The Journal of Arthroplasty 31 (2016) 64–69

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

# Evaluation of Automated Volumetric Cartilage Quantification for Hip Preservation Surgery

Austin J. Ramme, MD, PhD [a], Michael S. Guss, MD [a], Shaleen Vira, MD [a], Jonathan M. Vigdorchik, MD [a], Axel Newe [b,c], Esther Raithel, PhD [d], Gregory Chang, MD [e]

[a] Department of Orthopaedic Surgery, New York University Hospital for Joint Diseases, New York, New York
[b] Methodpark Engineering GmbH, Erlangen, Germany
[c] Chair of Medical Informatics, Friedrich-Alexander University, Erlangen-Nuremberg, Erlangen, Germany
[d] Siemens Healthcare GmbH, Erlangen, Germany
[e] Department of Radiology, Center for Musculoskeletal Care, NYU Langone Medical Center, New York, New York

ARTICLE INFO

*Article history:*
Received 15 May 2015
Accepted 10 August 2015

*Keywords:*
magnetic resonance imaging
automated segmentation
cartilage evaluation
variability
interrater

ABSTRACT

Automating the process of femoroacetabular cartilage identification from magnetic resonance imaging (MRI) images has important implications to guiding clinical care by providing a temporal metric that allows for optimizing the timing for joint preservation surgery. In this paper, we evaluate a new automated cartilage segmentation method using a time trial, segmented volume comparison, overlap metrics, and Euclidean distance mapping. We report interrater overlap metrics using the true fast imaging with steady-state precession MRI sequence of 0.874, 0.546, and 0.704 for the total overlap, union overlap, and mean overlap, respectively. This method was 3.28× faster than manual segmentation. This technique provides clinicians with volumetric cartilage information that is useful for optimizing the timing for joint preservation procedures.



Modern treatments of hip and knee osteoarthritis have focused on symptomatic relief until the point at which a total joint arthroplasty is performed. Femoroacetabular impingement, hip dysplasia, Perthes disease, and slipped capital femoral epiphysis have all been associated with the onset of degenerative osteoarthritis [1-3]. Joint preservation surgery aims to prevent or delay the onset of hip arthritis by modifying the mechanical factors that contribute to this disease [3]. A contraindication to performing hip preservation surgery is the presence of underlying osteoarthritis, which is predictive of failure of a hip preservation procedure.

The quantification of cartilage volume is one method to identify early stage osteoarthritis before radiographic, macroscopic bony changes. The primary imaging technique used to evaluate cartilage is magnetic resonance imaging (MRI) due to the high water and proteoglycan content of this tissue. Various methods have been developed for automated cartilage identification of the knee [4-7], but few methods have been developed for cartilage of the hip. This is due to difficulty in hip MRI based on its deep position and thin cartilage layer that overlie a spherical surface [3]. Furthermore, the magnetic resonance (MR) signal given from cartilage can closely resemble that of synovial fluid making cartilage difficult to visualize [5].

Manual segmentation (slice-by-slice tracing) of hip cartilage by a trained technician remains the gold standard for hip cartilage identification [8]. Nishii et al [9] have published a method for automatically determining the average cartilage thickness from spoiled gradient echo MR images using Hough transforms. Li et al [10] have presented a semiautomated method to quantify thickness and cartilage volume using active contours and a multiple-echo data image combination MR sequence with axial rotational acquisition. Fully automated methods of hip cartilage segmentation are still being developed. Various techniques have been attempted including thresholding [11], multitemplate-based label fusion [12], graph searching [13], and deformable models [14].

Automating the process of hip cartilage segmentation has important implications to guiding clinical care by providing a volumetric cartilage measure, which can be followed over time to support a patient's symptoms and timing of joint preservation surgery. One segmentation algorithm described by Chandra et al [14,15] uses a combination of tissue classification, shape models, and statistical model–based segmentation to segment the bone and cartilaginous surfaces; for technical details regarding the algorithm, we refer the reader to Chandra et al. In this paper, we evaluate this segmentation method in the hip joint on a patient population; this would be highly useful for clinical applications such as hip preservation surgery. Two MRI sequences commonly used at our institution for cartilage evaluation were applied in this study for a rigorous evaluation of the segmentation technique. We hypothesize that the

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to http://dx.doi.org/10.1016/j.arth.2015.08.009.

Reprint requests: Gregory Chang, MD, NYU Langone Medical Center, 660 First Ave, 3rd Floor, Room 333, New York, NY 10016.

http://dx.doi.org/10.1016/j.arth.2015.08.009
0883-5403/

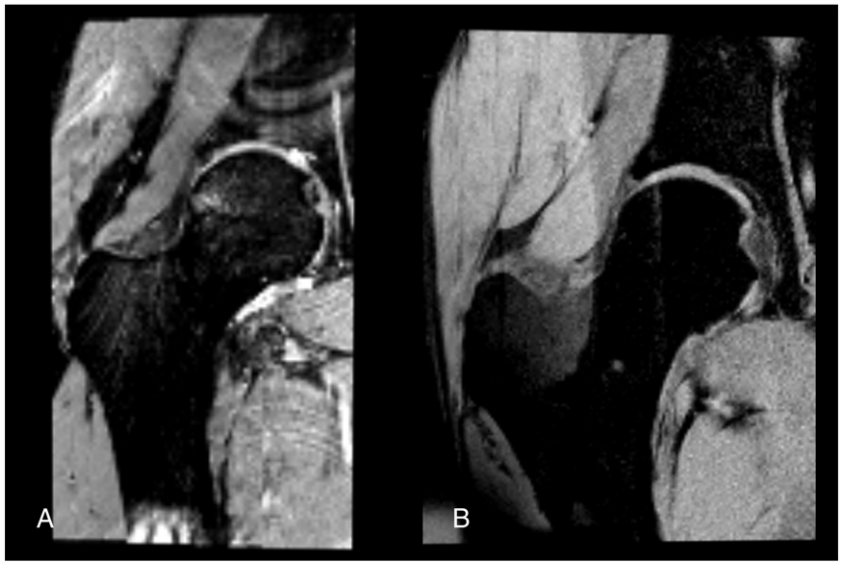


**Fig. 1.** Sample hip MRI image sequences used in the study including True FISP (A) and GRE (B).

image sequence that provides the best contrast between cartilage and surrounding tissues (ie, synovial fluid, labrum, bone) and that has a resolution high enough to resolve thin cartilage will perform the best with an automated segmentation technique. Given the shape and thickness of the hip's cartilage, we hypothesize that there will be variability in manual segmentations of the joint; however, the segmentations performed by an expert rater will closely resemble the automated approach.

## Materials and Methods

### *Sample Characteristics*

This study had institutional review board approval, and informed consent was obtained from all subjects. A total of 20 human hips were used in this investigation from 20 different patients. The patients ranged

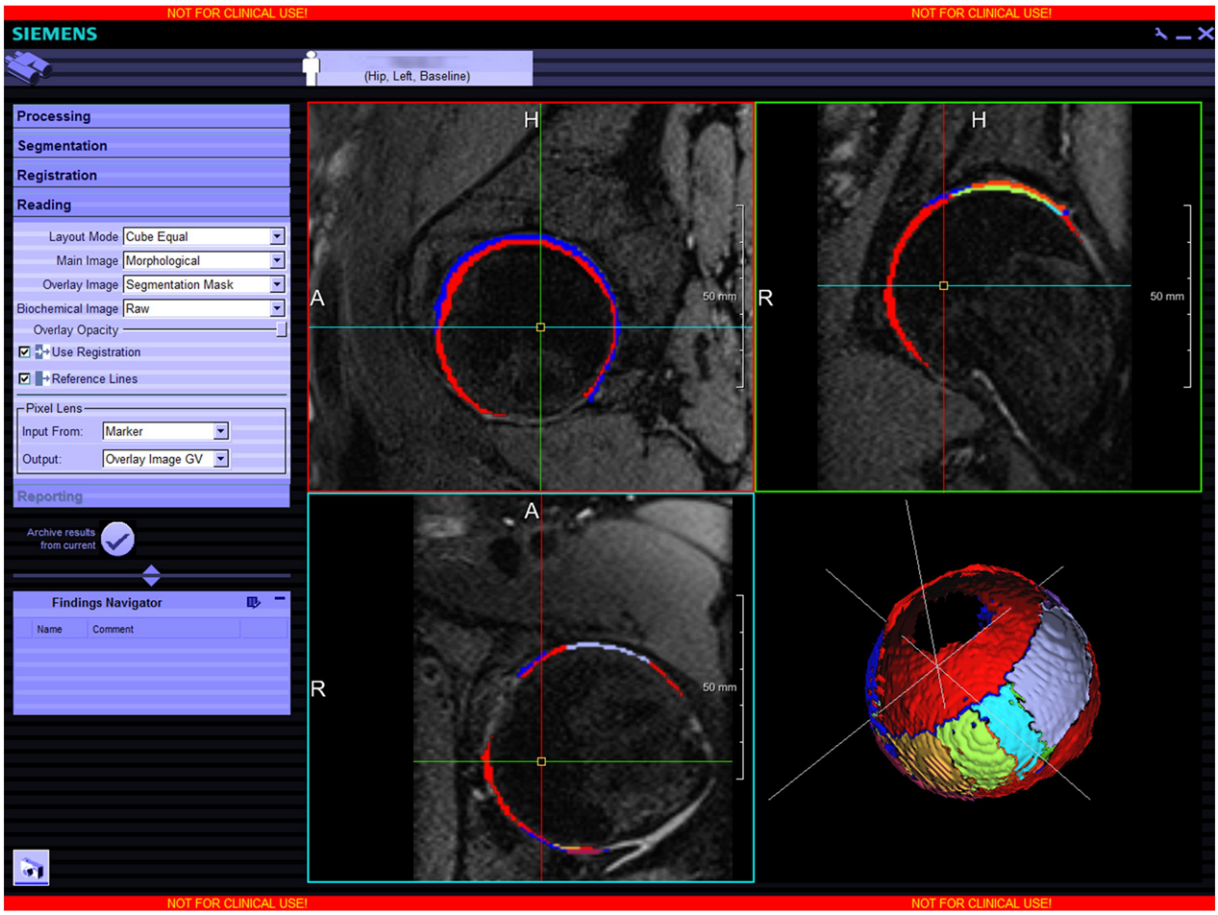


**Fig. 2.** Screenshot of the software used for automated cartilage segmentation.

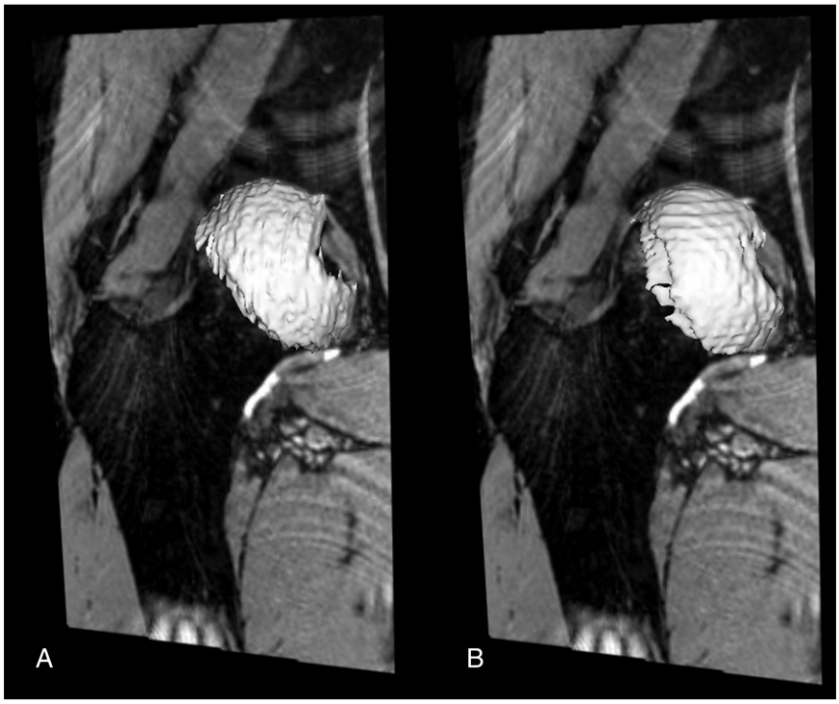


**Fig. 3.** Comparison in segmentation surfaces resulting from the MRI True FISP sequence for the expert rater (A) and automated system (B).

in age from 23 to 74 with 5 males and 15 females. The patients were referred to our imaging center because of hip pain or hip osteoarthritis.

*Magnetic Resonance Imaging*

All MRI was performed on a 128-channel 3T MRI scanner (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) using a 26-element coil setup composed of a flexible 18-element array coil anteriorly and 8 elements from a spine coil posteriorly (Siemens Healthcare, Erlangen, Germany). Ten hips were imaged using a 3-dimensional (3D) true fast imaging with steady-state precession (True FISP) sequence (Repetition time = 10.16 ms, Echo time = 4.28 ms, matrix 256 × 256, flip angle = 30°, pixel size = 0.625 mm × 0.625 mm, slice thickness = 0.629 mm). An additional 10 hips were imaged using a 3D gradient echo (GRE) sequence (TR = 28.0 ms, TE = 4.39 ms, matrix 320 × 320, flip angle = 10°, pixel size = 0.5 mm × 0.5 mm, slice thickness = 3 mm). The 3D GRE sequence is commonly used at our institution for cartilage imaging. Sample images for each sequence are demonstrated in Fig. 1.

*Image Analysis*

The freely available FireVoxel software package (NYU Center for Advanced Imaging Innovation and Research, New York, NY) was used to perform the manual segmentations of the proximal femur and acetabular cartilage (excluding any synovial fluid, labrum, and ligamentum) for all cases and to export DICOM files for use in the remainder of the study. The prototypical cartilage segmentation software (Chondral-Health version 1.3) from Siemens was used to perform automated cartilage segmentation. For this software, a maximum slice thickness of less than 1 mm to avoid partial volume effects on the thin hip cartilage is specified; it should be noted that the slice thickness of our institution's cartilage sequence falls outside this specification [16]. This software (Fig. 2) is based on MeVisLab [17] and encapsulates the automated cartilage segmentation algorithm described by Chandra et al [14,15]. The

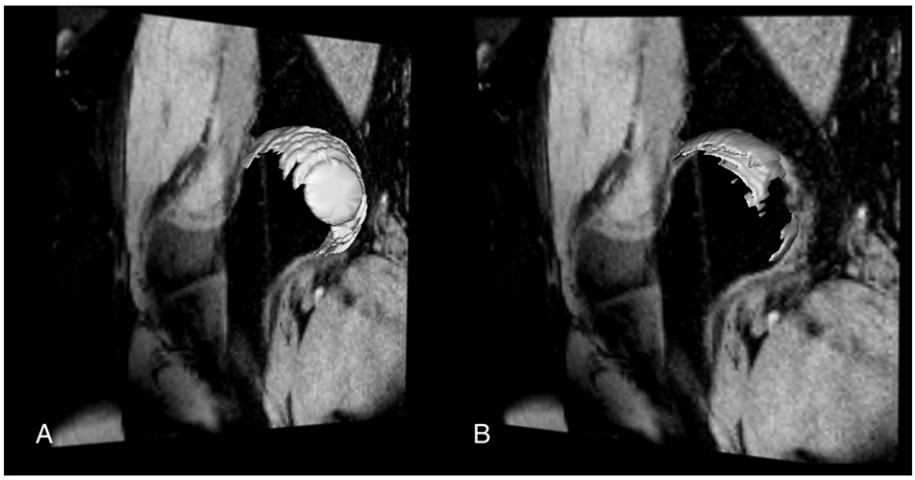


**Fig. 4.** Comparison in segmentation surfaces resulting from the MRI GRE sequence for the expert rater (A) and automated system (B).

output from the software is a label map image with the cartilage separated into femoral and acetabular components with 13 subregions (7 femoral cartilages and 6 acetabular cartilages) (Fig. 2). To give a fair comparison to the manual ratings, the 3D Slicer toolkit was used to binary threshold these regions to give a single label map representing the aggregate hip joint cartilage. The automated segmentation was performed on an AMD Phenom II X6 1090T Processor 3.20 GHZ with 8GB RAM.

*Raters and Qualifications*

Two raters of different experience levels were used in this study including an expert rater and a musculoskeletal physician rater. The third rater for the study was the automated cartilage identification software.

*Statistical Analysis and Evaluation Metrics*

This study was evaluated using 4 evaluation techniques. First, the time to complete each segmentation was recorded, and descriptive statistics were calculated. Second, the segmented region of interest volume was calculated using the FireVoxel software. For hip joint cartilage volumes, we used SPSS (IBM, Somers, NY) to compute the intraclass correlation coefficients (ICCs) (as described by Shrout and Fleiss) to assess reproducibility and reliability. As per Landis and Koch, ICC reliability values can be classified as follows: slight (0-0.2), fair (0.21-0.4), moderate (0.41-0.6), substantial (0.61-0.80), and almost perfect (0.81-1.0) [18].

Third, 6 segmentation overlap metrics were calculated using the C++ programming language and the Insight Toolkit. These metrics include the union overlap (Jaccard coefficient), total overlap (sensitivity), mean overlap (Dice coefficient), volume similarity, false negative, and false positive (1 − specificity). Perfect matching segmentations are represented by a value of 1 for the union overlap, total overlap, and mean overlap, and a value of 0 for false negatives, false positive, and volume similarity. The equations for these metrics are as follow [19]:

$$\text{Total Overlap} = \frac{\sum_r |\text{Segmentation1}_r \cap \text{Segmentation2}_r|}{\sum_r |\text{Segmentation2}_r|} \quad (1)$$

$$\text{Union Overlap(Jaccard Coefficient)} = \frac{\sum_r |\text{Segmentation1}_r \cap \text{Segmentation2}_r|}{\sum_r |\text{Segmentation1}_r \cup \text{Segmentation2}_r|} \quad (2)$$

$$\text{Mean Overlap (Dice Coefficient)} = 2 * \frac{\sum_r |\text{Segmentation1}_r \cap \text{Segmentation2}_r|}{\sum_r |\text{Segmentation1}_r| + |\text{Segmentation2}_r|} \quad (3)$$

$$\text{Volume Similarity} = 2 * \frac{\sum_r |\text{Segmentation1}_r| - |\text{Segmentation2}_r|}{\sum_r |\text{Segmentation1}_r| + |\text{Segmentation2}_r|} \quad (4)$$

$$\text{False Negative} = \frac{\sum_r |\text{Segmentation1}_r| / |\text{Segmentation2}_r|}{\sum_r |\text{Segmentation1}_r|} \quad (5)$$

$$\text{False Positive} = \frac{\sum_r |\text{Segmentation2}_r| / |\text{Segmentation1}_r|}{\sum_r |\text{Segmentation2}_r|} \quad (6)$$

The fourth method of evaluation involved calculating the average distance between the surfaces generated from the rater segmentations. The 3D Slicer toolkit was used to generate and smooth 3D surface representations of each region of interest using the marching cubes algorithm, 10 iterations of sinc smoothing, and 25% surface point decimation. With the cartilage surfaces overlapped, custom C++ code using the Visualization Toolkit was used to calculate a Euclidean distance map for the comparison expert rater vs physician rater and expert rater vs automated rater. In addition, mean distance values for each surface were calculated, and standard descriptive statistics were calculated.

**Table 1**
Comparison of Segmentation Times Between Manual and Automated Methods.

| Imaging Sequence | Average Manual Time | Average Automated Time |
|---|---|---|
| GRE | 8 min 3 s | 12 min 32.7 s |
| True FISP | 41 min 3 s | 12 min 29.4 s |

## Results

Segmentation-based surface representations of the hip joint cartilage were successfully created in all 20 cases for the expert rater, physician rater, and automated rater. Figs. 3 and 4 demonstrate surface representations overlaid on MRI slices from the expert and automated rater for the True FISP and GRE sequences, respectively.

The time trial results for the segmentations are available in Table 1. Using the GRE images, the manual segmentations were completed in 8:03 minutes, which was 1.56 times faster than the automated segmentation method on average. Using the True FISP images, the automated method completed the segmentation in 12:29 minutes, which was 3.28 times faster than the manual segmentation method on average.

Table 2 provides the results of the segmentation volume comparison for each of the 3 raters based on MRI sequence. Table 3 provides the ICC values for the volume comparisons between raters. Using the GRE sequence, the manual ratings had an ICC value of 0.81, but using the True FISP, they had an ICC of 0.312. The expert vs automated rater had an ICC of 0.286 for the GRE sequence and an ICC of 0.614 for the True FISP sequence.

Table 4 provides the overlap metrics comparing the expert and physician rater for both imaging sequences, and Table 5 provides the overlap metric comparing the expert and automated rater for both imaging sequences. In comparing the manual ratings, the interrater overlap metrics for the GRE sequence were 0.763, 0.610, and 0.756 and, for the True FISP, were 0.797, 0.744, and 0.744 for the total overlap, union overlap, and mean overlap, respectively. In comparing the expert and automated raters, the interrater overlap metrics for the GRE sequence were 0.940, 0.196, and 0.324 and, for the True FISP, were 0.874, 0.546, and 0.704 for the total overlap, union overlap, and mean overlap, respectively.

Table 6 provides the average Euclidean distance between surfaces created from the segmentations for the manual segmentations and the expert vs automated segmentations. Sample Euclidean maps for 1 case

**Table 2**
Comparison of Segmentation Volumes for Each of the Cases From the 3 Raters.

| Subject | MRI Sequence | Physician Rater Volume ($cm^3$) | Expert Rater Volume ($cm^3$) | Automated Rater Volume ($cm^3$) |
|---|---|---|---|---|
| 1 | GRE | 15.46 | 14.30 | 3.93 |
| 2 | GRE | 21.33 | 18.88 | 5.35 |
| 3 | GRE | 11.29 | 13.85 | 2.62 |
| 4 | GRE | 15.52 | 15.54 | 2.30 |
| 5 | GRE | 12.28 | 13.06 | 2.62 |
| 6 | GRE | 16.24 | 15.54 | 2.74 |
| 7 | GRE | 10.92 | 12.36 | 2.92 |
| 8 | GRE | 14.77 | 17.16 | 2.76 |
| 9 | GRE | 18.44 | 18.18 | 2.88 |
| 10 | GRE | 16.04 | 14.34 | 3.94 |
| 11 | True FISP | 9.83 | 8.29 | 6.69 |
| 12 | True FISP | 9.00 | 9.73 | 6.28 |
| 13 | True FISP | 8.80 | 13.88 | 8.56 |
| 14 | True FISP | 9.78 | 9.93 | 6.81 |
| 15 | True FISP | 8.68 | 9.31 | 6.68 |
| 16 | True FISP | 8.86 | 9.99 | 7.76 |
| 17 | True FISP | 12.32 | 10.82 | 6.94 |
| 18 | True FISP | 8.96 | 11.50 | 6.79 |
| 19 | True FISP | 8.42 | 11.24 | 7.80 |
| 20 | True FISP | 11.99 | 17.33 | 9.58 |

**Table 3**
Comparison of ICCs (Confidence Intervals) Between the 3 Raters.

| Comparison | GRE Sequence ICC Value | True FISP ICC Value |
|---|---|---|
| Expert vs physician rater | 0.81 (0.405-0.949) | 0.312 (−0.357-0.77) |
| Expert vs automated rater | 0.286 (−0.382-0.758) | 0.614 (0.018-0.888) |
| Physician vs automated rater | 0.351 (−0.319-0.787) | 0.223 (−0.438-0.727) |
| Aggregate | 0.556 (0.173-0.847) | 0.422 (0.029-0.783) |

of the GRE and True FISP sequences between the expert and automated raters are available in Fig. 5.

## Discussion

Automating the process of femoroacetabular cartilage segmentation from MRI images is difficult. In our evaluation of a novel automated cartilage segmentation module, several important findings are evident. Manual segmentation of the hip joint requires training for the raters. In our study, the expert rater was considered the gold standard. The other rater is a physician with a background in the musculoskeletal system, although this was the first time that this physician performed image segmentation. Based on the thin nature of the hip joint cartilage and overall small volume that is segmented, it is expected that there will be more variability in the volume comparison and overlap metrics than if the entire femur had been segmented, for example. The segmentation of the hip joint cartilage requires training, and accurate segmentations take time if manually performed.

The selected MRI sequence and scanning parameters that are chosen are important for accurate cartilage identification by both manual and automated methods. Because of the slice thickness of the GRE sequence, fewer slicers were required to be manually traced resulting in low manual segmentation times; this, however, did not alter the needed time for the automated process. Looking at the volume comparison, the manual ratings for the GRE sequence had "almost perfect" agreement based on the ICC but only "fair" agreement between the manual ratings and automated rating. The GRE sequence highlights the cartilage immensely in the images; however, this makes it difficult to differentiate it from tissues of similar image intensity: synovial fluid and labrum. The large slice thickness leads to partial volume effects, prohibiting exact volume measurements even by expert raters. If the absolute cartilage volume is of interest, the resolution has to be chosen such that the cartilage is resolved to at least some voxels in all dimensions.

The overlap metrics for the GRE sequence were overall lower both between the manual ratings and comparing the expert and automated rater when compared to the other imaging sequence. Furthermore, in comparing the average Euclidean distance, the GRE segmentation average distance was greater for both the manual comparison and the expert vs automated comparison. Overall, the GRE imaging sequence did not perform as well as the True FISP sequence in terms of our ability to reliably identify the articular cartilage of the hip joint.

The True FISP sequence provided improved contrast between the articular cartilage and surrounding tissues. This allowed for "substantial" agreement between the expert rater and the automated segmentation technique with regard to the ICC values for the volume comparison. The overlap metrics for the True FISP sequence comparing the manual ratings and the expert and automated rater were similar. The average cartilage surface difference between the segmentations demonstrates that the manual ratings are more similar. However, the automated technique was 3.28× faster than the manual ratings in this case. Perhaps, the combination of this automated technique with manual editing could produce the optimal conditions for accuracy and speed.

The number of automated hip cartilage segmentation methods for comparisons is limited. Xia et al [13] report mean overlap (Dice coefficient) for their approach using graph searching to automate hip cartilage segmentation with a value of 0.81 using the True FISP MRI sequence and 54 volunteers. Our study's mean overlap value of 0.704 is comparable given the size of the region of interest that is being considered and the greater sample size of the study by Xia et al. Chandra et al report a mean overlap of 0.73 when using a multispin echo T2 MRI sequence and statistical shape modeling with 24 subjects. This study chose a T2 imaging sequence, as it was hoped to be more representative of imaging sequences seen in the clinic. Our study's trial using the GRE sequence was performed for a similar rationale; however, our mean overlap of 0.324 falls short. The GRE sequence used clinically at our institution has a slice thickness of 3 mm, whereas 1 mm was used by Chandra et al.

Given the thin nature of the femoroacetabular cartilage and the spherical shape of the femoral head, the stair-step artifact using this slice thickness is seen in Figs. 4 and 5. It is important that orthopedists note that the "standard" cartilage imaging sequence at a given institution may not be appropriate for volumetric cartilage quantification. Scanning parameters must be matched to the geometry of the region of interest. Based on the literature and our own results, we recommend that MR sequences should have at most a slice thickness of 1 mm as reported by Chandra et al for volumetric analysis of the femoroacetabular cartilage. We recommend clear communication with MR physicists before implementation of an automated hip cartilage quantification algorithm to ensure adequate scanning parameters are performed in the clinical setting.

**Table 4**
Comparison of Overlap Metrics for All Subjects Between the Expert Rater and the Physician Rater.

| Metric | MRI Sequence | Minimum Value | Maximum Value | Average Value | SD |
|---|---|---|---|---|---|
| Total overlap | GRE | 0.709 | 0.811 | 0.763 | 0.037 |
| Total overlap | True FISP | 0.659 | 0.901 | 0.797 | 0.078 |
| Union overlap | GRE | 0.501 | 0.702 | 0.610 | 0.063 |
| Union overlap | True FISP | 0.515 | 0.639 | 0.744 | 0.039 |
| Mean overlap | GRE | 0.668 | 0.825 | 0.756 | 0.049 |
| Mean overlap | True FISP | 0.679 | 0.779 | 0.744 | 0.039 |
| Volume similarity | GRE | −0.122 | 0.204 | 0.017 | 0.113 |
| Volume similarity | True FISP | −0.170 | 0.448 | 0.117 | 0.214 |
| False negative | GRE | 0.189 | 0.291 | 0.237 | 0.037 |
| False negative | True FISP | 0.099 | 0.341 | 0.203 | 0.078 |
| False positive | GRE | 0.126 | 0.379 | 0.247 | 0.082 |
| False positive | True FISP | 0.151 | 0.432 | 0.290 | 0.094 |

**Table 5**
Comparison of Overlap Metrics for All Subjects Between the Expert Rater and the Automated Rater.

| Metric | MRI Sequence | Minimum Value | Maximum Value | Average Value | SD |
|---|---|---|---|---|---|
| Total overlap | GRE | 0.885 | 0.965 | 0.940 | 0.022 |
| Total overlap | True FISP | 0.778 | 0.944 | 0.874 | 0.052 |
| Union overlap | GRE | 0.129 | 0.270 | 0.196 | 0.050 |
| Union overlap | True FISP | 0.429 | 0.650 | 0.546 | 0.076 |
| Mean overlap | GRE | 0.228 | 0.425 | 0.324 | 0.070 |
| Mean overlap | True FISP | 0.601 | 0.788 | 0.704 | 0.065 |
| Volume similarity | GRE | 1.117 | 1.484 | 1.311 | 0.143 |
| Volume similarity | True FISP | 0.223 | 0.575 | 0.389 | 0.112 |
| False negative | GRE | 0.034 | 0.115 | 0.060 | 0.022 |
| False negative | True FISP | 0.0563 | 0.222 | 0.126 | 0.052 |
| False positive | GRE | 0.727 | 0.869 | 0.802 | 0.051 |
| False positive | True FISP | 0.290 | 0.522 | 0.408 | 0.078 |

**Table 6**
Comparison of Average Euclidean Distances in Millimeters (SD) for All Subjects Between the 2 Manual Ratings and Between the Expert Rater and Automated Rater.

| MRI Sequence | Average Euclidean Distance (mm) Expert Rater vs Physician Rater | Average Euclidean Distance (mm) Expert Rater vs Automated Rater |
|---|---|---|
| GRE | 0.682 (0.159) | 6.06 (0.680) |
| True FISP | 0.554 (0.255) | 1.540 (0.481) |

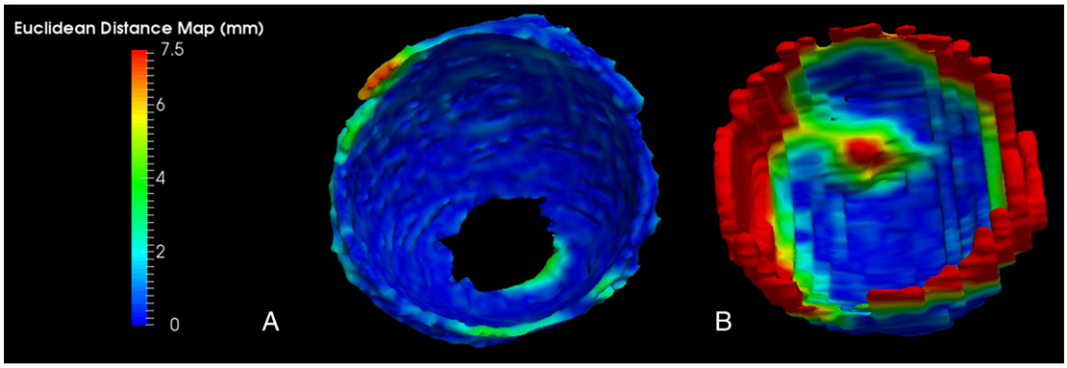


**Fig. 5.** Euclidean distance mapping of hip cartilage surfaces comparing the expert rater and automated rater for each hip MRI image sequence: True FISP (A) and GRE (B).

In the future, we aim to collect a large database of 3D hip MRI data sets with a focus on correlating the quantitative cartilage metrics with clinical outcomes. Furthermore, we propose to perform MRI on a subset of patients over time to evaluate patients at the first presentation for hip pain to the point at which hip preservation surgery or arthroplasty is performed. This will allow for temporal comparison of cartilage thickness, which will be useful for establishing quantitative thresholds to recommend surgical intervention. Now that we have demonstrated high measurement reproducibility, we envision this tool as being able to allow surgeons to closely monitor disease progression or treatment response in patients. For example, if surgeons knew that patients were losing 20% of their hip cartilage volume annually, then this could influence their recommendations for physical activity or weight loss, especially because radiographs are insensitive for monitoring cartilage loss.

In addition, this tool could allow a surgeon to assess the effect of interventions to preserve cartilage health. If a patient's cartilage volume stabilized after a surgical procedure, such as a periacetabular osteotomy in a patient with hip dysplasia, then the surgeon would feel more confident that the procedure was successful. This tool may even permit the more accurate comparison of the effects of different therapies in clinical trials, as cartilage health is currently not quantified in standard radiology reports. Overall, we feel that this tool has high potential to impact patient management, especially because the MRI protocols are already available on MRI scanners from the major MRI scanner vendors and, therefore, do not require any additional cost to patients.

## Conclusions

Overall, we have applied a prototypical segmentation algorithm to a clinical data set for the first time and have performed rigorous evaluation of this new automated cartilage segmentation technique. We have shown that the True FISP MR imaging sequence ($0.625 \times 0.625 \times 0.629$ mm) allowed for more accurate segmentations than the more commonly used clinical GRE MR imaging sequence ($0.5 \times 0.5 \times 3.0$ mm) at our institution. This method when used with the True FISP sequence is $3.28\times$ faster than manual segmentation and produces results comparable to that found in the literature given the region of interest's small volume. The overlap metrics comparing the automated method to that of the manual ratings were comparable. However, this automated segmentation method represents an automated method for femoroacetabular cartilage quantification, which could allow for temporal analysis of cartilage thickness in symptomatic patients to support surgical intervention. We recommend clear communication with MR physicists before implementation of an automated algorithm to ensure that adequate scanning parameters including resolution and contrast are clinically obtained.

## Acknowledgments

The authors would like to thank Mary Bruno for her assistance in collecting the MRI data sets.

## References

1. Bedi A, Kelly BT, Khanduja V. Arthroscopic hip preservation surgery: current concepts and perspective. Bone Joint J 2013;95-b:10.
2. Chu CR, Millis MB, Olson SA. Osteoarthritis: from palliation to prevention: AOA critical issues. J Bone Joint Surg Am 2014;96:e130.
3. Peters CL. Mild to moderate hip OA: joint preservation or total hip arthroplasty? J Arthroplasty 2015;30:1109.
4. Eckstein F, Glaser C. Measuring cartilage morphology with quantitative magnetic resonance imaging. Semin Musculoskelet Radiol 2004;8:329.
5. Fripp J, Crozier S, Warfield SK, et al. Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. IEEE Trans Med Imaging 2010;29:55.
6. Lee JG, Gumus S, Moon CH, et al. Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. Med Phys 2014;41:092303.
7. Pang J, Li P, Qiu M, et al. Automatic articular cartilage segmentation based on pattern recognition from knee MRI images. J Digit Imaging 2015 [Epub ahead of print].
8. McWalter EJ, Wirth W, Siebert M, et al. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. Osteoarthritis Cartilage 2005;13:48.
9. Nishii T, Sugano N, Sato Y, et al. Three-dimensional distribution of acetabular cartilage thickness in patients with hip dysplasia: a fully automated computational analysis of MR imaging. Osteoarthritis Cartilage 2004;12:650.
10. Li W, Abram F, Beaudoin G, et al. Human hip joint cartilage: MRI quantitative thickness and volume measurements discriminating acetabulum and femoral head. IEEE Trans Biomed Eng 2008;55:2731.
11. Abraham CL, Bangerter NK, McGavin LS, et al. Accuracy of 3D dual echo steady state (DESS) MR arthrography to quantify acetabular cartilage thickness. J Magn Reson Imaging 2015. http://dx.doi.org/10.1002/jmri.24902 [Epub ahead of print].
12. Siversson C, Akhondi-Asl A, Bixby S, et al. Three-dimensional hip cartilage quality assessment of morphology and dGEMRIC by planar maps and automated segmentation. Osteoarthritis Cartilage 2014;22:1511.
13. Xia Y, Chandra SS, Engstrom C, et al. Automatic hip cartilage segmentation from 3D MR images using arc-weighted graph searching. Phys Med Biol 2014;59:7245.
14. Chandra SS, Surowiec R, Ho C, et al. Automated analysis of hip joint cartilage combining MR T2 and three-dimensional fast-spin-echo images. Magn Reson Med 2015. http://dx.doi.org/10.1002/mrm.25598 [Epub ahead of print].
15. Chandra SS, Xia Y, Engstrom C, et al. Focused shape models for hip joint segmentation in 3D magnetic resonance images. Med Image Anal 2014;18:567.
16. Sato Y, Tanaka H, Nishii T, et al. Limits on the accuracy of 3-D thickness measurement in magnetic resonance images—effects of voxel anisotropy. IEEE Trans Med Imaging 2003;22:1076.
17. Heckel FS, M, Peitgen HO. Object-oriented application development with MeVisLab and Python. GI Jahrestagung 2009:1338.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159.
19. Tustison NJ, Gee JC. Introducing Dice, Jaccard, and other label overlap measures to ITK. Insight J 1-4 2009:17.