Faculty of Medicine
Biomedical Engineering

Master of Science Thesis

Deep Learning based Fully Automatic Quantification of Rotator Cuff Tears from MRI

by

Stefan Weber

of Basel BS

Supervisors
PD Dr. Kate Gerber, MSc Hanspeter Hess, Dr. Guodong Zeng

Institutions
University of Bern, sitem Center for Translational Medicine and Biomedical Entrepreneurship

Examiners
PD Dr. Kate Gerber and Dr. Nicolas Gerber

Bern, September 2021

Bern University of Applied Sciences

UNIVERSITÄT BERN

Abstract

Rotator cuff tears are the most common source of shoulder pain. Many factors can be considered to choose the right surgical treatment procedure. The most important factors are tear thickness (partial vs. full), tear size, tear shape, and muscle quality. The aim of this work was the fully automated quantification and classification of a full-thickness posterosuperior rotator cuff tear from MR images using a deep learning based approach.

A complete new approach to automatically quantify and classify a rotator cuff tear, based on the segmentation of the tear from MR images, was developed and validated. A neural network was trained to segment the rotator cuff tear from an MR image and automatic methods for calculating tear width and retraction and for classifying the tear according to pattern, extension and retraction were implemented.

The accuracy of the automatic segmentation and the automated tear analysis were evaluated relative to the ground truth of manual segmentations by a clinical expert, and analyzed based on the ground truth segmentations. Variance in the manual segmentations was assessed in a interrater variability study of two clinical experts. The error of the automatic segmentation to one of the two clinical experts are meant to be in the same region.

To make quantification accessible the whole pipeline was implemented in an existing webapp.

The results were also evaluated clinically by intraoperative measurements of the rotator cuff tear performed on a separate dataset of six patients.

The accuracy of the tear retraction calculation based on the developed automatic tear segmentation was 6.56 mm ± 6.48 mm in comparison to the interrater variability of tear retraction calculation based on manual segmentations of 3.77 mm ± 3.58 mm.

These results show that an automatic quantification of a rotator cuff tear is possible. The achieved accuracies of the quantification pipeline for rotator cuff tears need to be improved to make them clinically useable. The large interrater variability of manual segmentation based measurements, highlights the difficulty of the tear segmentations task in general. To improve the accuracy of an automated segmentation, a larger dataset for training may be required or a semi-automatic approach could be used.

Acknowledgements

It was a difficult time to write this thesis during a worldwide pandemic. Despite these aggravating circumstances, I received a lot of support from all sides. I am truly grateful for the support, the laughs, the difficulties, and the love I received during this time and throughout my studies.
First I want to thank the group heads, Dr. Kate Gerber, Dr. Nicolas Gerber. I am really impressed how you master everything and so successfully. I also enjoyed it when the whole family was present. It was then always a very cheering mood.
A big thanks to Hanspeter Hess. For the great trust you have put in me and for always supporting me. I am sure your whole PhD project will be a true success.
I also want to thank the rest of the personalized medicine group at SITEM. Especially Adrian Ruckli and Dr. Guodong Zeng for giving precious tips and ideas. All the conversations with you really helped me every time.
Many thanks to the clinical partners Dr. Mathias Zumstein, Dr. Tomàs Rojas and Philipp Gussarow. I enjoyed every meeting with you and always looked forward to it. It was interesting to see the problems from your points of view. Only thanks to you it was possible to deliver this thesis. A highlight of my thesis was attending rotator cuff repair surgery.
The greatest thank goes to my partner Carolina. You are always here for me and help me in every aspect. You brightened up my entire study period almost from start to finish. I look forward to starting a new, unknown chapter with you.
I would also like to thank my whole family. You support me in every phase of my life and I know that I can always count on you.

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche kenntlich gemacht. Mir ist bekannt, dass andernfalls der Senat gemäss dem Gesetz über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, September 17th 2021

Stefan Weber

Contents

1 Introduction

In a first step, a brief introduction to the general clinical background is given. The clinical background explains what a rotator cuff tear is, why and how often it occurs, and what steps are required from the detection of a tear to the treatment of a tear. Because there are several possibilities to quantify a rotator cuff tear the main focus of this work is the size of the tear defined by its anterior to posterior extension, which is called the tear width and also its lateral to medial extensions, which is called the retraction. In clinical field a rotator cuff tear is not quantified but rather classified. The most established classifications system of rotator cuff tears are explained in a second step. The last section of the introduction explains what others have already done in this field and why the approach of this work is a novelty. Finally the aim of this work will be stated.

1.1 Clinical Background

The shoulder joint is a ball and socket joint and has the highest range of motion of all joints in the human body. The main parts are the humeral head and the glenoid cavity of the scapula. Due to the high range of motion the shoulder joint must be stabilized by the rotator cuff. The rotator cuff consists of four muscles as seen in Figure 1. The muscles are the supraspinatus, infraspinatus, terese minor and the subscapularis which are all connected to the humeral head. The two main functions of the rotator cuff are to stabilize the humeral head in the glenoid fossa and to enable the humerus to change position relative to the scapula.[4]

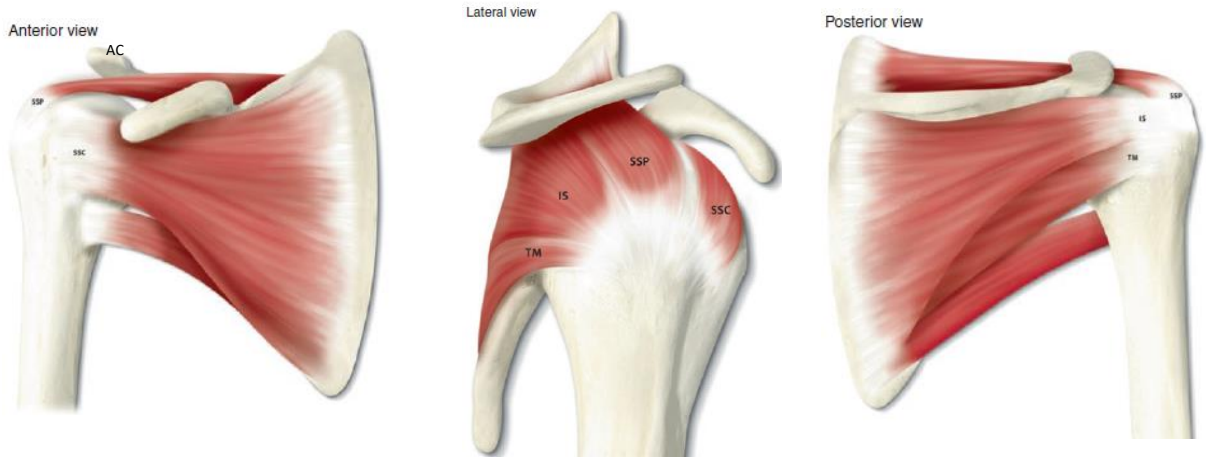

Figure 1: Anatomy of the rotator cuff muscles supraspinatus (SSP), infraspinatus (IS), terese minor (TM), subscapularis (SSC), acromion (AC)[6]

Tears of the rotator cuff tendons are the most common source of shoulder pain. Several studies are conducted to evaluate the prevalence of rotator cuff tears and it varies between 5 to 44 %. In addition to traumatic rotator cuff tears affecting younger patients, most studies have found that the prevalence of rotator cuff tears increases with age. Therefore, rotator cuff tear can be considered as a physiological condition related to the progressive degeneration of the tissue. One reason why prevalence varies is that a rotator cuff tear can be asymptomatic. In asymptomatic rotator cuff tears, there is no clinically significant loss of shoulder function compared to a healthy shoulder. An asymptomatic rotator cuff tear is a risk factor that can progress to a symptomatic tear. [17, 26]

The symptoms of a symptomatic rotator cuff tear are pain, discomfort, weakness, and reduced range of motion [20].
The etiopathogenesis of rotator cuff tears is multifactorial and can be grouped into extrinsic and intrinsic factors. An often described extrinsic factor is the impingement syndrome, where the acromion impinges into the bursal side of the rotator cuff tendon of the supraspinatus and the infraspinatus. Studies of cadavers have shown that most rotator cuff tears occur on the articular side of the tendon away from the acromion, where the tear is caused by a friction phenomenon. Other extrinsic factors are mechanical overuse, smoking habits, dislocations and fractures. It is assumed that the majority of the rotator cuff tears are primarily caused by intrinsic degeneration, as already stated before. There are several theories why this degenerations such as hypoperfusion theory, degenerative theory, degeneration-microtrauma, apoptotic theory, and extra cellular matrix modifications appear. [13, 31]
For clinical examination of a rotator cuff tear, various tests can be performed. These tests require the patient to perform a certain movement or hold the arm in a certain position. These tests only give some indication that a rotator cuff tear is likely to be present. In addition to the clinical examinations, radiographs are typically the initial imaging test for shoulder injury. If a rotator cuff tear is suspected, further imaging modalities are performed. Magnetic Resonance Imaging (MRI) and ultrasound are well suited for detecting rotator cuff tears. In addition, Computer Tomography (CT) may also be considered. MRI offers several advantages over ultrasound (US) in the study of the rotator cuff. MRI has a higher reliability in determining the size of the tear and the degree of retraction of the tendon compared with US and has been preferred for surgical planning. Imaging is performed either without contrast agent or with direct injection of contrast agent into the joint. A rotator cuff tear appears as a high intensity signal within the tendon on T2-weighted and proton density sequences with fat suppression on MRI images. This is due to the presence of synovial fluid and contrast agent (if used) in the injured tendon. [2, 6, 13, 33]
Spontaneous healing of the rotator cuff has only been clearly observed in animal studies. There are inconsistent results in human trials with a range of spontaneous healing between 0-37 %. Therefore, it is not clear whether a rotator cuff tear can heal spontaneously in humans. [13] Asymptomatic or small to medium tears may be treated by conservative treatment, which includes physical therapy for muscle strengthening and subacromial injections of different medications. If there is no improvement with conservative treatment a rotator cuff repair or shoulder replacement by reverse shoulder arthroplasty are the options. Rotator cuff repair is usually performed arthroscopically or via the miniopen technique.
To choose the right surgical treatment procedure many factors can be considered. The most important factors are tear thickness (partial vs. full), tear size, tear shape, and muscle quality. Muscle quality is assessed by the Goutallier grade, which is an indicator of fat infiltration of the rotator cuff muscles and by the extent of muscle atrophy. Full-thickness tears can be classified by the size of the tear, which is most commonly measured in the coronal plane on MRI images. PD TSE with fat suppression and T2-weighted are well suited MRI sequences for the identification and classification of the rotator cuff tears. With increasing tear size and decreasing muscle quality, the opportunity for a positive surgical outcome of a rotator cuff repair decreases. In particular, the medial to lateral extent of the tear caused by retraction of the rotator cuff muscles is an indicator of the potential success of surgery. If the tear width is large the tendon is more likely to retract. Also the age and the physical demand of the patient are important factors that must be considered. [8, 12, 13]
A rotator cuff tear can either be reparable or irreparable. In irreparable cases the reverse shoulder arthroplasty is the most common treatment option. In reparable cases the tendons of the involved rotator cuff muscles are reattached to the footprint anatomy of the humeral head. Currently, there is insufficient data in the scientific literature to establish an evidence-based treatment algorithm. Treatment is based on patient factors and associated pathology and therefore involves personal experience.[12, 13]

Some steps of a rotator cuff repair remain the same regardless of the surgical technique used. Following is a ruff overview which steps are involved [1, 13]:

1. Evaluation of the size, shape, and elasticity of the lesion.
2. Assessment of the extent of tear retraction.
3. Assessment of the thickness and fraying (delamination) of the tear edge.
4. Debridement of the nonviable tendon.
5. Insertion of suture anchors or creation of transosseous tunnels. The number of suture anchors or transosseous tunnels depends on the tear size and preferred technique of the surgeon.
6. Suturing of the cuff tear edges. The exact suturing technique depends on the shape and size of the tear. The differences can be seen in Figure 2.
7. Secure knots.

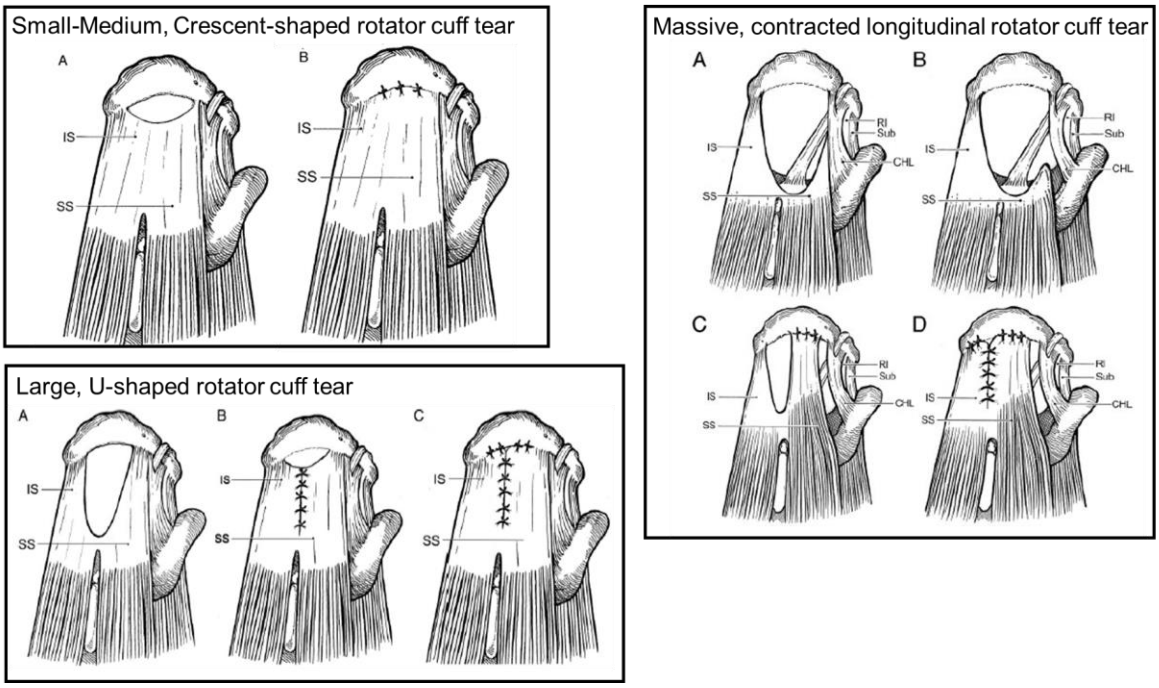

Figure 2: Illustration of how the surgical technique depends on the size and the shape of the tear. Supraspinatus tendon (SS), Infraspinatus Tendon (IS), rotator interval (RI), Subscapularis tendon (Sub), Coracohumeral ligament (CHL) [16]

The most common complication after rotator cuff repair is a re-tear of the tendon. This can be a consequence of a wrongly assessed rotator cuff tear. [13]

1.2 Classifications of rotator cuff tears

Rotator cuff tears can be classified and quantified under different aspects. In this section, all important classification systems are briefly explained and it is shown how the classification is included in this work.

"In general, a classification for rotator cuff tears should follow several principles. First, the classification system should be already in use, if possible, validated for reliability, and easily used by physicians. Second, it should be descriptive to define the location and anatomy of the tear, helping all surgeons to understand precisely its characteristics. Third, the classification should be useful to dictate appropriate treatment in each specific case, and fourth, it should also have a predictive value both to guide physicians and to transmit the patient realistic expectations of postoperative outcome. Finally classification also enables the surgeon to acquire data and to communicate clinical and functional results obtained with different treatments to the scientific community [13]."

1.2.1 Classification by involved components

Rotator cuff tears can be classified by the tendons involved. There are five possible types as illustrated in Figure 3: type A; supraspinatus and superior subscapularis tears, type B; supraspinatus and entire subscapularis tears, type C; supraspinatus, superior subscapularis, and infraspinatus tears, type D; supraspinatus and infraspinatus tears, and type E; supraspinatus, infraspinatus, and teres minor tears. Tears where the supraspinatus, the infraspinatus and the teres minor tendons are involved are called posterosuperior rotator cuff tears [3, 7, 27].

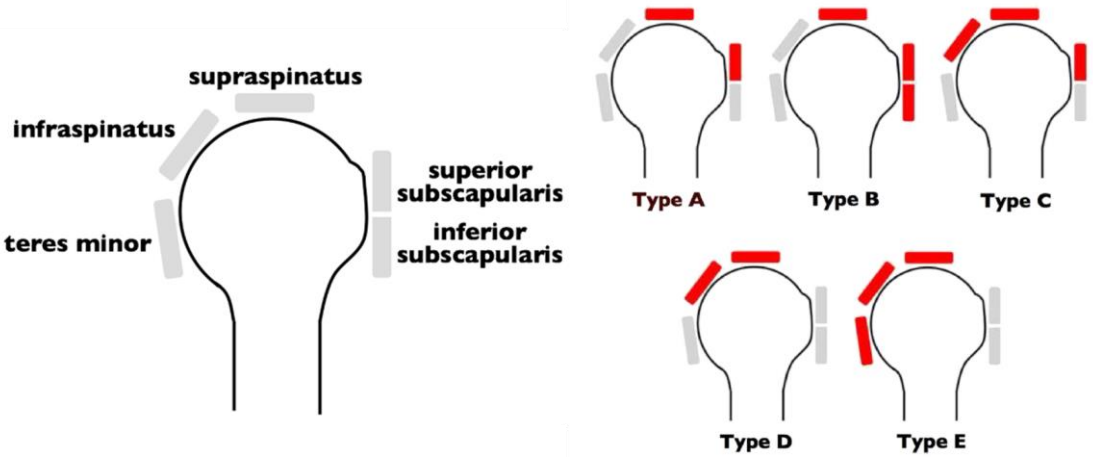

Figure 3: Classification by involved tendons by Collin et al. [7]

1.2.2 Classification by tear thickness

In the case of a rotator cuff tear, a distinction is made between partial and full thickness tears. In partial thickness tears the involved tendon is not fully torn along its thickness. They can be classified by the involved surface and by the extent of the tear. A partial tear can be either articular or bursal. An articular sided tear is facing towards the shoulder joint and an bursal sided tear is facing away from the shoulder joint. In the Ellman classification the extend of the partial thickness tear is divided into three categories: superficial disruption of tendinous fibers (grade 1), depth not exceeding one half of the tendon thickness (grade 2), and the continuity of the tendon appears insufficient (grade 3). An example of a grade 3 partial articular tear is shown in Figure 4 on the right. [6]

In a full thickness tear the tear is torn along it complete extend, as Figure 4 shows on the left.

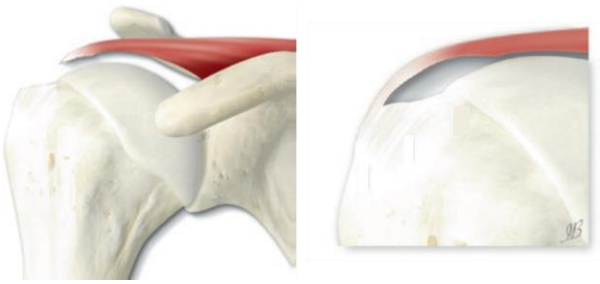

Figure 4: A full thickness tear of the supraspinatus tendon (left); A partial thickness tear of the supraspinatus tendon (right) [6]

1.2.3 Classification by tear size

There are few classification systems that consider only the size of the tear. The two best known classification systems for tear size were defined by DeOrio and Cofield and Patte. DeOrio and Cofield classified the tear sizes measured during surgery according to the tear width of the tendon torn off from the humeral head. The tears are classified in *small (>1 cm)*, *medium(1 – 3 cm)*, *large (3-5 cm)*, and *massive (> 5 cm)*. This system has the disadvantage that it is only one-dimensional and that the difficulty of a repair is overestimated.

The Patte system is a common classification for measuring retraction in the coronal plane on MRI images. It includes three levels used to measure retraction relative to the glenoid, as shown in Figure 5. Retraction is due to the shortening of tendons and muscles and is a good indicator of the reparability of the tear, since larger and older tears with marked muscle atrophy and fatty infiltration tend to have a higher retraction. [3, 6, 13, 18]

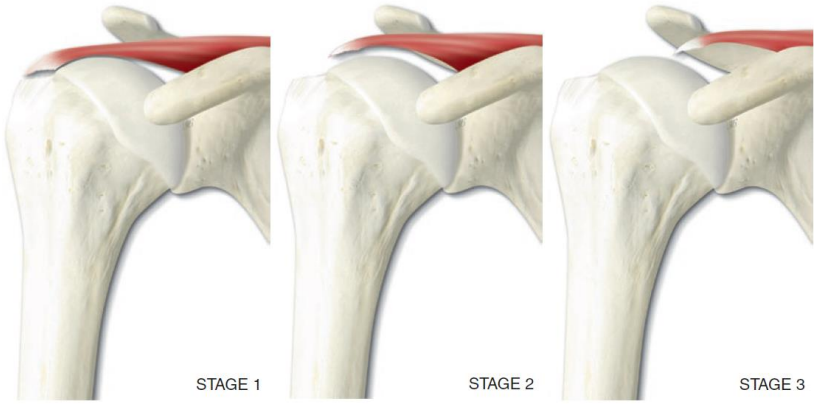

Figure 5: The three stages of the Patte classification system

1.2.4 Classification by tear pattern

The tear pattern of posterosuperior rotator cuff tears gives an indication of how the rotator cuff should be repaired, as showed earlier in Figure 2. There are four common tear patterns in Figure 6: crescent-shaped tears, U-shaped tears, L-shaped and reversed L-shaped (rL-shaped) tears, and massive tears. The pattern of a tear is defined by the tear width and the retraction and the location of the tear in the supraspinatus and infraspinatus tendons. The distinction between L-shaped and inverted L-shaped tears depends on the anatomical direction from which the tear originates. [9]

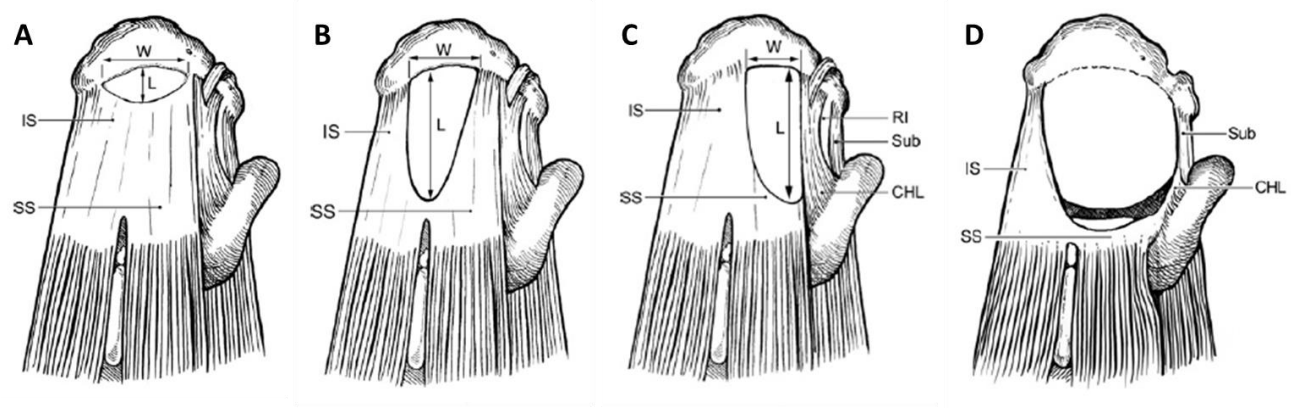

Figure 6: A) crescent-shaped tear; B) U-shaped tear; C) L-shaped tear; D) massive tear [10]

1.2.5 Classification by multiple factors

The Snyder and ISAKOS classification system take multiple factor into account to give an estimate of the outcome of a rotator cuff repair. The Snyder classification systems is develop for full thinks tears and is defined as follows:

Table 1: Snyder classification of complete tears posterosuperior tears [6, 13]

Class	Definition	Prognosis
C I	Small, complete tear, such as a puncture wound (<1 cm)	Repair is easily obtainable
C II	Moderate tear (usually <2 cm) that still encompasses only one of the rotator cuff tendons (supraspinatus) with no retraction of the torn ends	Repair is easily obtainable
C III	Large, complete tear (usually 3–4 cm) involving two tendons with minimal retraction of the torn edge	Repairable
C IV	Massive rotator cuff tear involving two or more rotator cuff tendons, frequently associated with retraction and scarring of the remaining tendon ends and fatty degeneration of the respective muscles.	Partially repairable or irreparable

The ISAKOS classification system described in Table 2 considers location, extension, pattern, fatty atrophy, and retraction. ISAKOS stands for International Society of Arthroscopy, Knee Surgery, and Orthopedic Sports Medicine and consist of surgeons all around the world. It is possible to classify all types of tears such as partial and full thickness tears and also posterosuperior and anterior tears. The creators of the system give no prognosis for the outcome of a rotator cuff repair but they state that it is a complete and straightforward, user-friendly method to describe all rotator cuff tears and give the surgeon the possibility to predict difficulties rotator cuff repair. [3, 6, 13]

Table 2: ISAKOS classification system [3]

Location (L)	Extension (E)	Pattern (P)	Fatty atrophy (A) [c]		Retraction (R) [d]
Partial thickness posterosuperior	>50 % thickness <50 % thickness	A (Articular) B (Bursal) I (Interstitial)	SS0 SS1	IS0 IS1	
Full thickness posterosuperior (SS-IS)	C I [a] C II C III C IV	Crescent U-Shape L-Shape rL-Shape	SS2 SS3 SS4	IS2 IS3 IS4	1
Anterior (SC)	1 [b] 2 3 4 5		SC0 SC1 SC2 SC3 SC4		2 3

SS: Supraspinatus; IS: Infraspinatus; SC: Subscapularis
[a] Snyder classification; [b] Lafosse classification; [c] Gutallier classification; [d] Patte classification

The classification according to Lafosse and Gutallier is not used in this thesis and therefore not explained further.

1.3 Current practice and related work

Currently all in 1.2 explained classifications are extracted either from preoperative imaging techniques such as MRI or intraoperatively during the surgery. To get an accurate classification highly experienced surgeons or radiologists are needed and the interrater reliability between different raters is only moderate.

A study conducted in 2012 found that the interrater reliability among three orthopedic surgeons in the Patte classification from T2 weighted coronal MRI images was only moderate with a Cohen's kappa coefficient of $\kappa = 0.58$ [19]. Another study from 2013 found that the distinction between no tear, partial tear or full thickness tear of the supraspinatus tendon from MRI among 16 orthopedic surgeons had also a moderate interrater reliability with $\kappa = 0.6$. If the classification was only done by grouping the tears into no full thickness tears and full thickness tear the interrater reliability was substantial with $\kappa = 0.77$ [11].

Deep Learning with convolutional neural network based methods is a well-suited technique for solving image-based problems in the medical field. Therefor it can also be used to identify of classify rotator cuff tears. In the field of musculoskeletal applications, the number of publications involving Deep Learning has increased significantly in the last decade. The applications can be classified into four categories: Segmentation, Classification, Lesion Detection, and Non-Interpretive Tasks. Segmentation is a process in which image content is grouped pixel by pixel into different categories. In this way, anatomical structures can be extracted from medical images. Some examples are the segmentation of cartilage, bone and the proximal femur. In classification, the algorithm outputs the probability of assigning a given input image to a particular category. Some previously published musculoskeletal examples include: Bone age assessment or sex determination. Lesion detection is a subcategory of classification tasks where the output of the algorithm is the presence or absence of a particular lesion, e.g., fracture or meniscal tear detection. The last category is non-interpretative tasks, which can be best described by some examples: Improving the image quality of low-dose CT images or low-sampling MRI images, or accelerating the acquisition of MRI images.[5]

Deep learning based approach to classify the fatty atrophy from MRI images were already used and had similar accuracy compared to classified images from surgeons [21, 28].

Fully automated classification of rotator cuff tears based on their tear thickness and by its size were done recently by Kim et al. and Shim et. al [22, 30].

Kim et al. fully automatically grouped coronal T2 weighted MRI images in to normal, partial thickness tear or full thickness tear by using an deep neural network. They were also able to localize the tear in the image by class activation maps. A Class activation map is an explanation method to detect on which pixels the convolution neural network focused to make is decision for the classification [32]. One drawback of class activation maps is that the location of the lesion only roughly can be detected. The network was trained on a data set with of total 2477 MRI images (1963 training, 242 Validation, 242 testing) and they achieved an accuracy of 0.87 on the testing set. The ground truth for the training set was two times labeled by a single orthopedic surgeon with more than 10 experience using an interval of two weeks in between. The dataset they used was skewed and had much higher number of normal shoulder then shoulder with a tear. [22]
Shim et al. did something similar to Kim et al. but they grouped the rotator cuff tears into five classes by also consider the size of the tear. The classes are: None, partial, small, medium, Large-to-Massive. To localize the tear they used also class activation maps. The dataset consist of 2124 patients (1924 training, 200 testing) with axial T1 weighted MRI images, coronal and sagittal T2 weighted MRI images. Their neural network achieved an accuracy of 0.69 and an top 1 ± 1 accuracy of 0.875 which was significantly better compared to the classification done by four orthopedic shoulder specialists. The diagnostics time of the algorithm is with 0.01 seconds significantly shorter comparted to the orthopedic shoulder specialists with a diagnostic time of 20.3 seconds. The ground truth was obtained retrospectively during surgery. [30]

1.4 Aim of this Thesis

The aim of this work is the fully automated quantification and classification of a full-thickness posterosuperior rotator cuff tear from MRI images using a deep learning based approach. The classification is based on the retraction and the tear width in millimeters. The obtained measurements are used to give indications of the pattern of the tear, to classify the extensions according to the Snyder classification and to classify the retraction according to the Patte classification. In this work, a different approach was used compared to Shim et al. and Kim et al. by first segmenting the tear and then extracting the measured values from the segmented area. This approach allows a neural network to be trained with only 57 patients. In contrast to a classification task, segmentation requires less data. An additional advantage is that the segmented area of the tear can be represented as a three-dimensional volume, which allows the surgeon to interpret the shape and extend more easily. The accuracy of the extracted measurements was compared with two orthopedic shoulder specialists. Within this thesis, the automatic algorithm for tear shape classification based on tear segmentation is presented along with verification of the algorithms accuracy as compared to manual and clinical measurements. To make the concept available to surgeons the algorithm was implemented in a webapp which is easy to use.

2 Material and Methods

A method for automatic size quantification and classification of full thickness rotator cuff tears was developed based on automatic segmentation of the tear. For automatic segmentation of the rotator cuff tear from MRI, a neural network was trained with ground truth manual segmentations of MRI images created by a clinical shoulder expert. An algorithm for automatic calculation of the tear width and retraction based on the segmented tear data was developed along with a method for automatic classification.
The accuracy of the automatic segmentation algorithm was evaluated against ground truth segmentations. To verify the entire automatic tear evaluation pipeline, tear size and classification calculated with the developed methodology were evaluated against intraoperative tear measurements in a preliminary clinical study.

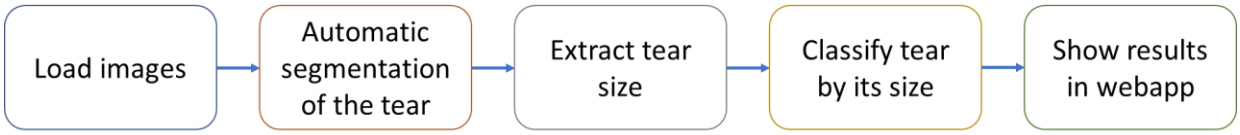

Figure 7: Overview of the whole pipeline for automatic quantification and classification of a rotator cuff tear.

2.1 Methods used to Quantify and Classify Segmentation

For quantification and classification, retraction and tear width were calculated in millimeters from the manual and automatic segmentations according to the formulas (2.1) and (2.2). The extracted measurements were further used to classify the rotator cuff tear by its extension, pattern, and retraction. Therefore, the classification of full thickness posterosuperior tears from ISKOS in was applied. Due to the lack of spatial context, some of the classifications could not be applied directly. For example, it is not clear from the predicted segmentation which tendons are affected because the tendons are not segmented.

2.1.1 Quantification of Tear Size from Segmentation

A method for automatically quantifying the tear size was developed. Tear parameter analysis from a segmentation in anisotropic data is performed in two different ways. The in-plane evaluation is done in the high-resolution plane of the MRI image. Additionally, the segmentation is evaluated in the direction orthogonal to the in-plane slices. However, this orthogonal shows a significantly lower resolution due to the high-slice thickness, and therefore results in a lower accuracy.

The in-plane calculation of the tear retraction is applied in the coronal slice, whereas the tear width is analyzed in the sagittal slice.
The in-plane tear measurement is done in the following procedure:
The algorithm evaluates in each segmented slice i the distance d_i between the two points on the border of the segmentation with the largest distance to each other p_i, q_i.
The tear retraction in the coronal plane and the tear width in the sagittal plane ($d_{InPlane}$) is defined as the greatest of theses distance (d_i) according to equation (2.1).

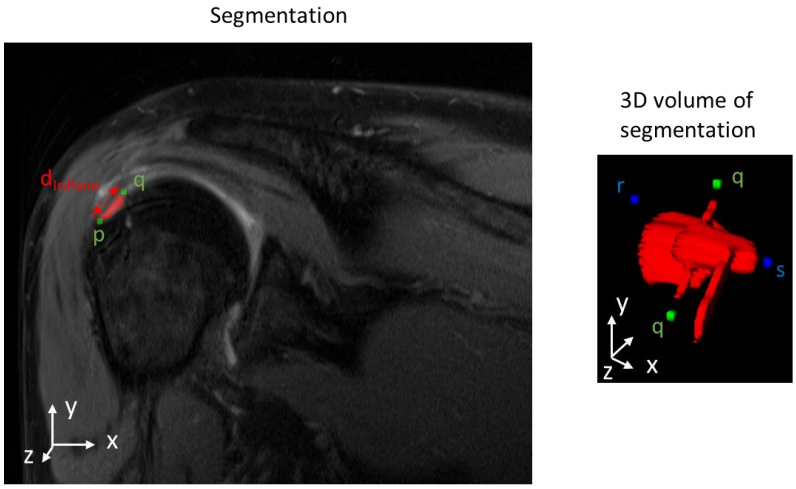

Figure 8: Quantification of the tear size in coronal view. To quantify the tear size the retraction and tear width are measured.

The tear size in the image plane is defined as

$$d_{InPlane}(p,q) = \sqrt{((q_x - p_x)ps_x)^2 + ((q_y - p_y)ps_y)^2} \quad , \tag{2.1}$$

where the position of the points p and q are described in pixels. To get the size in millimeters, the difference between the two points in *x*- and *y*-direction must be multiplied with the respective pixel spacing ps_x and ps_y.

Additionally, the tear is evaluated in an orthogonal manner along the low-resolution direction. To evaluate the tear in the orthogonal direction, the number of slices with a tear segmentation is multiplied with the slice thickness. This orthogonal measure $d_{Orthogonal}$ allows the measurement of the tear width in the coronal view and the tear retraction in the sagittal view.

The tear size orthogonal to the image plane was calculated by only considering the difference in the z direction, which is the same as the number of slices which are segmented, as shown in formula (2.2), where st_z describes the slice thickness.

$$d_{Orthogonal}(r,s) = \sqrt{((r_z - s_z)st_z)^2} \tag{2.2}$$

2.1.2 Classification of the Extension

For classification of the extension, the Snyder classification was used, which is described in Table 1. The Snyder classification had to be adapted because of lack of spatial context. Because tendons are not segmented by the algorithm or by clinical experts, no number of affected tendons can be determined when applying the Snyder classification. The complete classification decision tree is shown in Figure 9.

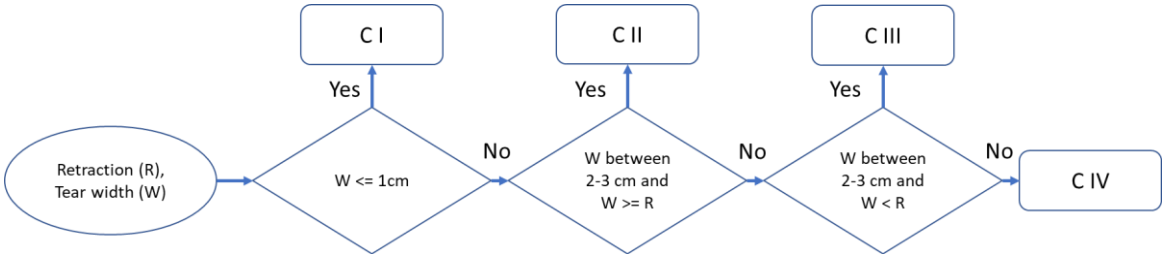

Figure 9: Adapted decision tree of the Snyder classification. The number of involved tendons are not considered in this decision trees.

2.1.3 Classification of the Tear Pattern

For identification of the tear pattern, the method suggested by Davidson and Burkhart [10] was used. The method was originally applied for manual prediction of the tear pattern from MRI images. The patterns were classified into crescent-shaped tears, longitudinal tears, and massive tears. L-shaped, rL-shaped, and U-shaped tears are all classified as longitudinal tears. The decision tree is shown in Figure 10.

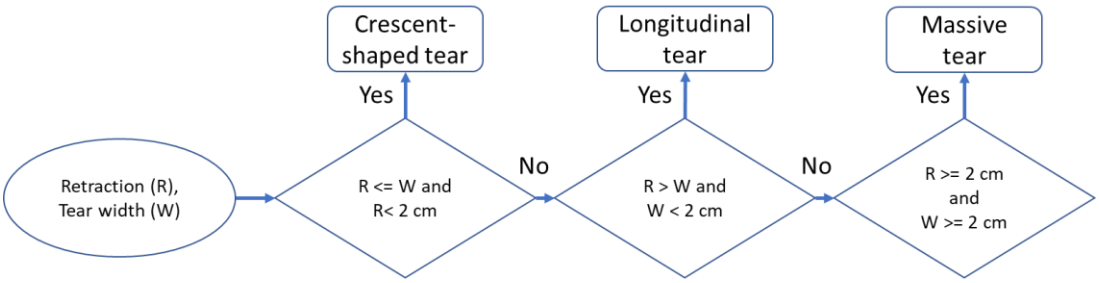

Figure 10: Decision three to predict the tear pattern by Davidson and Burkhart

2.1.4 Classification of the Retraction

The retraction is classified by the Patte classification. Usually, the retraction is classified by its relative position to the humeral head. It was not possible to apply the original Patte classification, since the humeral head was not segmented. Thus, the grades of the retraction were determined relative to the mean diameter of the humeral head. As reference a mean diameter of 45 mm was used [15, 23]. The tear was classified as grade I if the retraction was smaller than 22.5 mm. If the retraction was between 22.5 mm and 45 mm it was classified as grade II and if the retraction was larger than 45 mm it was classified as grade III.

To evaluate the accuracy of the neural network compared to the ground truth, all measures and classifications described are determined from the predicted segmentation and the ground truth segmentation.

2.2 Data Acquisition and Patient Selection

For all tasks a dataset with coronal and sagittal slices of PD TSE fs, PD Blade and T2w MRI images was used. The images from the dataset were acquired from different MRI scanners in different institutes and thus show a high variance. Nevertheless, all patients underwent surgery in the same clinic. Table 3 shows the reasons of these variances. Reasons for this variance are the different demographic factors of the patients and the different acquisition of the MRI images. In general, the larger the variance and the smaller the area to be segmented, the more data is needed for an accurate segmentation result [24].

Table 3: Summary of patient demographics and image acquisition

Patient demographics		Image acquisition	
# Patients	57	Anatomical Direction (Cor/Sag)	57/46
Age (years)	58.6 (22-77)	# Institutions	17
Gender (M/F)	37/20	# Manufacturers	3
Laterality (R/L)	37/20	# Models	11
Tear width (mm)	20.3 (6.8-68.3)	Magnetic Field Strength (1.5T/3T)	49/54
Tear retraction (mm)	24.3 (8.6-57.8)	Sequence Cor (PD TSE fs/PD Blade fs/T2w)	54/2/1
		Sequence Sag (PD TSE fs/PD Blade fs/T2w)	39/4/3
		Spacing between slices (mm)	3.5 (2.4-4.4)
		Pixel spacing (mm)	0.34 (0.21-0.63)

The selection of the patients for the dataset is described in Figure 11. As a first step 121 patients who underwent surgery in the same clinic and suffered from a reparable rotator cuff tear were selected. Only full thickness tear patients with coronal PD TSE fs, PD Blade fs or T2w MRI images were selected. If a sagittal view of the selected patients was available, it was also included into the dataset. The final number of selected patients was 57.

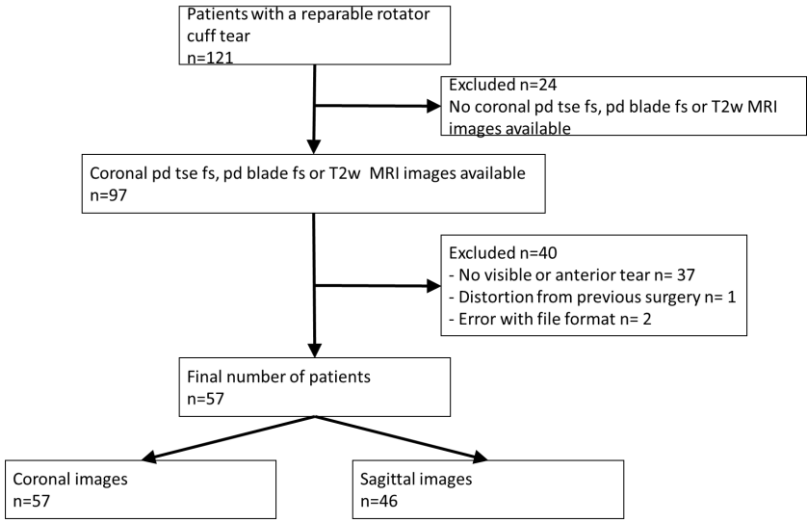

Figure 11: Patient selection for rotator cuff tear quantification

All MRI images were transferred from the DICOM format to NIFTI files. This allowed a more secure anonymization and an easier organization of data. NIFTI files do not save patient specific information, but save all image information in one single file, containing only metadata which are relevant for the representation of the images.
The ground truth segmentation was performed by an orthopedic surgeon, specialized in shoulder surgeries.

2.3 Network Architecture, Software and Libraries

For the realization of the ground truth segmentation, the software *ITK Snap* version 3.8.0 was used.
A 2D patch-based U-net was used to automatically segment a rotator cuff tear. The U-net is a convolutional neural network developed for medical image segmentation by Ronneberger et al. [29]. It is the current state of the art method to perform a segmentation task in this field. A patch-based segmentation algorithm allows efficient training by extracting 2D patches from 3D MRI images.

The MRI images from the dataset described in 0 have a spacing between slices of around 3.5 mm between each coronal or sagittal slice, respectively. A rotator cuff tear is a small lesion compared to the whole image volume and the spacing between slices. A 2D U-net can still handle such spacings and lesion sizes, which is why it was the preferred architecture of the neural network.
The entire algorithm was implemented in *Python* version 3.8.2. For the preprocessing and the implementation of the patch-based pipeline *TorchIO* version 0.18.36 was used. For the training of the neural network the deep learning framework *PyTorch* version 1.8.1 was used.

2.4 Segmentation Pipeline for the Training Phase

Figure 12shows the entire segmentation pipeline in the training phase. Since the images were acquired from different MRI scanners in different institutes, the images had to be preprocessed to ensure a similar distribution and range of intensities of all images [25]. At first, a histogram normalization was applied to the images. Furthermore, a z-score normalization was applied to the images to have a mean of zero and a unit variance. To account for the occurrence of left and right shoulder in the dataset, the images were randomly flipped along the medial axis with a probability of 0.5. To address for the different spacing of the pixels, random affine and elastic transformations were applied to the images, involving the scaling of the images. The advantage of all these transformations was the artificial enlargement of the dataset. The preprocessing being performed on the whole MRI image and not only on specific 2D slices guarantees a more concise preprocessing especially for the normalization of the images.
2D slices with the shape of 256x256 were randomly picked by a label sampler from the preprocessed 3D images. This ensured that the probability of randomly selecting a field with labeled voxels in the center was 0.5. The patch size was chosen large enough to ensure that all relevant features to detect a rotator cuff tear are contained in the sampled patch.
To evaluate the accuracy of the neural network, a seven-fold cross validation was performed. The dataset was divided into seven groups. Consequently, the network was trained seven times. Each training iteration a different dataset group was the validation dataset. The other six groups were used as the training dataset. Thanks to that process the performance of the network was evaluated on unseen data and on the whole dataset.

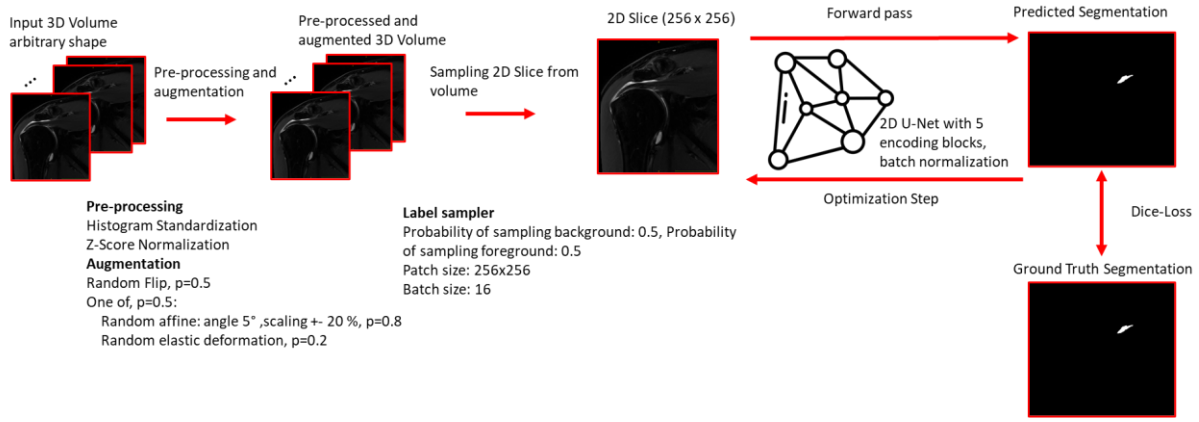

Figure 12: Illustration of the segmentation pipeline during the training process.

2.5 Segmentation Pipeline for the Prediction

The segmentation pipeline used to perform a prediction is shown in Figure 13. The input MRI images were preprocessed with a histogram normalization and a z-score normalization. They were fed into the trained neural network and split up in several 2D image patches with a shape of 256x256.
The grid sampler was applied to ensure that for every patch of an image a segmentation was predicted. After the prediction, all segmentations were aggregated back to a 3D image of the same shape as the input image. Only the largest connected component of the prediction was kept.
The size of the rotator cuff tear was extracted by finding the largest extend in every coronal or sagittal slice, depending on which anatomical direction the input MRI image was acquired. The size of the tear orthogonal to the image plane was found in the same matter. These measurements would be less accurate due to the high slice thickness of the images.

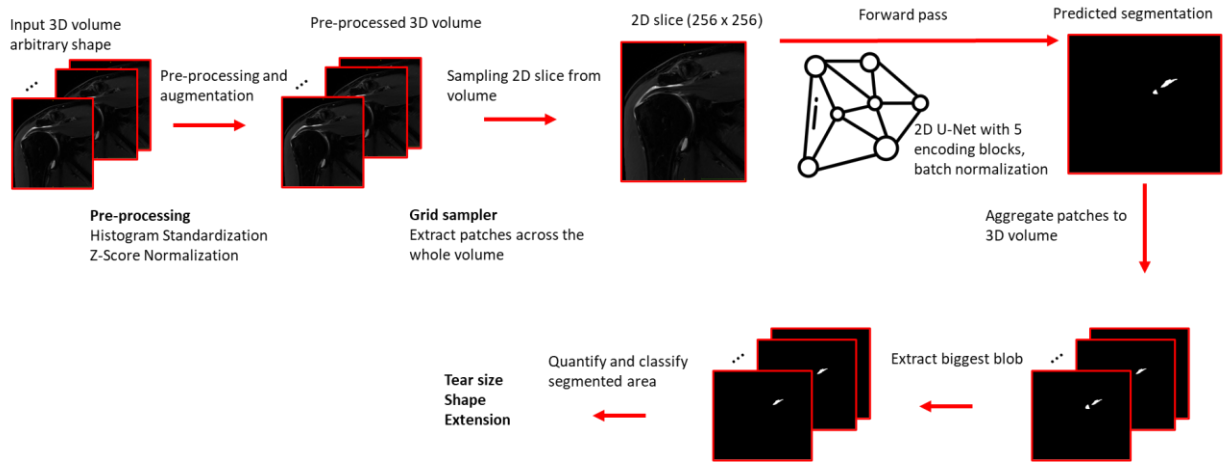

Figure 13: Illustration of the segmentation pipeline when performing a prediction with the trained neural network and postprocessing.

2.6 Manual Segmentation

To evaluate the manual segmentation 27 patients from the dataset were segmented by two surgeons. This allows to determine how difficult the segmentation of the a rotator cuff tear is. For this task only the images in coronal view were used. Images acquired in this view are of high interest, because it is possible to measure the retraction of a rotator cuff tear more accurately. As the retraction lies in plane with the coronal view, the formula (2.1) was used to calculate the retraction. The retraction is a more important indicator of the reparability of the tear, compared to the tear width. The width is not in plane with the coronal view, thus was calculated with the formula (2.2).

2.7 Pre- & Intraoperative Measurements

Since the patients in the dataset described in Figure 11 had already had their shoulder surgery it was impossible to measure pre- or intraoperatively. Therefore, for seven additional patients, the retraction and the tear width were measured in a prospective manner.
The conventional preoperative measurement was performed using the MRI images. To exclude measurement bias, these measurements were performed by a third surgeon.
The surgical tool for the intraoperative measurement was a probe with engravings similar to the one shown in Figure 14. As this probe is used to assess the integrity of anatomical structures, there was no

additional risk to the patient when measuring the size of rotator cuff tear. The pattern and the Patte grade were also determined during the surgery.

Figure 14: Probe for intraoperative measurements of the rotator cuff tear. [13]

2.8 Metrics

The segmentations were evaluated with the metrics described in this section. To make a comparison possible, these metrics were used for the manual segmentations as well as the automatic segmentations.

2.8.1 Dice Score

The dice score is a measure to evaluate the similarity of the segmented area and is defined as follows:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.3}$$

Where X is the first segmented volume and Y is the second segmented volume.

2.8.2 Absolute error of Euclidean distance

The absolute error of Euclidean distance is a measure to get the difference between two distances. The absolute error of Euclidean distance is defined as follows:

$$d_{Error} = \sqrt{(d_1 - d_2)^2} \tag{2.4}$$

2.8.3 Accuracy

The accuracy is a measure used for classification tasks. It states how many samples were classified right. The accuracy is defined as follows:

$$Accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \tag{2.5}$$

Where $\hat{y}$ is the true class of the sample, y is the predicted class of the sample.

2.8.4 Cohen kappa

The Cohen's kappa coefficient is a measure for interrater reliability of categorical classifications tasks. Cohen's kappa coefficient κ is defined as follows:

$$\kappa = \frac{p_0 - p_c}{1 - p_c} \tag{2.6}$$

Where p_0 relative observed agreement and p_c is the hypothetical probability of chance agreement. The results can be interpreted in the following way: poor agreement $\kappa < 0$, slight agreement $\kappa = [0, 0.2]$, fair agreement $\kappa =]0.2, 0.4]$, moderate agreement $\kappa =]0.4, 0.6]$, substantial agreement $\kappa =]0.6, 0.8]$, almost perfect agreement $\kappa =]0.8, 1]$.

2.9 Evaluation of Manual Segmentation (Interrater Reliability)

In this section the comparison of the manual segmentation of two surgeons is described. To compare the segmented volumes, the dice score was used. For the comparison of the distances of retraction and tear width, the absolute error of Euclidean distance was calculated. The accuracy and Cohens Kappa were calculated to evaluate the classifications.
To obtain the error of manual segmentation, the interrater reliability of the retraction and the tear width was determined. On the basis of the manually measured distances of retraction and tear width, the classification of the tear according to Snyder, the pattern and Patte was made for both surgeons and then evaluated.

2.10 Evaluation of Automatic Segmentation

In this section the comparison of the automatic segmentation and the ground truth is described. To compare the segmented volumes, the dice score was used. For the comparison of the distances of retraction and tear width, the absolute error of Euclidean distance was calculated. The accuracy and Cohens Kappa were calculated to evaluate the classifications.
To obtain the error of automatic segmentation, the distances of the retraction and the tear width was determined. On the basis of the manually and automatically measured distances of retraction and tear width, the classification of the tear according to Snyder, the pattern and Patte was made and then evaluated.

2.11 Comparison Between Manual and Automatic Segmentation

In this section the comparison of the manual segmentation with of the automatic segmentation is described. To compare the segmented volumes, the dice score was used. For the comparison of the distances of retraction and tear width, the absolute error of Euclidean distance was calculated. The accuracy and Cohens Kappa were calculated to evaluate the classifications.
The error of manual segmentation was obtained as described in 2.9.
To obtain the error of the automatic segmentation the same calculations and classifications were performed as for the error of the manual segmentation. Here, the measurements and classifications were taken between the automatic segmentation and the ground truth. The ground truth was segmented by one of the above surgeons.

2.12 Comparison Between Pre- & Intraoperative Measurements and Automatic Segmentation

In this section the comparison of the pre- and intraoperative measurements and the measurements from the automatic segmentation is described. For the comparison of the distances of retraction and tear width, the absolute error of Euclidean distance was calculated. No manual segmentation was performed preoperatively, thus it is not possible to calculate a dice score. The intraoperative classifications and the classifications from the automatic segmentations are evaluated with the accuracy and Cohens Kappa.

2.13 Visualization in Webapp

In order to make the automatic evaluation and classification of a rotator cuff tear accessible to surgeons, it was integrated into the existing web app Shouldeer [14]. The focus was on simple

usability and a high level of data security. Shouldeer was only used to visualize the results of this thesis.

3 Results

In this section, the results of the manual and automatic segmentation as well as the pre- and intraoperative measurements are presented. At the end of this section, the user interface of the web app is explained. To give a sense of what to expect from automatic segmentation, the results of manual segmentation are shown first.

3.1 Evaluation of Manual Segmentation (Interrater Reliability)

The results of manual segmentation were evaluated using the metrics presented in section 2.8 and 2.9. The segmented volumes are quantified by calculating the retraction and the tear width from the segmentation as described in section 2.1.1. Then, the differences of the obtained retraction and tear width are presented and finally the classification accuracies are shown. All discussed results in this section were obtained from 27 patients manually segmented in the coronal direction. The segmentations were held out by an orthopedic surgeon and compared to orthopedic surgeon who obtained the ground truth.

3.1.1 Segmentation Volumes

The results of the dice score between the two orthopedic surgeon are shown in Figure 15. The results of the individual patients are represented by dots in the graphic. A mean dice score of 0.50 with a standard deviation of 0.16 was achieved.

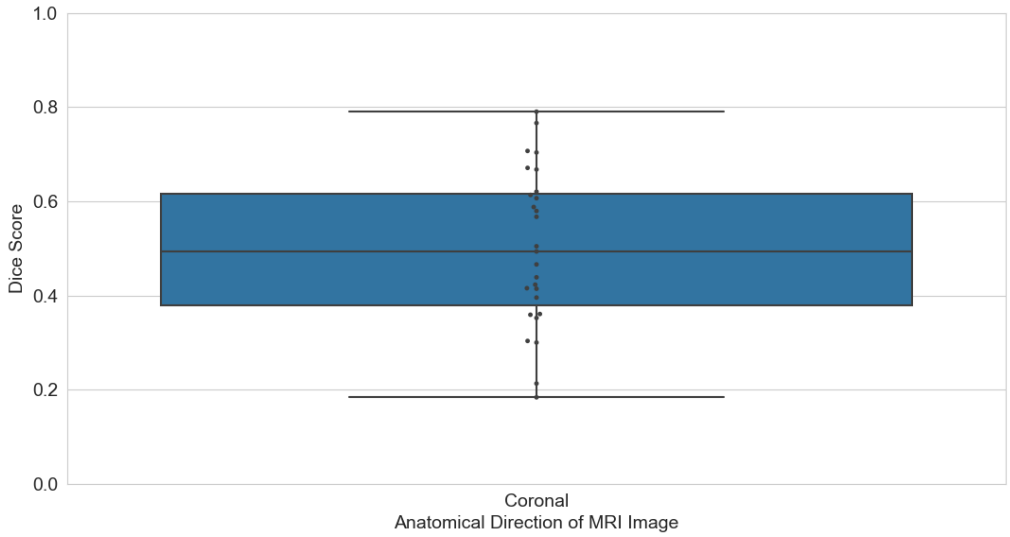

Figure 15: Boxplots of the dice score between the two surgeon. All MRI images were acquired in the coronal view.

3.1.2 Distances of Retraction and Tear Width

For the evaluation of the retraction and tear width, the differences of the distances between the two surgeons are shown in Figure 16. The two boxplots on the right show the absolute difference. The

retraction was determined in the image plane. Since the surgeon's segmentations were only in the coronal direction, the tear width was determined orthogonal to the image plane.

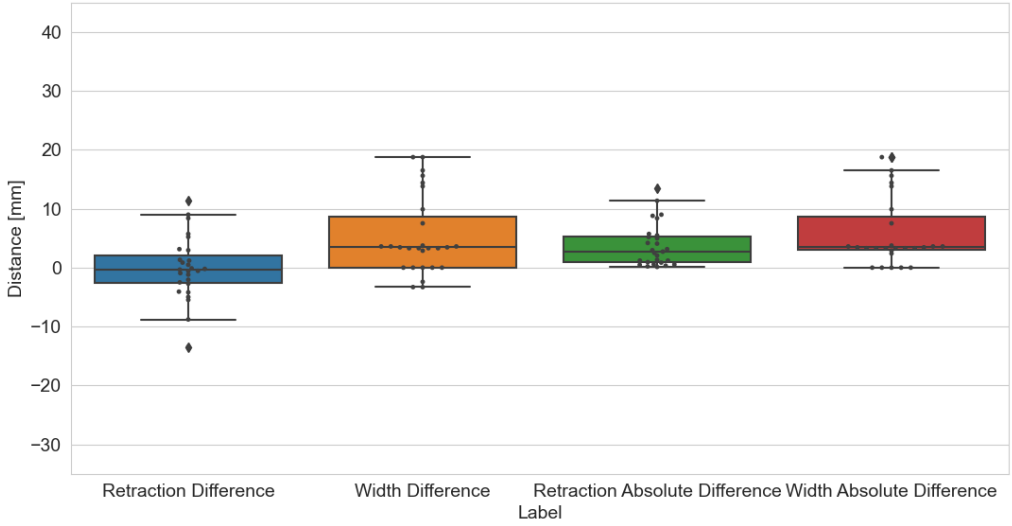

Figure 16: Boxplots of the difference of the retraction and the tear width between the two surgeons. Retraction was calculated in the image plane. The tear width was calculated orthogonal to the image plane.

Table 4 shows the mean value and the standard deviation for the same results as shown in the boxplots in Figure 15.

Table 4: Mean value and standard deviation for the difference and the absolute difference of the two surgeons.

Retraction difference	Width Difference	Retraction Absolute Difference	Width Absolute Difference
-0.093 mm ± 5.20 mm	5.32 mm ± 6.56 mm	3.77 mm ± 3.58 mm	6.00 mm ± 6.00 mm

The median and the mean of the retraction difference lies around 0 mm. The retraction difference compared to the width difference is significantly more accurate. The positive difference of the tear width indicates that the surgeon segmented the tear on average 5.32 mm smaller than the ground truth. The absolute difference of the retraction is 3.77 mm and 6.00 mm for the tear width. With a p-value of 0.1032 it is not significantly smaller.

3.1.3 Classification by Snyder, the Pattern and Patte

The classification of the manual segmentation according to Snyder, the pattern and Patte are represented in Table 5. The classification were acquired from the extracted retraction and tear width from both, the manual segmentation and the ground truth, as described in section 2.1.

Table 5: Results of the classification, with the corresponding accuracy and Cohens kappa. The retraction was extracted in the image plane and the tear width orthogonal to the image plane of a coronal MRI image.

Classification System		Classifications (Number of patients)	Accuracy	Cohens Kappa
Snyder	Ground truth	CI: 5 CII: 11 CIII: 8 CIV: 3	0.63	0.46
	Surgeon	CI: 7 CII: 12 CIII: 7 CV: 1		
Pattern	Ground truth	Crescent Shaped Tear: 2 Longitudinal Tear: 14 Massive Tear: 11	0.78	0.57
	Surgeon	Crescent Shaped Tear: 1 Longitudinal Tear: 20 Massive Tear: 6		
Patte	Ground truth	Grade I: 10 Grade II: 11 Grade III: 6	0.93	0.88
	Surgeon	Grade I: 9 Grade II: 13 Grade III: 5		

The Cohens kappa value for the Snyder classification is 0.46, indicating only moderate agreement in the lower range between the surgeon and the ground truth.
A Cohens kappa value of 0.57 is obtained for the classified patterns of the tear. This indicates moderate agreement close to the border of substantial agreement. The accuracy of this classification is 0.15 higher compared to the Snyder classification.
An almost perfect agreement with a Cohens kappa value of 0.88 and an accuracy of 0.93 was obtained for the Patte classification. For all classification tasks, the Patte classification was the most accurate in manual segmentation.

3.2 Evaluation of Automatic Segmentation

The results of automatic segmentation were evaluated in the same way as the results for the manual segmentation. This allowed a comparison of the results between manual and automatic segmentation. All discussed results in this section were obtained automatically from 57 patients automatically segmented in the coronal direction and from 46 patients automatically segmented in the sagittal direction. The automatic segmentations were done by the network pipeline described in 2.5. The results were compared to the ground truth.

3.2.1 Segmentation Volumes

Figure 17 shows boxplots of the dice score. The dice score of the coronal view was calculated for all 57 patients compared to the ground truth and the dice score of the sagittal view was calculated for all 46 patients having also a sagittal MRI image. The right boxplot shows the combined results.

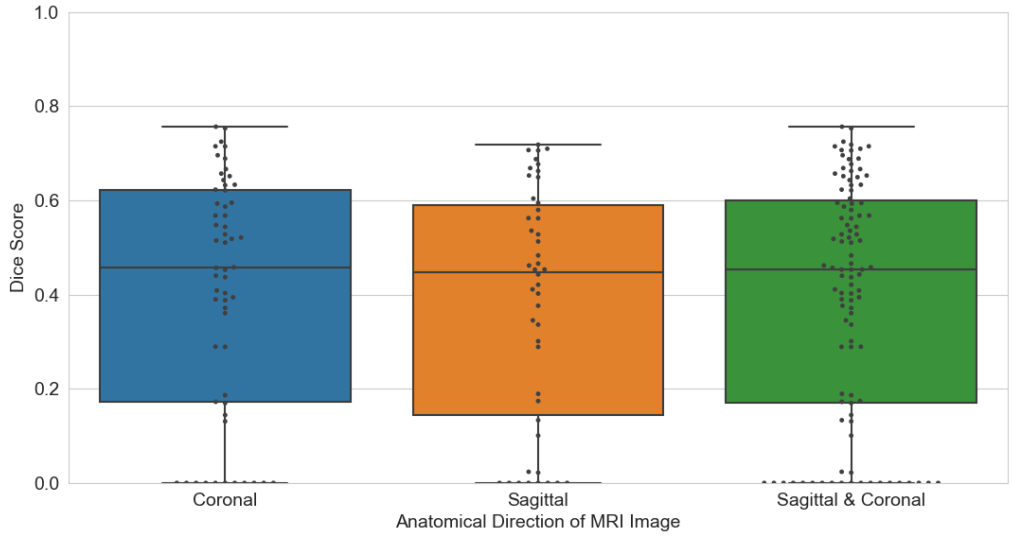

Figure 17: Boxplots of the dice score between the ground truth and the automatic segmentation. All MRI images were acquired in the coronal and sagittal direction. Additionally a boxplot of all the dice scores combined is shown on the right.

There are no significant differences in the dice scores related to the direction of the MRI image. As shown by the results of the boxplots and in Table 6, a dice score of 0.41 was obtained in the coronal direction and 0.38 in the sagittal direction. Automatic segmentation detected falsely no tear in 12 of 57 cases in the coronal direction and in 10 of 46 cases in the sagittal direction.

Table 6: Mean value and standard deviation for the dice score between automatic segmentation and the ground truth in the coronal and sagittal direction as well as both results combined.

Dice Coronal Direction	Dice Sagittal Direction	Dice Sagittal & Coronal Direction
0.41 ± 0.25	0.38 ± 0.25	0.40 ± 0.25

3.2.2 Distances Retraction and Tear Width

For the evaluation of retraction and tear width, the differences of the distances to the ground truth are shown in Figure 18 and Figure 19. Since automatic segmentation was performed on all coronal and sagittal acquired MRI images, it was possible to obtain the retraction and tear width in the image plane. To allow comparison with manual segmentation, tear width was also obtained from all coronal MRI images orthogonal to the image plane. Since the evaluation of the manual segmentation was only performed on coronal MRI images.

Figure 18 shows boxplots of the differences to the ground truth for all 57 coronal MRI images. The retraction was evaluated in plane and the tear width was evaluated orthogonal to the image plane.

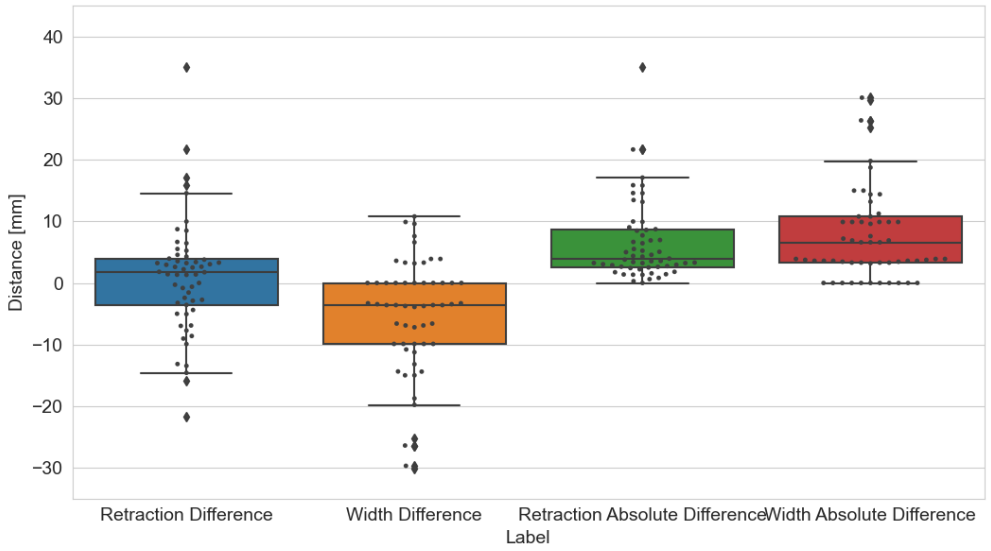

Figure 18: Boxplots of the difference of the retraction and the tear width between ground truth and automatic segmentation. The retraction was calculated in the image plane. The tear width was calculated orthogonal to the image plane.

The mean values of the same results as shown in Figure 18 are represented in Table 7. The mean value of the retraction difference is 0.93 mm with a standard deviation of 9.18 mm. The positive difference means that the automatic segmentation overestimates the retraction by 0.93 mm. The negative difference of -5.74 mm with a standard deviation of 9.65 mm indicates that the automatic segmentation underestimates the tear width by 5.74 mm.
Figure 19 shows boxplots of the differences from the ground truth. Retraction was calculated in the same way and based on the same MRI images as in Figure 18. The tear width was calculated from the 46 sagittal images of the dataset. The sagittal images allow the tear width to be calculated in the image plane as well.

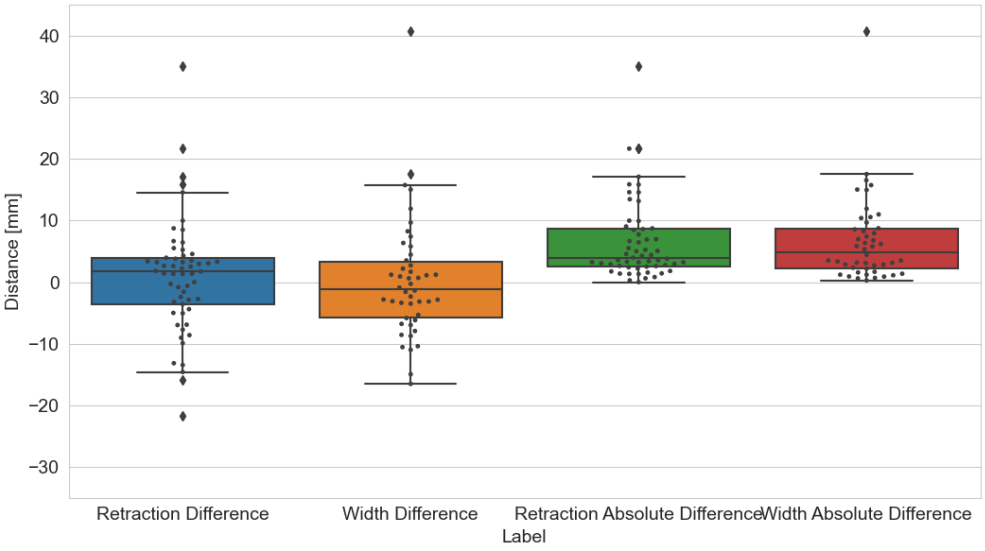

Figure 19: Boxplots of the difference of the retraction and the tear width between ground truth and automatic segmentation. The retraction and tear width were calculated in the image plane.

The mean values of the same results as shown in Figure 19 are represented in Table 7. The mean value of the retraction difference is the same as in Figure 18. Since, the calculation were the same as in Figure 18. The difference of the tear width is 0.23 mm with a standard deviation of 9.62 mm. Therefore the tear width is insignificantly underestimated.
The absolute difference of the calculated tear width in image plane is 1.38 mm smaller compared to the tear width calculated orthogonal to the image plane. The absolute difference of the tear width compared to the absolute difference of the retraction is 0.11 mm and thus insignificant higher if both are calculated in the image plane.

Table 7: Mean value and standard deviation for the difference and the absolute difference of the automatic segmentation to the ground truth.

	Retraction Difference	Width Difference	Retraction Absolute Difference	Width Absolute Difference
Retraction in plane / Width orthogonal	0.93 mm ± 9.18 mm	-5.74 mm ± 9.65 mm	6.56 mm ± 6.48 mm	8.05 mm ± 7.83 mm
Retraction and Width in plane	0.93 mm ± 9.18 mm	0.23 mm ± 9.62 mm	6.56 mm ± 6.48 mm	6.67 mm ± 6.93 mm

3.2.3 Classifications by Snyder, the Pattern and Patte

The classification for the automatic segmentation according to Snyder, the pattern and Patte are represented in Table 8 and Table 9. The classifications were acquired form the extracted retraction and tear width from the automatic segmentation and the ground truth as described in section 2.1.
The classification was once done from automatic segmentation of only coronal MRI images and once from coronal and sagittal MRI images from the same patient. To be able to compare the results of the automatic segmentation to the results of the manual segmentation the classification must be applied in the same way. To estimate the most accurate segmentation the paired coronal and sagittal automatic segmentation were considered. Table 8 shows the results of the classifications obtained only from coronal MRI images. In this case, the retraction was calculated in the image plane and the tear width was calculated orthogonal to the image plane.

Table 8: Results of the classification, with the corresponding accuracy and Cohens kappa. The retraction was extracted in the image plane and tear width orthogonal to the image plane from a coronal MRI image.

Classification System		Classifications (Number patients)	Accuracy	Cohens Kappa
Snyder	Ground truth	CI: 15 CII: 26 CIII: 12 CIV: 4	0.61	0.41
	Automatic segmentation	CI: 10 CII: 33 CIII: 8 CV: 6		
Pattern	Ground truth	Crescent Shaped Tear: 5 Longitudinal Tear: 36 Massive Tear: 16	0.58	0.34
	Automatic segmentation	Crescent Shaped Tear: 13 Longitudinal Tear: 20 Massive Tear: 24		
Patte	Ground truth	Grade I: 32 Grade II: 21 Grade III: 4	0.75	0.54
	Automatic segmentation	Grade I: 31 Grade II: 25 Grade III: 1		

The kappa value for the Snyder classification is 0.41, indicating only moderate agreement in the lower range between the automatic segmentation and the ground truth.
A kappa value of 0.34 is obtained for the classified patterns of the tear. This indicates fair agreement. The accuracy of this classification is 0.03 lower compared to the Snyder classification.
A moderate agreement with a kappa value of 0.54 and an accuracy of 0.75 was obtained for the Patte classification. For all classification tasks, the Patte classification was the most accurate in automatic segmentation from only coronal MRI images.
Table 9 shows the results of the classifications obtained from paired coronal and sagittal MRI images. This has the consequence that the retraction and the tear width was calculated in the image plane.

Table 9: Results of the classification, with the corresponding accuracy and Cohens kappa. The retraction and tear width for the classification were extracted in plane from paired coronal and sagittal MRI images.

Classification System		Classifications (Number patients)	Accuracy	Cohens Kappa
Snyder	Ground truth	CI: 9 CII: 19 CIII: 7 CIV: 4	0.56	0.33
	Automatic segmentation	CI: 7 CII: 21 CIII: 9 CV: 2		
Pattern	Ground truth	Crescent Shaped Tear: 3 Longitudinal Tear: 21 Massive Tear: 15	0.64	0.41
	Automatic segmentation	Crescent Shaped Tear: 8 Longitudinal Tear: 17 Massive Tear: 14		
Patte	Ground truth	Grade I: 21 Grade II: 16 Grade III: 2	0.82	0.67
	Automatic segmentation	Grade I: 19 Grade II: 19 Grade III: 1		

The Cohens kappa value for the Snyder classification is 0.33, indicating only fair agreement in the lower range between the automatic segmentation and the ground truth.
A Cohens kappa value of 0.41 is obtained for the classified patterns of the tear. This indicates moderate agreement. The accuracy of this classification is 0.08 higher compared to the Snyder classification.
A moderate agreement with a Cohens kappa value of 0.67 and an accuracy of 0.82 was obtained for the Patte classification. For all classification tasks, the Patte classification was the most accurate in automatic segmentation from only coronal MRI images.

3.3 Comparison Between Manual (Interrater Reliability) and Automatic Segmentation

In this section the results of the manual and automatic segmentations are compared. To make the comparison as meaningful as possible only the results which were obtained in the same way were compared. Since the manual segmentation was only done on 27 coronal MRI images all the results of

the automatic segmentation were calculated on coronal images as well. The 27 images of the manual segmentation was a subset of all 57 coronal MRI images used to evaluate the automatic segmentation.

3.3.1 Segmentation Volumes

Figure 20 shows boxplots of the dice score of the manual and automatic segmentations. The dice scores were calculated by comparing the results to the ground truth.

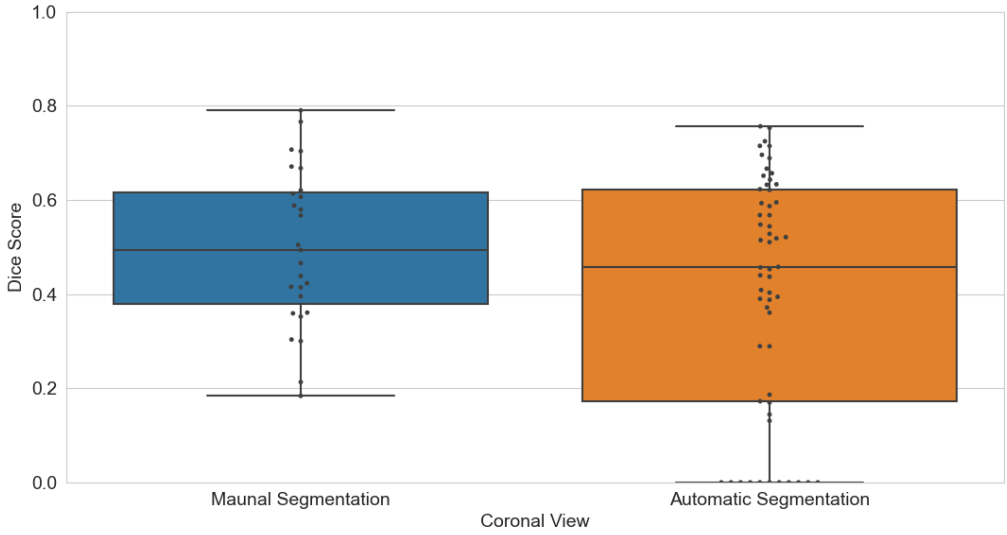

Figure 20: Boxplots of the dice score for the manual segmentation and the automatic segmentation compared to the ground truth which is obtained by one of the surgeons. The dice scores were only calculated in the coronal view.

All the results of Figure 20 are summarized in Table 10. The mean dice score of the manual segmentation is 0.50 with a standard deviation of 0.16. The mean dice score of the automatic segmentation is 0.41 with a standard deviation of 0.25. The mean manual dice score is 0.09 higher compared to the automatic segmentation and the standard deviation is 0.09 lower.

Table 10: Mean value and standard deviation for the dice score of the manual and automatic segmentation. Both are compared to the ground truth which is obtained by one of the surgeons and segmentation in coronal MRI images.

Dice Manual Segmentation	Dice Automatic Segmentation
0.50 ± 0.16	0.41 ± 0.25

3.3.2 Distances Retraction and Tear Width

Figure 21 shows the absolute differences of the retraction and the tear width for the manual and automatic segmentation. The retraction is calculated in the image plane and the tear width is calculated orthogonal to the image plane.

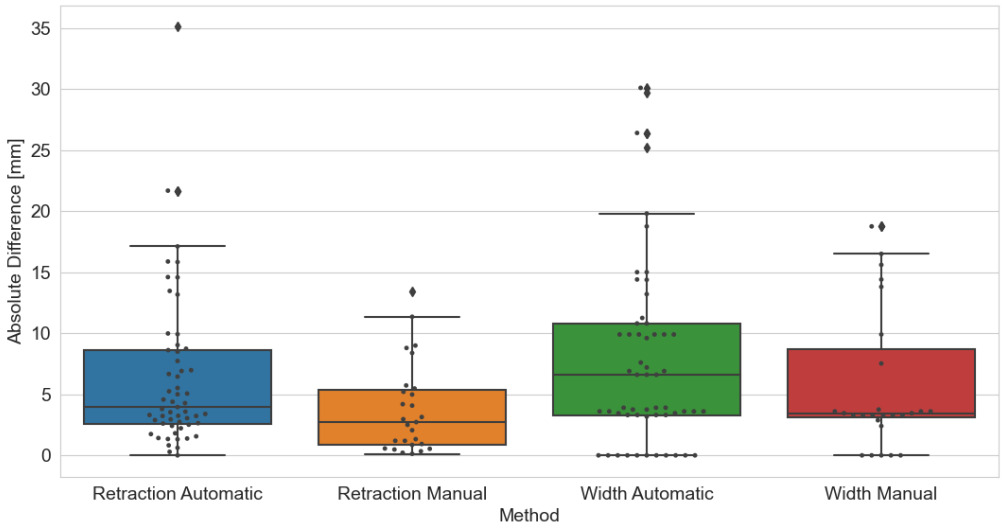

Figure 21: Boxplots of the difference of the retraction and the tear width between ground truth and automatic segmentation and between the two surgeons. The retraction was calculated in the image plane and the tear width orthogonal to the image plane.

All results of Figure 20Figure 21 are summarized in Table 11. The mean absolute difference for the retraction of the automatic segmentation is 6.56 mm with a standard deviation of 6.48 mm. The mean absolute difference for the retraction of the manual segmentation is 3.77 mm with a standard deviation of 3.58 mm. Compared to the automatic segmentation, is the manual segmentation 2.79 mm more accurate and the standard deviation is 2.90 mm lower. Similar results can be obtained for the tear width. The mean absolute difference for the automatic segmentation is 8.05 mm with a standard deviation of 7.83 mm. The mean absolute difference for the tear width of the manual segmentation is 6.00 mm with a standard deviation of 6.00 mm. Compared to the automatic segmentation, is the manual segmentation 2.05 mm more accurate and the standard is 1.83 mm lower.

Table 11: Mean value and standard deviation for the absolute difference of the automatic segmentation and the manual segmentation compared to the ground truth which is obtained by one of the surgeons.

Retraction Automatic	Retraction Manual	Width Automatic	Width Manual
6.56 mm ± 6.48 mm	3.77 mm ± 3.58 mm	8.05 mm ± 7.83 mm	6.00 mm ± 6.00 mm

3.3.3 Classifications by Snyder, the Pattern and Patte

The comparison of the classification results between the manual and automatic segmentation are shown in Table 12. All results for the manual and automatic segmentation were obtained in the same way.

The Snyder classification has the smallest difference between the automatic and manual segmentation with 0.02 in the accuracy and 0.05 in the kappa value.

The pattern classification shows a significant difference between automatic and manual segmentation. The Cohens kappa value is 0.23 lower for the automatic segmentation, which is a reduction from a moderate agreement to a fair agreement.

Also for the Patte classification, is the agreement lower in the automatic classification. It is reduced from an almost perfect agreement by a moderate agreement.

Table 12: Results of the classification for the automatic and manual segmentation, with the corresponding accuracy and Cohens kappa. The retraction was extracted in the image plane and the tear width orthogonal to the image plane from coronal MRI images.

Classification System	Metric	Manual	Automatic	Difference
Snyder	Accuracy	0.63	0.61	0.02
	Cohens Kappa	0.46	0.41	0.05
Pattern	Accuracy	0.78	0.58	0.2
	Cohens Kappa	0.57	0.34	0.23
Patte	Accuracy	0.93	0.75	0.18
	Cohens Kappa	0.88	0.54	0.34

3.4 Comparison Between Pre- & Intraoperative Measurements and Automatic Segmentation

For a further evaluation of the accuracy of the automatic segmentation, additionally pre- and intraoperative measurements were obtained. The preoperative measurements were performed on a coronal MRI image of seven patients by an orthopedic surgeon. The intraoperative measurements were obtained by a second surgeon as described in section 2.7. In this way the measurements are not biased. The measurements of the automatic segmentation were only calculated from the coronal MRI image of these patients. All the results are shown in Table 13. If no measurement was obtained it is indicated with a "-".

Table 13: Comparison of pre- & intraoperative measurements and automatic segmentation for retraction, tear width, pattern and Patte classification.

Patient		Retraction	Width	Pattern	Patte
502	Preoperative	41.78	39.6	-	-
	Intraoperative	25	40	Massive	Grade III
	Automatic	42.13	36.00	Massive	Grade II
503	Preoperative	16.12	19.8	-	-
	Intraoperative	15	18	rL	Grade II
	Automatic	16.7	19.8	Crescent	Grade I
504	Preoperative	26.86	32.4	-	-
	Intraoperative	26	37	Massive	Grade II
	Automatic	24.86	32.40	Massive	Grade II
505	Preoperative	38.44	33	-	-
	Intraoperative	22	32	Massive	Grade II
	Automatic	37.52	26.40	Massive	Grade II
507	Preoperative	-	-	-	-
	Intraoperative	23	33	Massive	Grade III
	Automatic	43.98	39.60	Massive	Grade II
508	Preoperative	-	-	-	-
	Intraoperative	14	17	Massive	Grade II
	Automatic	18.92	24.15	Crescent	Grade I

The classification of the tear pattern differs two times to the intraoperative classification. In the Patte classification only two times are the same classes obtained.
The retraction measured manually preoperatively has a difference of less than 2 mm compared to all automatic preoperative measured retractions. For the tear width the results are similar. Only one case had a higher difference than one slice thickness of 3.6 mm.

The highest difference from the automatic segmentation to the intraoperative measurements for the retraction is 20.98 mm. The smallest difference is 1.14 mm. The mean difference is 10.23 mm.
Compared between the pre- & intraoperative measurements, the highest difference of the retraction is 16.78 mm and the smallest difference is 0.58 mm. The mean difference is 5.07 mm.
The highest difference from the automatic segmentation to the intraoperative measurements for the tear width is 7.15 mm and the smallest difference is 1.8 mm. The mean difference is 5.0 mm.
Compared between the pre- & intraoperative measurements, the highest difference of the retraction is 5.4 mm and the smallest difference is 0.4 mm. The mean difference is 2.08 mm.

3.5 Visualization in Webapp

The user interface of the webapp *Shouldeer* is shown in Figure 22. All anatomical views of the MRI images are shown in the window with a red, green and blue border. The segmentation of the tear is visualized in all the anatomical views. For a better spatial context of the tear a 3D volume is represented on the window with the yellow border. The results of the retraction, tear width and classifications are shown on the right side.

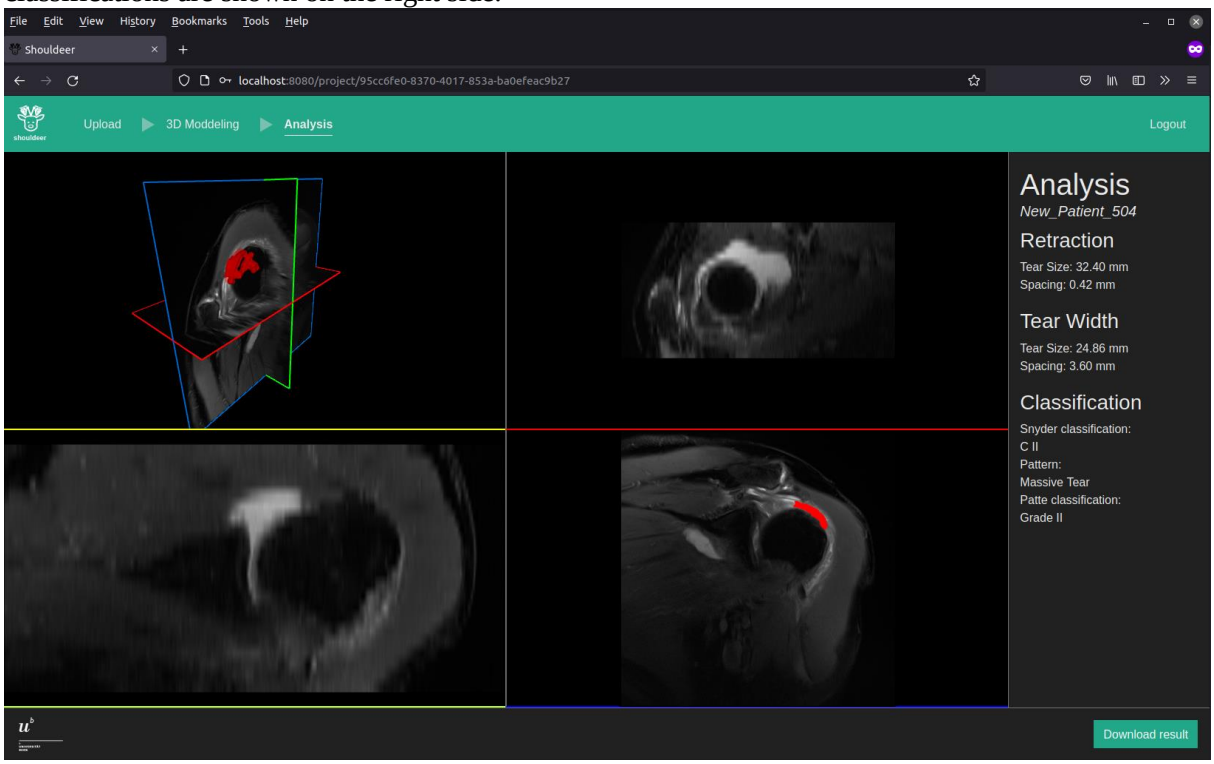

Figure 22: User interface of the webapp *Shouldeer* for the quantification of a rotator cuff tear.

4 Discussion and Conclusions

In the discussion the main findings are pointed out and summarized. In general the results show that an automatic quantification and classification of a rotator cuff tear is possible. Its comparison to the interrater reliability shows that some improvements are still necessary to be able to use it in a clinical environment.

4.1 Discussion

In this section all the results are discussed and the main foundings are pointed out for each conducted comparison and experiment.

4.1.1 Evaluation of Manual Segmentation (Interrater Reliability)

A dice score of 0.5 is quite low compared to other segmentations tasks in the medical field. As an example, for bone segmentation a dice score of 0.95 is realistic [5]. This indicates that the task of segmenting a rotator cuff tear is more difficult, even for clinical experts. This thesis was the first attempt to quantify a rotator cuff tear from a segmentation. To get a higher dice score between two clinical experts a clear instruction of the segmentation task must be elaborated.
Despite the low dice score, the retraction and the tear width could be determined accurately between two specialists. The results also show that, whenever the tear size is quantified in the image plane, it becomes more accurate.
The highest agreement was achieved in the Patte classification. To determine the grade of a tear in the Patte classification, only the retraction is used. Once the retraction and the tear width are used to classify a rotator cuff tear, the classification becomes less accurate because another source of error is added.

4.1.2 Evaluation of Automatic Segmentation

The dice score for automatic segmentation was 0.4. Additionally, the automatic segmentation falsely did not detect the rotator cuff tear for some patients. The low dice score is probably due to the high variance and the relatively small size of the dataset. The smaller the structure to be segmented, the larger the data set must be in order to be able to map all the characteristics that occur in the different rotator cuff tears.
Automatic segmentation was shown to work equally well in coronal as in sagittal MRI images, even though a rotator cuff tear does not look the same in the two anatomical directions.
When calculating both, the retraction and the tear width, in the image plane, the accuracy is increased for both results. Therefore, it is recommended to determine the retraction always from coronal images and the tear width always from sagittal images in order to obtain the most accurate result. Particularly, this is due to the high slice thickness of the MRI images.
In the classifications, the accuracy could be increased when retraction and tear width were extracted from images acquired in coronal and also sagittal view from the same patient.

4.1.3 Comparison Between Manual and Automatic Segmentation

As expected, the dice score of the automatic segmentation is 0.1 lower than that of the manual segmentation. Relatively many rotator cuff tears were not detected by automatic segmentation. This

never occurred with manual segmentation. Probably because the surgeon already knew that it must be a rotator cuff tear.
The retraction was on average 2.71 mm less accurate in the automatic segmentation compared to the manual segmentation. With a mean retraction of 20.3 mm in the entire dataset, the difference of 2.71 mm represents an error of 13% of the total retraction. In the tear width, the mean difference between the automatic and the manual segmentation was 2.02 mm. This is because the tear widths were determined orthogonally to the image plane and thus both committed the same errors. It is also noticeable that the standard deviation in the automatic segmentation is higher and therefore the result is less reliable.
Since the classification is determined from the retraction and the tear width, its error is propagated from the segmentation error and the calculation error of the retraction and the tear width.

4.1.4 Comparison Between Pre- & Intraoperative Measurements and Automatic Segmentation

The intraoperative measurements provide information on how accurately retraction and tear width can be determined manually and automatically from MRI images. Preoperatively, the retraction and tear width can be determined automatically as accurate as a surgeon does. For all patients, the preoperative measurements agreed very well with the automatic segmentation results. For the automatic segmentation, half of the results agreed very well with the intraoperative measurements. However, it is not clear which measurement is the most accurate. The intraoperative measurements depends on who is measuring.
A dataset of six patients is too small for a deep analysis of the results. However, the trends show the possibility of tear measurements from MRI images. With a larger dataset to train the neural network, automatic segmentation can be brought to a similar level as manual segmentation.

4.1.5 Visualization in Webapp

Consultations and demonstrations with various shoulder surgeons showed that a web app like *Shouldeer* can be a great help in planning a rotator cuff tear procedure. The actual accuracy of the sizes of the rotator cuff tears needs to be further determined together with clinical partners. *Shouldeer* provides a concrete example of what a final solution could look like and how simple and user-friendly it is.

4.2 Conclusions

A fully automatic quantification and classification of a rotator cuff tear is possible. Compared to related work [22, 30] this approach is completely new due to the segmentation of the tear. This allows the quantification of the retraction and tear width by its distance in millimeters. This is not possible when the tear is only classified by a neural network and not segmented. An advantage of this approach is that the tear can be represented as a 3D volume. This can help the surgeons to assess the tear shape more accurate and before the rotator cuff repair.
A quantification and classification of a rotator cuff tear can help choosing the right treatment option and to plan the surgery more precisely. This is important, because nowadays, many decisions such as the sutures technique and the exact placing of the anchors are made during the surgery.
Due to the high variability in the MRI images and the relatively small dataset this thesis is a proof of concept and not a final solution that could be implemented in a preoperative software.
The webapp enables a further discussion to define the necessary accuracy and the exact use of a preoperative quantification and classification of a rotator cuff tear further.

5 Outlook

The accuracy of the quantified measures can probably be further improved. Since all quantifications and classifications rely on the automatic segmentation of the neural network, this should be the first point addressed. A larger dataset can better capture the high variance of a rotator cuff tear. Inclusion of similar MRI sequences such as T2w and PD Blade fs would increase the dataset. However, a good balance must be found in order to not increase the variance of the data unnecessarily. Furthermore, it must be ensured that the manual segmentation of the ground truth is exactly defined. There must be no doubt about what needs to be segmented.
Narrowing down the field of views during segmentation could be beneficial for automatic segmentation. For this, a landmark must be detected and the volume around the landmark must be cropped.
For a further clinical evaluation more intraoperative measurements are necessary. At the moment it is not sure which measurement is the most accurate, since there were high differences between the pre- and intraoperative measurements. Also, the extracted classifications according to the tear size should be validated by a clinical expert. This allows a more precise classification algorithm for the extensions, the pattern and the retraction.
An additional approach is to use the automatic segmentation as a first suggestion in a semi-automatic tool. In this way the preoperative quantification of a rotator cuff tear can be easier and faster to use.

6 References

[1] Abrams, J. S. and Bell, R. H. 2008. *Arthroscopic rotator cuff surgery*. *A practical approach to management*. Springer, New York.

[2] Anirudh Chawla. 2017. Ultrasound, MRI and Arthroscopic Correlation of Rotator Cuff Tears. *International Journal of Contemporary Medical Research* 4, 3, 650.

[3] Arce, G., Bak, K., Shea, K. P., Savoie III, F., Kibler, W. B., Itoi, E., Mazzocca, A. D., Beitzel, K., Calvo, E., and Ejnisman, B. 2013. *Shoulder Concepts 2013: Consensus and Concerns*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[4] Bozkurt, M. and Açar, H. İ. 2017. *Clinical Anatomy of the Shoulder*. Springer International Publishing, Cham.

[5] Chea, P. and Mandell, J. C. 2020. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal radiology* 49, 2, 183–197.

[6] Chillemi, C., Castagna, A., and Osimani, M. 2018. *Arthroscopic Transosseous Rotator Cuff Repair*. Springer International Publishing, Cham.

[7] Collin, P., Matsumura, N., Lädermann, A., Denard, P. J., and Walch, G. 2014. Relationship between massive chronic rotator cuff tear pattern and loss of active shoulder range of motion. *Journal of shoulder and elbow surgery* 23, 8, 1195–1202.

[8] Dang, A. and Davies, M. 2018. Rotator Cuff Disease: Treatment Options and Considerations. *Sports medicine and arthroscopy review* 26, 3, 129–133.

[9] Davidson, J. and Burkhart, S. S. 2009. The geometric classification of rotator cuff tears: a system linking tear pattern to treatment and prognosis. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 26, 3, 417–424.

[10] Davidson, J. and Burkhart, S. S. 2010. The geometric classification of rotator cuff tears: a system linking tear pattern to treatment and prognosis. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 26, 3, 417–424.

[11] Grant, J. A., Miller, B. S., Jacobson, J. A., Morag, Y., Bedi, A., and Carpenter, J. E. 2013. Intra- and inter-rater reliability of the detection of tears of the supraspinatus central tendon on MRI by shoulder surgeons. *Journal of shoulder and elbow surgery* 22, 6, 725–731.

[12] Greenspoon, J. A., Petri, M., Warth, R. J., and Millett, P. J. 2015. Massive rotator cuff tears: pathomechanics, current treatment options, and clinical outcomes. *Journal of shoulder and elbow surgery* 24, 9, 1493–1505.

[13] Gumina, S. 2017. *Rotator Cuff Tear*. Springer International Publishing, Cham.

[14] Herren, M. 2021. Deep learning-based segmentation and fat fraction analysis of the shoulder muscles using quantitative MRI.

[15] Hertel, R., Knothe, U., and Ballmer, F. T. 2002. Geometry of the proximal humerus and implications for prosthetic design. *Journal of shoulder and elbow surgery* 11, 4, 331–338.

[16] Ian K. Y. Lo, MD, and FRCSC and Stephen S. Burkhart. Current Concepts in Arthroscopic Rotator Cuff Repair.

[17] Keener, J. D., Steger-May, K., Stobbs, G., and Yamaguchi, K. 2010. Asymptomatic rotator cuff tears: patient demographics and baseline shoulder function. *Journal of shoulder and elbow surgery* 19, 8, 1191–1198.

[18] Lädermann, A., Burkhart, S. S., Hoffmeyer, P., Neyton, L., Collin, P., Yates, E., and Denard, P. J. 2016. Classification of full-thickness rotator cuff lesions: a review. *EFORT open reviews* 1, 12, 420–430.

[19] Lippe, J., Spang, J. T., Leger, R. R., Arciero, R. A., Mazzocca, A. D., and Shea, K. P. 2012. Inter-rater agreement of the Goutallier, Patte, and Warner classification scores using preoperative magnetic resonance imaging in patients with rotator cuff tears. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 28, 2, 154–159.

[20] Longo, U. G., Vasta, S., Maffulli, N., and Denaro, V. 2011. Scoring systems for the functional assessment of patients with rotator cuff pathology. *Sports medicine and arthroscopy review* 19, 3, 310–320.

[21] Medina, G., Buckless, C. G., Thomasson, E., Oh, L. S., and Torriani, M. 2021. Deep learning method for segmentation of rotator cuff muscles on MR images. *Skeletal radiology* 50, 4, 683–692.

[22] Mijung Kim, Ho-min Park, Jae Yoon Kim, Seong Hwan Kim, Sofie Van Hoeke, and Wesley De Neve. 2020. MRI-based Diagnosis of Rotator Cuff Tears using Deep Learning and Weighted Linear Combinations.

[23] Milner, G. R. and Boldsen, J. L. 2012. Humeral and femoral head diameters in recent white American skeletons. *Journal of forensic sciences* 57, 1, 35–40.

[24] Narayana, P. A., Coronado, I., Sujit, S. J., Wolinsky, J. S., Lublin, F. D., and Gabr, R. E. 2020. Deep-Learning-Based Neural Tissue Segmentation of MRI in Multiple Sclerosis: Effect of Training Set Size. *Journal of magnetic resonance imaging : JMRI* 51, 5, 1487–1496.

[25] Nyúl, L. G., Udupa, J. K., and Zhang, X. 2000. New variants of a method of MRI scale standardization. *IEEE transactions on medical imaging* 19, 2, 143–150.

[26] Oliva, F., Piccirilli, E., Bossa, M., Via, A. G., Colombo, A., Chillemi, C., Gasparre, G., Pellicciari, L., Franceschetti, E., Rugiero, C., Scialdoni, A., Vittadini, F., Brancaccio, P., Creta, D., Del Buono, A., Garofalo, R., Franceschi, F., Frizziero, A., Mahmoud, A., Merolla, G., Nicoletti, S., Spoliti, M., Osti, L., Padulo, J., Portinaro, N., Tajana, G., Castagna, A., Foti, C., Masiero, S., Porcellini, G., Tarantino, U., and Maffulli, N. 2015. I.S.Mu.L.T - Rotator Cuff Tears Guidelines. *Muscles, ligaments and tendons journal* 5, 4, 227–263.

[27] Peter J. Millett, Ryan J. Warth. 2014. Posterosuperior Rotator Cuff Tears. Classification, Pattern, Recognition, and Treatment.

[28] Ro, K., Kim, J. Y., Park, H., Cho, B. H., Kim, I. Y., Shim, S. B., Choi, I. Y., and Yoo, J. C. 2021. Deep-learning framework and computer assisted fatty infiltration analysis for the supraspinatus muscle in MRI. *Scientific reports* 11, 1, 15065.

[29] Ronneberger, O., Fischer, P., and Brox, T. 2015. *U-Net: Convolutional Networks for Biomedical Image Segmentation*.

[30] Shim, E., Kim, J. Y., Yoon, J. P., Ki, S.-Y., Lho, T., Kim, Y., and Chung, S. W. 2020. Automated rotator cuff tear classification using 3D convolutional neural network. *Scientific reports* 10, 1, 15632.

[31] Via, A. G., Cupis, M. de, Spoliti, M., and Oliva, F. 2013. Clinical and biological aspects of rotator cuff tears. *Muscles, ligaments and tendons journal* 3, 2, 70–79.

[32] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. 2015. *Learning Deep Features for Discriminative Localization*.

[33] Zoga, A. C., Kamel, S. I., Hynes, J. P., Kavanagh, E. C., O'Connor, P. J., and Forster, B. B. 2021. The Evolving Roles of MRI and Ultrasound in First-Line Imaging of Rotator Cuff Injuries. *AJR. American journal of roentgenology*.