Seminars in Arthritis and Rheumatism 51 (2021) 623–626

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Volumetric quantitative measurement of hip effusions by manual versus automated artificial intelligence techniques: An OMERACT preliminary validation study

Jacob L. Jaremko[a,*], Banafshe Felfeliyan[b], Abhilash Hareendranathan[a], Bashiar Thejeel[a], Quinn-Laurin Vanessa[a], Mikkel Østergaard[c,d], Philip G. Conaghan[e], Robert G.W. Lambert[a], Janet L. Ronsky[f], Walter P. Maksymowych[g]

[a] Department of Radiology & Diagnostic Imaging, University of Alberta, 2A2.41 WMC, 8440 – 112 St. NW, Edmonton, AB, Canada
[b] Biomedical Engineering, Schulich School of Engineering, University of Calgary, Canada
[c] Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen, Denmark
[d] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[e] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, UK
[f] Mechanical & Manufacturing Engineering, Schulich School of Engineering, University of Calgary
[g] Department of Medicine, University of Alberta, Canada

ARTICLE INFO

Keywords:
Hip joint
Osteoarthritis
MRI
Effusion
OMERACT
Artificial intelligence

ABSTRACT

Objective: Preliminary assessment, via OMERACT filter, of manual and automated MRI hip effusion Volumetric Quantitative Measurement (VQM).
Methods: For 358 hips (93 osteoarthritis subjects, bilateral, 2 time points), 2 radiologists performed manual VQM using custom Matlab software. A Mask R-CNN artificial-intelligence (AI) tool was trained to automatically compute joint fluid volumes.
Results: Manual VQM had excellent inter-observer reliability (ICC 0.96). AI predicted hip fluid volumes with ICC 0.86 (status), 0.58 (change) vs. 2 human readers.
Conclusion: Hip joint fluid volumes are reliably assessed by VQM. It is feasible to automate this approach using AI, with promising initial reliability.

Introduction

Joint effusion is a feature of osteoarthritis visible on MRI that represents a tempting objective target for therapy. Increased fluid secretion by inflamed synovium is associated with active arthritis [1]. Effusion is traditionally assessed semi-quantitatively on MRI as “mild, moderate, or severe”, as seen in the Hip OsteoArthritis MRI Scoring system (HOAMS; grades 0–3) [2]. Even when graded in this semi-quantitative, reader-dependent fashion, hip effusion shows moderate associations to pain [3], but correlations to stiffness, disability and clinical outcomes are limited [4]. These correlations may be improved by measuring joint fluid more precisely on MRI, counting all the voxels containing fluid signal intensity along the joint. This process, termed “Volumetric Quantitative Measurement” (VQM), has become easier recently with improving computer tools, and can now be automated using artificial intelligence (AI). We sought to begin the process of applying the OMERACT filter [5] to evaluate manual human and automated AI hip joint effusion VQM. This pilot exercise was performed within the OMERACT MRI in Arthritis Working Group, presented at OMERACT 15 (virtual meeting, 29 October 2020).

Material & methods

Data available

The existing University of Alberta Steroid Injection in Hip Osteoarthritis (STIHO) cohort includes 97 adults with symptomatic hip OA who presented to a radiology clinic for fluoroscopically guided steroid injection. With ethical approval (UofA HREB Pro00039139) and written informed consent, each subject underwent MRI of both hips pre-injection and 8 weeks post-injection. Of 97 enrolled patients, 4 had no images, 6 had only 1 time point available and 1 had only one hip that could be analyzed, leaving 358 hip image sets in 93 patients, who were 55% male, age 59±13 years (mean±standard deviation, SD). We used wide field-of-view coronal STIR images of both hips (repetition/echo/inversion times TR/TE/TI 4530/50/150 ms, matrix size 384 × 250, slice thickness 4 mm, field-of-view 350×350 mm).

* Corresponding author.
E-mail address: jjaremko@ualberta.ca (J.L. Jaremko).

https://doi.org/10.1016/j.semarthrit.2021.03.009
0049-0172/

Scans were assessed in random order blinded to chronology and clinical data.

VQM tool

We previously developed a computer tool using custom Matlab software (The Mathworks, Natick, MA, v. R2019). A user quickly (~5 s) outlines a loose region of interest (ROI) around the hip joint in each MRI slice, within which an automated tool identifies and counts all voxels of signal intensity greater than an automatically determined threshold representing fluid within the ROI. The sum of voxels across all MRI slices indicates the joint fluid volume. We have used this tool to quantify normal fluid volumes in 140 pediatric hips [6] and to measure joint effusions in hip osteoarthritis patients [7]. VQM reading times averaged 3.9 min per hip [7]. For this exercise, we used the previously published VQM readings in the latter (STIHO) cohort from two musculoskeletal radiologists with 6 and 4 years experience (VQL, BT) as human expert gold-standard fluid volumes to compare against AI.

AI tool and training

For this exercise we developed a convolutional neural network (CNN) that operates analogously to the VQM tool above, except that once trained the CNN functions automatically, without further human input. We followed Dreizin et al. [8], who proposed two-stage

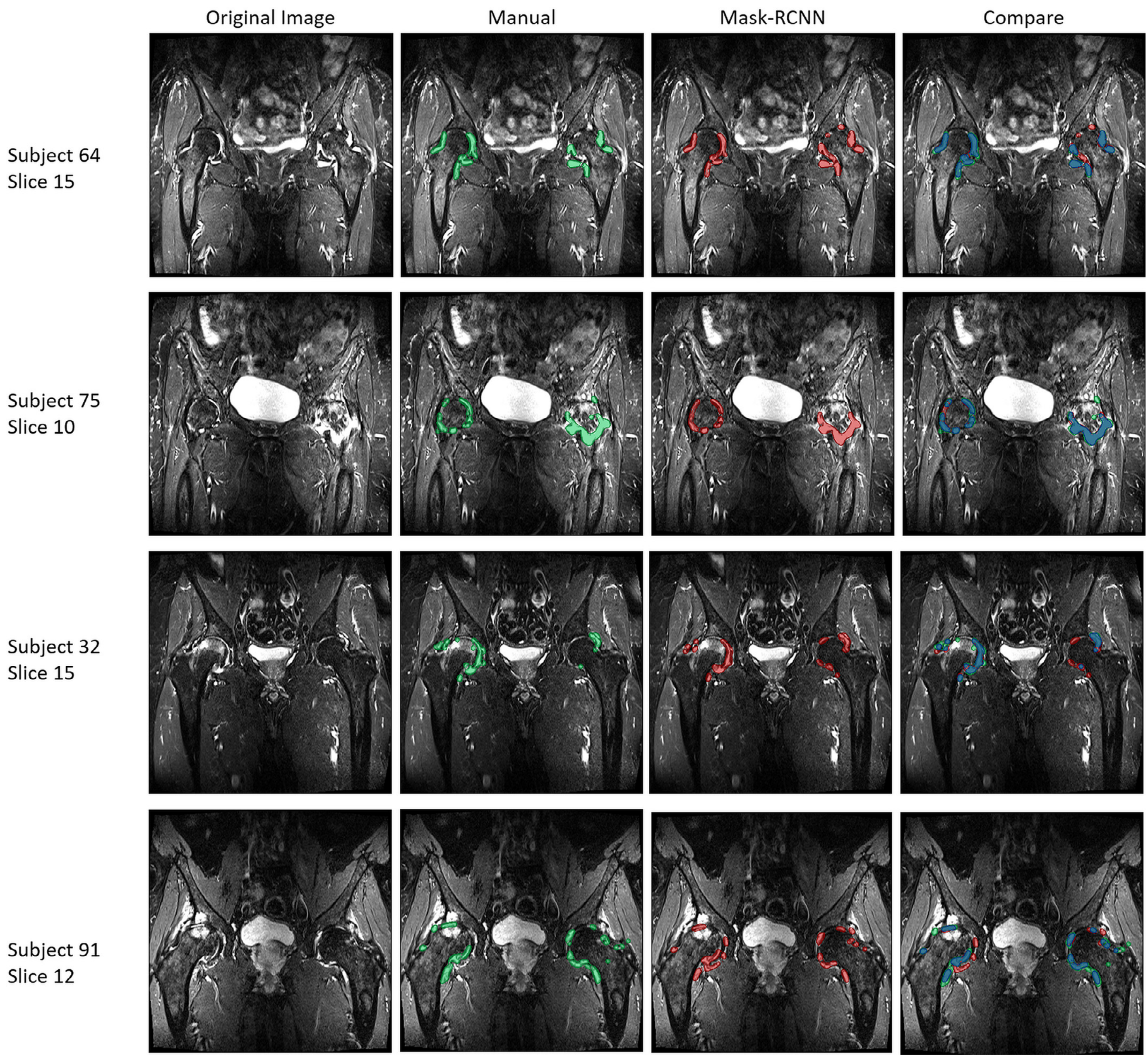

Fig. 1. Mask overlays demonstrating manual and AI identification of hip joint fluid in selected patients. First column = raw images. Second column = masks from human reader 1, with green voxels representing hip joint fluid. Third column = masks from AI, with red voxels representing AI predicted locations of joint fluid. Fourth column = comparison between human and AI masks. Blue = intersection voxels (identified by both methods); red = only AI; green = only human reader. Note that there is generally high visual fidelity between manual and AI results, with the AI identifying slightly more fluid overall than our human readers. For subject 32, AI mis-identified bright cartilage at the medial and inferior aspects of the left hip joint as fluid, while for subject 91, AI mis-identified blood vessels at the inferior medial aspect of right hip as joint fluid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

segmentation for identifying pelvic fluid collections, in which the first network generated a coarse ROI bounding box around the fluid and a second network was then responsible for fine segmentation, building on the first result.

We deployed Mask R-CNN architecture, consisting of a backbone/region proposal network, which extracts multi-scale feature maps to recognize effusion location (generating a bounding box for each effusion), and then uses the generated feature maps cropped to this bounding box to extract pixel-wise image masks. The Mask R-CNN code (downloadable at [9]) is written for Tensorflow v2.0/Python v3.6. The network is trained on one GPU (NVIDIA V100) for 300 epochs, with learning rate 0.001, using a loss function combining classification loss, bounding-box loss, and mask-loss. We used ~75% of STIHO scans for training (135 hip image sets) and validation (n = 123), withholding a random 25 patients (2 hips each at 2 time points) for testing. Network output was a mask overlay identifying hip joint fluid in each MRI slice, and the estimated total hip fluid volume.

Statistics

We considered 2 human experts and AI as 3 readers, each estimating the actual hip joint fluid volume. We calculated mean±standard deviation (SD) of fluid volumes and of differences between measurements for each reader pair, and Coefficients of Variation (CoV) between each pair. We computed intraclass correlation coefficients (ICC) (two-way random, single measures, ICC(2,1)) between reader pairs and between AI and the average of the 2 human readers, for fluid volume at each time point and for change between time points.

Results

Slice-by-slice visual image review showed AI and human readers to have generally high agreement on joint fluid locations, with AI tending to overestimate the presence of joint fluid vs. the human readers (Fig. 1).

Quantitative analysis confirms the visual impression that AI tended to overestimate fluid volumes (Table 1), but despite this systematic error, AI demonstrated high correlation to human readings (Table 2).

Discussion

In this study we performed initial pilot evaluation of a tool to automatically calculate MRI hip joint fluid volume by artificial intelligence.

Feasibility

Automated VQM of joint effusion has been developed previously for the knee as early as 2010, but that approach was of limited utility in clinical imaging because it required adding a pair of dedicated axial MRI sequences (7 min scan time) and cumbersome post-processing (45 min per knee with specialized software)[10]. In contrast, our deep learning approach uses an MRI sequence routinely acquired in clinical hip MRI (coronal STIR). It requires a powerful GPU-enabled computer to initially train AI, but once network weights have been established, it generates results nearly instantaneously (~1 second per slice) on a typical Windows personal computer running free publicly available software (e.g., Python, JavaScript). With validation and wide distribution of this package, joint fluid volume could become routinely reported as part of any joint MRI.

Reliability

The VQM method was previously demonstrated to be highly reliable for expert human readers [6,7]. The AI tool is fully automated, removing inter-reader variability. It showed high concordance with human readers in terms of baseline fluid volume and detection of change over time. However, the AI tool had slightly lower agreement with human manual assessment than the agreement between two human experts, and AI systematically overestimated fluid volumes, mainly by identifying other T2-intense structures such as blood vessels near the hip as joint fluid. Since AI was trained on just 135 hip scans, further AI training on additional data, careful attention in labeling and intensity thresholding to separating fluid vs. cartilage, and addition of a heuristic rule that a bright region is only counted as joint fluid if it is in direct contact with bone, are strategies we are testing to improve AI results.

Limitations

The data set was small for AI, which works best with thousands of cases. The MRI sequence used is relatively low-resolution. No external pathologic tissue gold standard for effusion volume was available in this study, although an older VQM technique was shown to correlate well to volumes at joint fluid aspiration (r = 0.88) [10].

Conclusion

An artificial intelligence tool to automatically quantify hip joint fluid volume is highly feasible for clinical application, and shows promising initial reliability. Further refinements to AI network design and training, and more extensive validation vs. imaging and clinical data are needed.

Declaration of Competing Interest

The co-authors have no relevant conflicts of interest to declare.

Acknowledgments

JJ is supported by the AHS Chair in Diagnostic Imaging. Academic time for JJ, RL and VQL is made available by Medical Imaging Consultants (MIC), Edmonton, Canada. BF is supported by an Alberta Innovates Graduate Student Scholarship for Data-Enabled Innovation. PGC is supported in part by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre, United Kingdom. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. We thank the members of the OMERACT MRI in Arthritis Working Group for their participation and support in this project.

Table 1
Hip fluid volumes quantified by 2 human readers & automatically by AI. Data are in mL (mean±standard deviation SD) for the test set (n=100 hips; not used during AI training).

	Fluid volumes (Baseline)			Difference in volumes (Week 8 vs. Baseline)			Coefficient of Variation (CoV)		
	Right	**Left**	**Overall**	**Reader 1**	**Reader 2**	**AI**	**Reader 1**	**Reader 2**	**AI**
Reader 1	7.72 ± 5.33	6.17 ± 6.28	6.94 ± 5.85	0	1.13 ± 0.90	3.06 ± 1.77	0	0.21	0.33
Reader 2	8.65 ± 5.05	6.63 ± 6.02	7.64 ± 5.62	1.13 ± 0.90	0	2.26 ± 1.31	0.21	0	0.22
AI	10.05 ± 3.76	8.49 ± 3.56	9.27 ± 3.73	3.06 ± 1.77	2.26 ± 1.31	0	0.33	0.22	0

Table 2
Inter-reader agreement of hip effusion volumetric quantitative measurements between two human readers and between human vs. automated AI assessment. Values in the table are ICC(2,1), mean[95% confidence interval], for the test set (n=100 hips; not used during AI training).

Baseline Effusion Volume					
Interreader agreement: ICC between 2 Human Readers			AI versus Mean of 2 Human Readers		
Left Hip	Right Hip	per Patient	Left Hip	Right Hip	per Patient
0.98 [0.97, 1.0]	0.99 [0.99, 1.0]	0.99 [0.98, 1.0]	0.82 [0.59, 0.92]	0.93 [0.84,0.97]	0.86 [0.68,0.94]
Difference in Effusion Volumes (Baseline versus Week 8)					
Interreader agreement: ICC between 2 Human Readers			**AI versus Mean of 2 Human Readers**		
Left Hip	**Right Hip**	**per Patient**	**Left Hip**	**Right Hip**	**per Patient**
0.87 [0.69,0.94]	0.62 [0.15,0.85]	0.80 [0.53, 0.92]	0.65 [0.19,0.85]	0.66 [0.21,0.85]	0.58 [0.03,0.82]

References

[1] Atukorala I, Kwoh CK, Guermazi A, Roemer FW, Boudreau RM, Hannon MJ, et al. Synovitis in knee osteoarthritis: a precursor of disease? Ann Rheum Dis 2016;75:390–5.

[2] Roemer FW, Hunter DJ, Winterstein A, Li L, Kim YJ, Cibere J, et al. Hip Osteoarthritis MRI Scoring System (HOAMS): reliability and associations with radiographic and clinical findings. Osteoarthr Cartil 2011;19:946–62.

[3] Ahedi H, Aitken D, Blizzard L, Cicuttini F, Jones G. Quantification of hip effusion-synovitis and its cross-sectional and longitudinal associations with hip pain, MRI findings and early radiographic hip OA. BMC Musculoskelet Disord 2020;21:533.

[4] Steer KJD, Bostick GP, Woodhouse LJ, Nguyen TT, Schankath A, Lambert RGW, et al. Can effusion-synovitis measured on ultrasound or MRI predict response to intra-articular steroid injection in hip osteoarthritis? Skelet Radiol 2019;48:227–37.

[5] Boers M, Beaton DE, Shea BJ, Maxwell LJ, Bartlett SJ, et al. OMERACT Filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. J Rheumatol 2019;46:1021–7.

[6] Quinn-Laurin V, Thejeel B, Chauvin NA, Brandon TG, Weiss PF, Jaremko JL. Normal hip joint fluid volumes in healthy children of different ages, based on MRI volumetric quantitative measurement. Pediatr Radiol 2020;50:1587–93.

[7] Quinn-Laurin V, Bostick GP, Thejeel B, Mandegaran R, Steer KJD, Lambert RGW, et al. Development of a technique for MRI gold-standard direct volumetric measurement of complex joint effusion, and validation at the hip. Skelet Radiol 2020. doi: 10.1007/s00256-020-03630-6.

[8] Dreizin D, Zhou Y, Zhang Y, Tirada N, Yuille AL. Performance of a deep learning algorithm for automated segmentation and quantification of traumatic pelvic hematomas on CT. J Digit Imaging 2020;33:243–51.

[9] Abdulla W. Mask R-CNN for object detection and instance segmentation on keras and tensorflow Computer software, 2017; accessed at: https://github.com/matterport/Mask_RCNN, 22 Mar 2021.

[10] Li W, Abram F, Pelletier J-P, Raynauld J-P, Dorais M, d'Anjou M-A, et al. Fully automated system for the quantification of human osteoarthritic knee joint effusion volume using magnetic resonance imaging. Arthritis Res Ther 2010;12:R173.