes in intensity in the presence of a tear. In the presence of a cuff tear the space that should be occupied by the cuff is filled with joint fluid, which is seen as a high intensity white image on PD



(a) (b) (c) (d)

Figure 2: a) Coronal T1 MR image of shoulder. Green line stands for rotator cuff footprint on humerus. The red arrow stands for supraspinatus tendon. b) PD weighted MR image of shoulder. Black arrow indicates supraspinatus tendon. The transition zone between humerus and supraspinatus tendon may be compared between a and b. c) PD weighted MR image demonstrating supraspinatus degeneration and edema. d) Tear of supraspinatus detected in PD weighted shoulder MR (blue arrow).

(a)

Figure 3: a)PD weighted coronal MR images of the shoulder. For the sake of demonstration of differences in each group of patients three examples belonging to the same group were given in a row. The three images in the first row represent normal rotator cuff, second row contains degenerated cuff images and the third row includes shoulders with rotator cuff tear. In each of the nine images the surrounding soft tissues and bones reflect multiple shape and texture variances in forms of arthrosis or edema. Additionally as may be confronted from the MR image of left shoulder in the second column of the first row there is a rotational variance from the other images which are images of right shoulders.

weighted MRI. On the other hand, in the presence of degeneration or a partial tear, effusion around the ligament is seen as patchy area, with uneven intensity making visual discrimination difficult (Fig.2).

The anatomical, pathological and rotational differences in the MR images

of individual patients must be considered when constructing a CAD system. These differences may be significant, even in patients in the same group. The presence of right and left shoulders in the dataset is another important factor that may require geometric translation (Fig.3). Therefore, a CAD system for the classification of the rotator cuff on MR imaging must successfully interpret different poses, views and orientations.

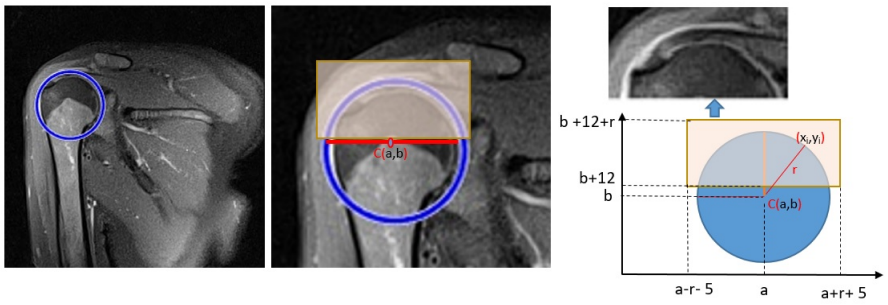
2.1. Automatic determination of Region of Interest

The he subacromial space was selected as the region of interest (ROI) (Fig.1). The first area to be identified when an evaluator analyzes coronal PD-weighted MRI of the shoulder is the subacromial space where the rotator cuff is located. This 1 to 2- centimeter wide space between the humerus and the acromion is covered in bony structures and includes nearby tissues with similar characteristics. The rotator cuff muscles share the subacromial space with other anatomical structures such as the subacromial bursa. Unfortunately, the shape and volume of the subacromial space are not standard due to anatomical differences in bony structures such as the acromion, and narrowing of the space in case of acromioclavicular or glenohumeral arthritis and cuff tears (Fig.3).

Therefore, to identify the region of interest the Circular Hough Transform (CHT), an extensively used pattern recognition and computer vision technique for the detection of regular curves, was used. CHT can detect circular shapes based on a parametric equation of circles [25]. In our study CHT located the humeral head in all cases on coronal PD weighted MR images of the shoulder even in the presence of image noise. A demonstration of humeral head detection on a PD weighted MR image is shown in Fig.4.

Our goal was to define a ROI that included both the rotator cuff muscles and the surrounding bony structures, in particular the superior part of humerus where the cuff muscles are attached, the superior corner of the glenoid and the bony roof of the subacromial space including the acromion and distal clavicula. An optimal ROI should include the rotator cuff from the point of insertion on the humerus to the glenoid, which is found between these anatomical structures.The ROI was determined by using the humerus as the central point. It was defined automatically using parameters of the center and radius obtained by CHT.

A rectangular ROI window was initially placed on the center of the humeral head whose inferior border was situated on the radius of the circle determined by CHT. However, the initial ROI did not include the subacromial space. For an optimal ROI it was shifted slightly upwards from the center of the humeral head along the y axis. After confirming that the inferior border of the ROI remained just below the superior margin of the glenoid in all images, the lower margin of the ROI was placed 12 pixels above the center of the humeral head. The ROI area was also enlarged to include the distal attachment and the proximal orientation of the rotator cuff muscles (Fig.4) Finally, a ROI was constructed with a short edge equal to the radius of the circle and a long edge that was twice the radius of the circle but elongated by 5 pixels on each side. The ROI was constructed instead of using a complicated segmentation algorithm for the rotator cuff. Automatic determination of the ROI made it possible to locate

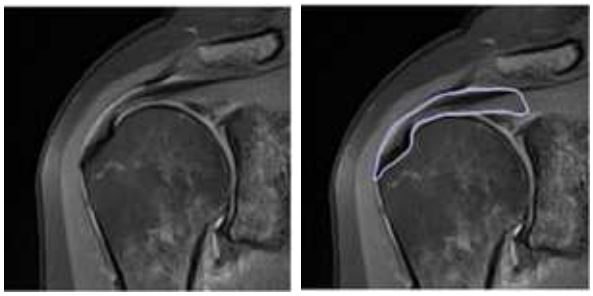


(a)

Figure 4: a) Detection of humeral head with Circular Hough Transform and determination of ROI by using the center information gathered from Circular Hough Transform.

the rotator cuff with the fewest possible unnecessary components and with less noise. Each ROI was re-sized to 64x64 to feed the capsule network and CNNs. The use of a ROI also decreased the area of interest by down-sampling the MR images from 256 X 256 to 64 X 64 pixels. As stated in the original CapsNet paper by Hinton et al., capsules tend to model everything in the input image [11]. Thus, decreasing the background area also decreased unnecessary fields, avoided complexity and provided a precise determination of the rotator cuff region.

Rotator cuff areas were determined by an expert in all images to obtain control data to compare our classification results. One of the manual segmentations by the expert on the screen was demonstrated in Fig.5.



(a) (b)

Figure 5: a) PD weighted MR image of right shoulder. b) Manual segmentation result of supraspinatus tendon by the expert.

2.2. Capsule Network

Hinton et al. changed the point of view in image analysis from invariance to equivariance while still taking into consideration invariance [11, 12]. Invariance and equivariance are two major components of visual representation. Invariance

is designed for a specific task while equivariance is designed for various tasks such as location, pose or orientation. In other words, equivariance keeps nearly all the information needed to represent an image.

The CapsNet model is one of the newest research topics in the literature. The basic unit of CapsNet is a capsule, which contains a group of organized neurons. The length of the capsule depends upon invariance, while the number of features to reconstruct the image are measurements of equivariance. Capsules produce vectors of the same magnitude but with different orientations. The orientation of a vector represents its parameters, that is, the information of the properties preserved from an image.

While a regular neural network needs additional layers to increase detail and accuracy, with CapsNet, a single layer can nest other layers. The capsules effectively represent different types of visual information, so called instantiation parameters, such as pose, which is a combination of position, size and orientation. The output of a capsule is a vector, which can be sent to a layer above to match its appropriate parent. The output of a capsule i was considered to be u_i , and transformation matrix W_{ij} was applied to capsule output to calculate the prediction of parent capsules j by transforming u_i to the prediction vector $\hat{U}j|i$.

$$\hat{U}j|i = W_{ij}u_i \tag{1}$$

where $\hat{U}j|i$ is the prediction vector of the output of the j_{th} capsule in a higher level computed by capsule i in the layer below, and W_{ij} is the weighting matrix that needs to be learned in the backward pass. The parameter s_j is a weighted sum over all prediction vectors $u_{j|i}$ where c_{ij} are the coupling coefficients calculated by the dynamic routing process to determine the degree of conformation between the capsules in the layer below and the parent capsules. Dynamic routing is the process of producing parent capsules by coupling the capsules via a "routing by agreement" approach. This relationship is not implemented by "max pooling" of standard CNNs. Unlike in max pooling, all the necessary information from details is preserved, thus increasing success by overlapping images. Dimensionality of the capsules also increases as the hierarchy ascends.

An activation function called squashing shrinks the final output vectors to zero if they are small and to unit vectors if they are large, to produce capsule lengths. Activity vector v_j is calculated with the following non linear squashing function.

$$v_j = \frac{||sj||^2}{1 + sj||^2} \frac{sj}{||sj||} \tag{2}$$

The c_{ij} is computed as the softmax of b_{ij}. The coupling coefficient is defined as the degree of conformation between the capsule and the parent capsule.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k exp(b_{ik})} \tag{3}$$

b_{ij} is the similarity score that takes into account both likeliness and feature properties rather than just likeliness in neurons.

$$b_{ij} = b_{ij} + \hat{U}_{j|i} v_j \tag{4}$$

2.3. Our Proposed Capsule Network

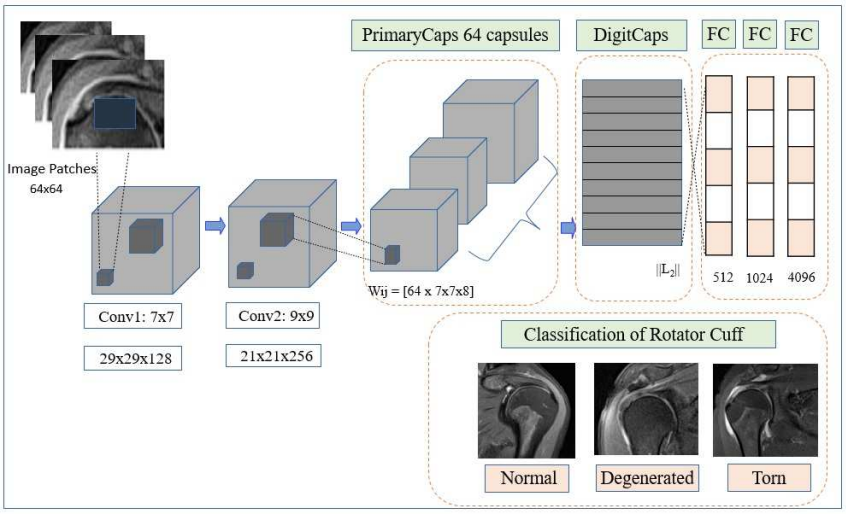
CapsNet offers a different approach from other computer vision methods. Indeed, existing computer aided image classification systems based on deep learning and/or hand designed methods cannot adequately handle the problems generated by local and global rotations in images obtained in different orientations. Rotation information plays an important role in the successful classification of many medical images, including rotator cuff lesions. In our study right or left shoulders could be handled as the mirror image of each other. Moreover, besides the positioning of the patient in the MR imaging machine, anatomical and pathological variances may also have affect the rotation of the arm in the image.

Table 1: Accuracy rates of different Capsule network models for classification of rotator cuff pathologies.

Models	Accuracy rate (%)
MNIST Model Baseline (256 feature maps + 32 Primary caps)	90.50
One Convolution layer with 512 feature maps + 32 Primary capsules of dimension 6, number of routing is 3	91.77
One Convolution layer with 256 feature maps + 64 Primary capsules of dimension 11, number of routing is 3	92.86
Two Convolution Layers with 128 and 256 feature maps + 64 Primary capsules of dimension 9, number of routing is 5	94.75

Automatically determined image patches were given to the original capsule network architecture designed for the MNIST dataset, which has one layer of convolution, two layers of capsules and three fully connected layers to classify rotator cuff MR images. In this study different CapsNet models were applied to evaluate the classification of rotator cuff pathologies (Table 1). Two strategies were used to increase the accuracy of the system. First, the number of primary capsules was increased. In this way, the learning of discriminative features was improved. Additionally, two convolution layers were cascaded before the capsule layer, which could also improve accuracy by creating more complex image coding before feeding it into the capsule layer (Fig.6).

Two cascading convolution layers were applied and each of the image patches was convolved with different sized kernels to obtain feature maps. One hundred



(a)

Figure 6: a) Our proposed Capsule network model for classification of rotator cuff lesions.

twenty-eight size 7 x 7 feature maps, were generated by the first convolution layer with a stride (S) size of 2, with no padding (P) and the spatial dimension was reduced to 29x29. Output of the first convolution layer was given to the second convolution layer as input to extract useful low level features. At this stage 256 size 9 x 9 feature maps were obtained with the parameters S=1. The next layer is a primary capsule layer, which generated 64 8-D capsules using 9 x 9 kernels with S=4. Each primary capsule contains 6 convolution units with a 11 x 11 kernel. 64 x 5 x 5 capsules were obtained and fed into digitcaps.

The final capsule layer included 3 capsules, referred as "Class Capsules", one for each type of rotator cuff. The dimension of these capsules was 16. The decoder element included three fully connected layers with 512, 1024 and 4096 neurons, respectively. The number of neurons in the last fully connected layer was the same as the number of pixels in the input image, because the goal was to minimize the sum of squared differences between input and reconstructed images. The success rate of the optimized CapsNet model was 94.75%, which was higher than the original MNIST CapsNet.

2.4. Comparative Methods

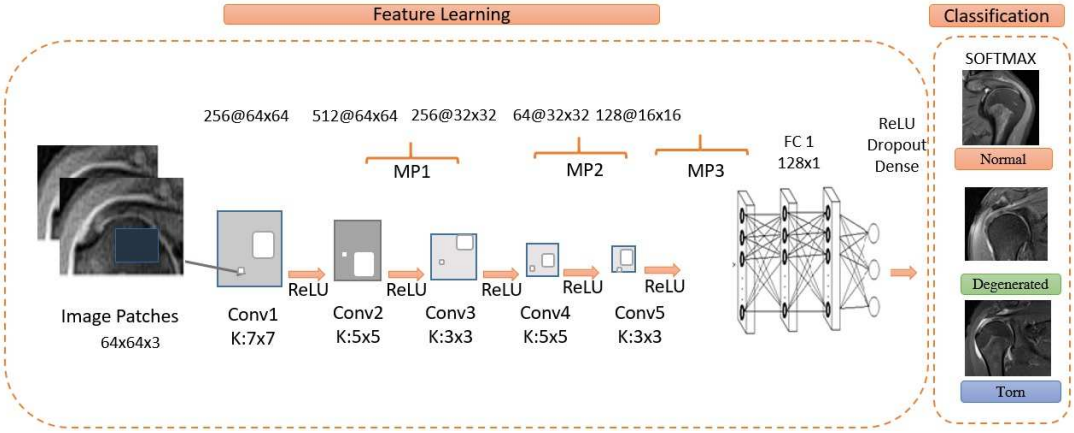
The success of our CapsNet system was compared to that of CNN for the discrimination of rotator cuff lesions. The progress of CNNs in the image recognition field has been remarkable. The success rate in different types of image datasets including in medical images is high. The success rate of a CNN for a specific task depends upon the optimization of its architecture. The depth and breadth of CNN architecture affects extractions of differentiating texture

and shape features. For this reason, different CNN models were tested for the classification of rotator cuff lesions.

The most successful CNN model included five convolutional and three max pooling layers in the feature learning step. Size 64 x 64 image patches were used in the first convolutional layer and 256 size 64x64 feature maps were generated. A rectified linear unit (ReLU) which replaces the negative value of feature maps with zero next to the first convolutional layer, was placed. The second convolutional layer with a kernel size of 5x5 was placed and 512 size 64x64 feature maps were generated. The pooling operation was employed after two cycles of convolution and activation function (ReLU). Max-pooling was applied to feature maps with a stride size of two pixels. The third and fourth convolutional layers generated 256 and 64 size 32x32 feature maps, respectively. The last convolutional layer was used to obtain 128 size 16x16 feature maps. Finally, a fully connected layer with 128 hidden units with max-out activation function was stacked to label each image patch (Fig.7). Exponential linear unit (ELU) was introduced to the computer vision field with the capability to speed up learning in CNNs with higher classification accuracy levels. ELUs alleviate vanishing gradient problem of deep architectures by pushing mean unit activations closer to zero via utilization of both positive and negative values. By this way gradient values get closer to the unit natural gradient, in other words the bias shift effect decreases and the learning speed increases. Unlike leaky ReLU and parametrized ReLUs which have negative values, as well, ELUs ensure a noise robust deactivation state. ELUs have the tendency to model the degree of presence of particular object while not coding its absence. This one-sided saturation property of ELUs was stated to lead to a better convergence [30]. For sake of comparison, we trained our dataset with the same proposed network model with ELU. Our goal was to obtain the probability of three groups of rotator cuffs, normal, degenerated and torn, from the softmax function of the classification step of our proposed CNN.

Transfer learning was another comparative method, used for the classification of rotator cuff pathologies. Transfer learning performs well with small training datasets, which is a common situation in medical imaging. In our study collecting and labeling a large number of rotator cuff images was time consuming. AlexNet [31] and GoogLenet [32] have been shown to provide successful results if the training and architecture hyper-parameters were optimally tuned.

GLCM has been routinely used in the classification of ultrasound images of rotator cuff lesions in literature. GLCM calculates the offset values of an image in a given displacement (d:distance) and orientation (θ :angle). Different offset values will result in changing second order GLCM matrices and different co-occurrence distributions causing different image features for the same reference image. In our study Haralick features were extracted for different offset values. The best result was obtained for $d = 4$ and $\theta : 0^0, 45^0, 90^0, 135^0$.



(a)

Figure 7: a) Details of proposed CNN model (Conv: Convolutional Layer,MP: Max pooling layer, FC: Fully connected layer)

3. Experimental Results

In this study the subacromial space, which includes the rotator cuff, was automatically and precisely defined by CHT in the entire dataset. The center and radius parameters obtained from CHT were used to determine images patches.

A 10-fold cross-validation method was used to check the stability and reliability of the proposed CapsNet. The dataset was split into two, as 70% training and 30% test sets. Random sampling was performed ten times to produce different training and testing sets from the data-set and decrease selection-bias. Our implementation was in TensorFlow and we used the Adam optimizer [33] with learning rate of 0.001 for 50 full epoches.

The success rate of our classification system was 90.50% with the original baseline architecture of CapsNet designed for the MNIST dataset. Certain components of the baseline MNIST CapsNet model were changed to extract discriminative features. Increasing the number of primary capsules and convolution layers before the capsule layer improved classification. When one convolution layer and 512 feature maps were used and the number of primary capsules was increased from 32 to 64, the overall accuracy of our system increased to 92.24% from 91.77%. Table 2 shows the confusion matrix of our proposed fully automated system to classify rotator cuff lesions as normal, degenerated or torn. The overall accuracy of our system was 94.75%.

The relevance of a CAD system can be measured by positive predictive (precision) and recall (sensitivity) values and F1-scores. The precision of a system is the rate of relevant instances among retrieved instances. The recall value of a system is the rate of relevant instances over the total amount of relevant values. The F1 score is the harmonic mean of the precision and recall values. The overall recall rates of our proposed CAD system were 95%. However,

the recall values of the cuff tear and degeneration groups were 96% and 97%, respectively, which confirms the power of the proposed system for the recognition of pathological rotator cuffs (Table 2).

Table 2: The confusion matrix of overall accuracy for our CapsNet model with 10 iterations (automatically determined ROI)

	Confusion Matrix For Our Proposed CapsNet				Classification result of rotator cuff lessions by our proposed CapsNet		
	Total Number of Data	Normal	Degeneration	Cuff Tear	Precision	Recall	F1-Score
Normal	316	88.60	1.91	9.49	0.93	0.89	0.91
Degeneration	311	0.32	96.46	3.22	0.94	0.97	0.95
Cuff Tear	379	2.12	1.84	96.04	0.95	0.96	0.96

Table 3: The confusion matrix of overall accuracy for our CapsNet model with 10 iterations (manually segmented images)

	Confusion Matrix For Our Proposed CapsNet				Classification result of rotator cuff lessions by our proposed CapsNet		
	Total Number of Data	Normal	Degeneration	Cuff Tear	Precision	Recall	F1-Score
Normal	316	85.32	2.21	11.45	0.92	0.81	0.86
Degeneration	311	1.92	90.67	7.34	0.94	0.91	0.92
Cuff Tear	379	1.85	4.22	93.93	0.9	0.94	0.92

The purpose of CAD systems is to help clinicians obtain a correct diagnosis, in particular by eliminating the underestimation of disease. A low false negativity rate is a measure of reliability and is a prerequisite for the use of CAD systems in the clinical setting. Our system labeled 1 degenerated cuff and 8 torn cuffs as normal in the entire dataset. Nine out of 690 patients were misclassified as normal. The rate of diseased cuffs misclassified as normal was 1.34% with our system (Table 2). These promising results suggests that this system can be used in clinical settings as a powerful diagnosis tool.

To validate our fully automated rotator cuff classification system, the success of our system in manually segmented shoulder MR images was evaluated. Table 3 demonstrates the confusion matrix of our CapsNet method for classifying rotator cuff lesions as normal, degenerated or torn in manually segmented images. The overall accuracy rate of our CapsNet system in the manually segmented dataset was 92.07%. Our CapsNet correctly labeled 356 out of 379 rotator cuff

Table 4: Overall accuracy rate of different methods for classification of rotator cuff lesions

Methods For Diagnosis of Rotator Cuff Lesion	Overall Accuracy Rate of the system (%)
CNN Model with original dataset	93.21
CNN Model with augmented dataset by transformation	94.42
CNN Model with augmented dataset by adding noise	94.87
CapsNet Model with original dataset	94.75
CapsNet Model with augmented dataset by transformation	95.36
CapsNet Model with augmented dataset by adding noise	95.48
AlexLeNet	88.45
GoogleLeNet	87.63
Gray Level Co-occurrence Matrix (GLCM)	85.20

tears (93.93% CI:95%) in the manually segmented dataset. On the other hand, 364 out of 379 rotator cuff tears (94.75% CI:95%) were correctly labeled by our system when the ROI was determined automatically. Thus, the performance of our proposed system was better in automatically determined ROI patches than in rotator cuff areas that were selected by an expert. This might be explained by the preservation of the spatial relationship of bones and the changes in the overall structure of the subacromial space in case of cuff rupture in automatically determined ROI.

The accuracy rates of our system with GoogLeNet and AlexNet pre-trained CNNs were also compared. GoogLeNet and AlexNet can be good solutions in the small training datasets that are common in the classification of medical images. However, there is no consensus in the literature for the optimization of hyperparameters and to determine the optimal learning rate of AlexNet and GoogLeNet for different layers. There are numerous architectural and training hyperparameters whose relevance has not been clearly defined. The approach by Sharif et al. using very small learning rates compared to default was applied [34]. After careful parameter tuning, the most successful model in this situation was obtained. As shown in Table 4, the accuracy rates of GoogLeNet and AlexNet were 88.45% and 87.63%, respectively.

The incremental gradient descent is the mostly used method for training a deep learning network. It is an iterative method to minimize the given objective function using batch samples. Optimum size of mini batches is very important to decrease memory requirements and increase the training speed [35]. However, determination of the optimum mini batch size is still accomplished by a trial and error approach. Proposed CNN model had 3 hidden layers of 128 units each. It was trained 300 epochs by stochastic gradient descent with mini batches of size 32 and learning rate of 0.01 by non-linear activation function of ReLU and ELU. In order to prevent over-fitting and reduce error we added dropout only to the fully connected layers with the probability of 0.02.

The classification accuracy of CNN were 92.50% and 93.21% with ReLU and

ELU, respectively. The overall success rate of our CNN was 93.21%, which was better than that of GoogLeNet and AlexNet for identifying rotator cuff lesions. One possible explanation for this may be that AlexNet and GoogLeNet were pre-trained on natural images, which are completely different from shoulder MR images.

Increasing amount of training data is presumed to have a positive effect on the success rate of the CNN. Though it may be achieved with some data multiplication protocols, such as rotation, mirroring, and scaling. In this study data augmentation by scale and translation transformation were used to overcome the effects of data scarcity and over-fitting. The overall success rate was increased by using augmented data by transformation to 94.42% from 93.21% and 95.36% from 94.75 for CNN and CapsNet, respectively (Table 4).

The effectiveness of data augmentation in rotator cuff image classification was assessed by increasing noise, as reported by Ding et al. [13]. The overall accuracy rates of the system were 94.87% and 95.48% for the augmented dataset based on adding noise for CNN and CapsNet, respectively. Data augmentation prevented over-fitting and increased the success rate to a degree for two proposed models. The results demonstrate that CapsNet is more successful than CNN in all cases. The application of transformation-based augmentation protocols may result in pose and orientation problems of the anatomical relationships, however, CapsNet seems to be robust to transformation and require far less training data compared to CNN.

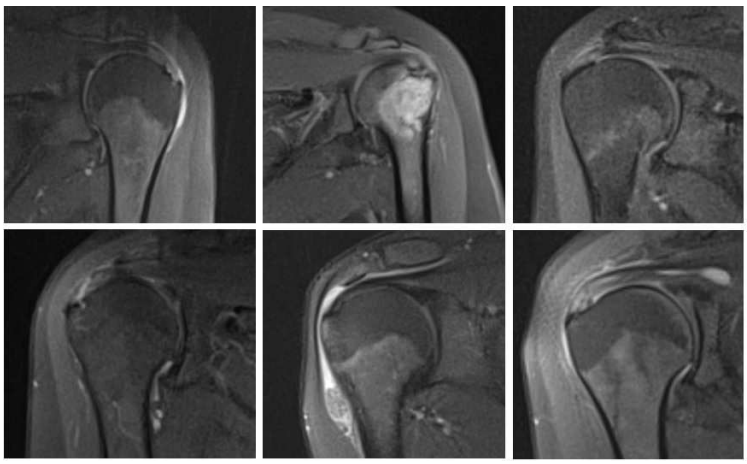
The performance of CapsNet models were tested on the fashion dataset which comprises of 28 28 gray-scale images of 70,000 fashion products from 10 categories. The overall accuracy rate was 92.15% and 93.86% for original and proposed CapsNet models, respectively.

GLCM has been shown to successfully classify ultrasound images of the rotator cuff. However, the overall success rate of GLCM for the classification of the rotator cuff was 85.20% with our automatically determined ROI. CapsNet was more successful than deep CNN, pre-trained networks and GLCM based methods in handling the problems of different poses and orientations of rotator cuff images and in determining the complex texture and shape information needed to classify rotator cuff injuries.

4. Discussion

There are a few studies in the literature evaluating the segmentation and classification of rotator cuff lesions. Most studies were performed in ultrasound images and were based on commonly used methods of texture analysis. This study was performed in coronal PD weighted MR images of the shoulder.

The use of MR imaging in rotator cuff injuries has been shown to be significantly better than ultrasound in the literature. Lenza et al. compared MR imaging and ultrasound for the assessment of rotator cuff tears. The sensitivity and specificity of MRI was 98% (95% CI 92% to 99%) and 79% (95% CI 68% to 87%), respectively (6 studies, 347 shoulders), and 91% (95% CI 83% to 95%)



(a)

Figure 8: a) Demonstration of images labeled incorrectly by the proposed CAD system. The first column contains torn rotator cuff images misclassified as degeneration, the second column demonstrates images classified as torn rotator cuff due to adjacent intensity changes and the third column contains images misclassified as torn rotator cuff due to high level degeneration.

and 85% (95% CI 74% to 92%), respectively, for ultrasound (13 studies, 854 shoulders) [8].

MR imaging is extensively used as a reference clinical imaging technique. Although the anatomy of the rotator cuff is visualized in detail in PD weighted sequences, there are some difficulties in pattern recognition due to anatomical and pathological features. First, the intensity of the bony cortex in PD weighted images is very close to that of the rotator cuff. PD weighted sequences may also be difficult to evaluate because of the low signal to noise ratio, which may be increased by changes in intensity from trauma, degeneration or edema. Moreover, CapsNet uses the entire image to represent imaging features and may be affected by image complexity. Therefore, automatic determination of rotator cuff ROI is essential for PD weighted sequences. CHT was used to automatically determine ROI to overcome unnecessary workload and irrelevant features by CNN and CapsNet.

CapsNet successfully classified automatically determined ROI with an overall sensitivity and specificity of 98.66% (95%CI: 97.48% to 99.39%) and 88.61% (95% CI: 84.58% to 91.89%), respectively, and a confidence level of 95%. Sensitivity and specificity are two important values when a CAD classification system is put into practice. Although the sensitivity rate of our system was high its specificity must be improved in cuff pathologies.

The results of our CAD system were compared to the most common state of the art methods; a custom-designed CNN, pre-trained CNNs [31, 32] and GLCM. The overall success rate of our CapsNet system for the classification of rotator cuff images was 94.75% while the success rates of CNN, GoogLeNet and

AlexNet were 93.21%, 87.63%, 88.45%, respectively. The F1-score of normal, degeneration and tear groups were 91%, 95%, 96%, respectively with the proposed CAD system. On the other hand CNN performed F1-score of 88%, 95%, 94% for normal, degenerated and tear groups, respectively which were slightly lower than the proposed CAD system. Both of the proposed CAD system and CNN performed higher F1-scores for degeneration and tear groups than normal group. The F1-score of degeneration group was surprisingly the highest contrary to the clinical performance of MRI which is higher in discrimination of the tear and normal groups [8].

The proposed CAD system successfully classified the rotator cuff MRIs given in Fig.3 but could not correctly classify the images demonstrated with Fig.8. In the first column there are two images which were labeled as cuff tears by the expert but as degenerations by the CapsNet system (Fig.8). The image in the first row contains a torn rotator cuff with underlying degeneration and the image below is a nonretracted tear resembling intratendinous density changes like degeneration. In the second column the images were misclassified as tear instead of normal cuff. The image in the first row contains fluid accumulation around the cuff and in the image below there is extensive edema in the humeral head. Those two images represent increased intensities in the ROI which is possibly the reason of misclassification. The third column contains two images misclassified as tear although those cuff tissues were homogeneously degenerated with white appearance.

There are an increasing number of image classification and segmentation [25] studies with CNNs in the literature. The success rates of CNN are mainly dependent upon obtaining an adequate number of high quality training datasets. Although data augmentation seems to be a solution for deep CNN, collecting images with all possible poses for each class is unfeasible, especially in the medical field. Although 1006 MR images of the shoulder were included in this study, classification could be improved with a larger dataset. Images with higher quality may also provide better results. In this study 1.5 Tesla MR images were utilized. Regarding the fact that 3 Tesla MRI provide a higher signal to noise ratio [36] further studies with 3 Tesla MR images would increase the image quality and the success of the proposed CAD system.

Transfer learning based CNNs of GoogLeNet [32] and AlexNet [31] successfully classifies images in small datasets. GoogLeNet and AlexNet are trained on natural images that are completely different from our datasets. To our knowledge there are no available pre-trained CNNs similar to our dataset.

Previous studies of the classification of rotator cuff lesions have been based on hand-crafted methods such as GLCM, gray level run length matrix or texture analysis with histogram [18, 19, 20]. All of these texture analysis-based CAD systems used manually segmented ultrasound images from a small group of data as well as GLCM. Although GLCM is highly sensitive to the selected offset direction and orientation values, the data was analyzed with GLCM for comparison. While the size of our dataset was favorable, the success rate of GLCM was 85.20% compared to a 94.75% success rate with our system.

Cho et al. conducted a study on CT images of different body parts in 6000

patients to estimate the minimum training data requirement in a medical image deep learning system to achieve high accuracy [37]. They reported that training data over 200 may not yield significantly better results in the classification success. The classification success increases steeply over 100 samples and forms a plateau after 200 samples. The only apparent advantage of increasing the training data size over 200 is a decrease in the standard deviation.

With regard to the state of the art methods results of this study suggest that capsule networks can be trained with less amount of data for the same or better performance and are more robust to an imbalanced class distribution [38], which makes our approach very promising for the medical imaging community. Although the training time for the capsule network is longer compared to the CNN [39] which may be attributed to its computational complexity, it provides a superior performance of classification in the diagnosis of rotator cuff lesions from MRI.

In this study, a ROI that included surrounding bony structures was automatically determined. The success rate of manually and automatically determined ROI with our CapsNet system was 92.07% and 94.75%, respectively. The success rate of the classification of MR images of the shoulder was increased by preserving local anatomical information about neighboring structures that were on the periphery of injured areas. Therefore, under disease conditions, which result in peripheral spatial effects, image modalities that can provide detailed anatomical information such as MR imaging may be more effective than ultrasound. Furthermore, a wider ROI including adjacent anatomical structures may provide a better overall picture of disease because of the dynamic peripheral impact of tendon ruptures. Therefore, in musculotendinous disorders such as the rotator cuff we recommend defining a wider ROI that contains bony relationships instead of manual segmentation of the diseased structure alone.

Selection of a ROI that is large enough to include the whole attitude of the rotator cuff pathologies and its surroundings is very important for the representability of it. However, there are intractable pitfalls which lead to misclassification but still located in the ROI. Local intensity changes like bone edema or fluid collections around the rotator cuff and highly degenerated cuff tissue itself were the main reasons of misclassification of rotator cuff pathologies in this study.

5. Conclusion

The increasing prevalence and clinical severity of rotator cuff lesions are a public health problem. We proposed a new CAD system based on CapsNet for the classification of rotator cuff injuries from coronal PD weighted MR images. The proposed CapsNet system was found to be successful and the results were quite similar to those of an expert with a sensitivity of 98.66%. The results of proposed automatic CapsNet based CAD system were promising compared to CNN, transfer learning and handcrafted methods. The use of ELU as the activation function increased the success rate of the proposed CNN from 92.50 to 93.21 which was still lower than the CapsNet. This system could be a useful

decision-making tool for less experienced radiologists and orthopedic experts. One limitation of this study is the inclusion of 2D images from one institution. A bigger dataset with higher quality images may yield better results. CapsNet may be considered one of the most innovative and successful models in the field and should be evaluated in further studies.

6. References

[1] R.-F. Chang, C.-C. Lee, C.-M. Lo, Computer-aided diagnosis of different rotator cuff lesions using shoulder musculoskeletal ultrasound, Ultrasound in medicine & biology 42 (9) (2016) 2315–2322.

[2] J. Hermans, J. J. Luime, D. E. Meuffels, M. Reijman, D. L. Simel, S. M. Bierma-Zeinstra, Does this patient with shoulder pain have rotator cuff disease?: The rational clinical examination systematic review, Jama 310 (8) (2013) 837–847.

[3] J. Y. Kim, J. S. Park, Y. G. Rhee, Can preoperative magnetic resonance imaging predict the reparability of massive rotator cuff tears?, The American journal of sports medicine 45 (7) (2017) 1654–1663.

[4] D. S. Weiner, I. Macnab, Superior migration of the humeral head: a radiological aid in the diagnosis of tears of the rotator cuff, The Journal of bone and joint surgery. British volume 52 (3) (1970) 524–527.

[5] C. A. Bretzke, J. R. Crass, E. V. Craig, S. B. Feinberg, Ultrasonography of the rotator cuff. normal and pathologic anatomy., Investigative radiology 20 (3) (1985) 311–315.

[6] T. Liu, J. Ma, H. Cao, D. Hou, L. Xu, Evaluation of the diagnostic performance of the simple method of computed tomography in the assessment of patients with shoulder instability: a prospective cohort study, BMC medical imaging 18 (1) (2018) 45.

[7] M. R. Abreu, M. Recht, Mr imaging of the rotator cuff and rotator interval, in: Musculoskeletal Diseases 2017-2020, Springer, 2017, pp. 203–214.

[8] M. Lenza, R. Buchbinder, Y. Takwoingi, R. V. Johnston, N. C. Hanchard, F. Faloppa, Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered, Cochrane Database of Systematic Reviews (9).

[9] E. Goceri, N. Goceri, Deep learning in medical image analysis: recent advances and future trends, in: International conferences computer graphics, visualization, computer vision and image processing, 2017, pp. 305–311.

[10] D. CireşAn, U. Meier, J. Masci, J. Schmidhuber, Multi-column deep neural network for traffic sign classification, Neural networks 32 (2012) 333–338.

[11] G. E. Hinton, S. Sabour, N. Frosst, Matrix capsules with em routing.

[12] S. Sabour, N. Frosst, G. E. Hinton, Dynamic routing between capsules, in: Advances in Neural Information Processing Systems, 2017, pp. 3856–3866.

[13] J. Ding, B. Chen, H. Liu, M. Huang, Convolutional neural network with data augmentation for sar target recognition, IEEE Geoscience and remote sensing letters 13 (3) (2016) 364–368.

[14] T. DeVries, G. W. Taylor, Dataset augmentation in feature space, arXiv preprint arXiv:1702.05538.

[15] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Data augmentation using synthetic data for time series classification with deep residual networks, arXiv preprint arXiv:1808.02455.

[16] T. Kooi, G. Litjens, B. Van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Medical image analysis 35 (2017) 303–312.

[17] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint arXiv:1712.04621.

[18] M.-H. Horng, Performance evaluation of multiple classification of the ultrasonic supraspinatus images by using ml, rbfnn and svm classifiers, Expert Systems with Applications 37 (6) (2010) 4146–4155.

[19] C.-F. Chao, M.-H. Horng, The construction of support vector machine classifier using the firefly algorithm, Computational intelligence and neuroscience 2015 (2015) 2.

[20] B. E. Park, W. S. Jang, S. K. Yoo, Texture analysis of supraspinatus ultrasound image for computer aided diagnostic system, Healthcare informatics research 22 (4) (2016) 299–304.

[21] R. Gupta, I. Elamvazuthi, S. C. Dass, I. Faye, P. Vasant, J. George, F. Izza, Curvelet based automatic segmentation of supraspinatus tendon from ultrasound image: a focused assistive diagnostic method, Biomedical engineering online 13 (1) (2014) 157.

[22] Y. Kang, G. Y. Lee, J. W. Lee, E. Lee, B. Kim, S. J. Kim, J. M. Ahn, H. S. Kang, Texture analysis of torn rotator cuff on preoperative magnetic resonance arthrography as a predictor of postoperative tendon status, Korean journal of radiology 18 (4) (2017) 691–698.

[23] H. R. Roth, J. Yao, L. Lu, J. Stieger, J. E. Burns, R. M. Summers, Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications, in: Recent advances in computational methods and clinical applications for spine imaging, Springer, 2015, pp. 3–12.

[24] S. Geng, S. Jia, Y. Qiao, J. Yang, Z. Jia, Combining cnn and mil to assist hotspot segmentation in bone scintigraphy, in: International Conference on Neural Information Processing, Springer, 2015, pp. 445–452.

[25] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, R. Leonardi, Deep learning for automated skeletal bone age assessment in x-ray images, Medical image analysis 36 (2017) 41–51.

[26] E. Goceri, Diagnosis of alzheimer's disease with sobolev gradient based optimization and 3d convolutional neural network, International journal for numerical methods in biomedical engineering (2019) e3225.

[27] K. Qiao, C. Zhang, L. Wang, B. Yan, J. Chen, L. Zeng, L. Tong, Accurate reconstruction of image stimuli from human fmri based on the decoding model with capsule network architecture, arXiv preprint arXiv:1801.00602.

[28] A. Mobiny, H. Van Nguyen, Fast capsnet for lung cancer screening, arXiv preprint arXiv:1806.07416.

[29] R. LaLonde, U. Bagci, Capsules for object segmentation, arXiv preprint arXiv:1804.04241.

[30] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint arXiv:1511.07289.

[31] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[33] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

[34] A. Sharif Razavian, H. Azizpour, J. Sullivan, S. Carlsson, Cnn features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2014, pp. 806–813.

[35] E. Göçeri, A. Gooya, On the importance of batch size for deep learning, Minisymposium on Approximation Theory &Minisymposium on Math Education 3-6 July 2018, Istanbul, Turkey 99.

[36] R. Wood, K. Bassett, V. Foerster, C. Spry, L. Tong, 1.5 tesla magnetic resonance imaging scanners compared with 3.0 tesla magnetic resonance imaging scanners: systematic review of clinical effectiveness, CADTH technology overviews 2 (2).

[37] J. Cho, K. Lee, E. Shin, G. Choy, S. Do, How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?, arXiv preprint arXiv:1511.06348.

[38] A. Jiménez-Sánchez, S. Albarqouni, D. Mateus, Capsule networks against medical imaging data challenges, in: Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, 2018, pp. 150–160.

[39] R. Mukhometzianov, J. Carrillo, Capsnet comparative performance evaluation for image classification, arXiv preprint arXiv:1805.11195.

Aysun Sezer: I'm a post doctorate researcher on machine learning for embedded classification of plancton images in the ENSTA-ParisTech, Université de Paris-Saclay, France, in the frame of the BRIDGES European Project. My research area of interest is biomedical image analysis, with a specific emphasis on segmentation of MR, CT and ultrasound images of human musculoskeletal system.

Hasan Basri Sezer: I am an orthopaedic surgeon since 2007. I have been working in Şişli Hamidiye Etfal training and research Hospital since 2010. I am working as a fellow in clinique du sport/ Nollet Institute, Paris. My research fields are sport surgery, trauma and pediatric orthopedics.