

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/68704>

Author: Aizenberg, E.

Title: Computer-aided techniques for assessment of MRI-detected inflammation for early identification of inflammatory arthritis

Issue Date: 2019-03-14

Computer-aided techniques for assessment of MRI-detected inflammation for early identification of inflammatory arthritis

Evgeni Aizenberg

ISBN: 978-94-6182-931-3

Printing and cover layout by Off Page, Amsterdam

© 2019, Evgeni Aizenberg, Leiden, The Netherlands. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

Computer-aided techniques for assessment of MRI-detected inflammation for early identification of inflammatory arthritis

Proefschrift

Ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 14 maart 2019
klokke 13:45 uur

door

Evgeni Aizenberg
geboren te Moskou in 1990

Promotores:

Prof. dr. ir. B.P.F. Lelieveldt

Prof. dr. A.H.M. van der Helm–van Mil

Co-promotor:

Dr. B.C. Stoel

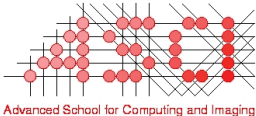
Leden promotiecommissie:

Dr. M. Reijnierse

Prof. dr. J.L. Bloem

Prof. dr. W. Lems (VUmc)

Dr. ir. K.L. Vincken (UMC Utrecht)



This work was carried out in the ASCI graduate school.
ASCI dissertation series number 404.

The research described in this thesis was conducted at the Division of Image Processing (LKEB), Department of Radiology of the Leiden University Medical Center, Leiden, The Netherlands and was supported by the Dutch Technology Foundation STW, under grant number 13329. STW (currently TTW) is part of the Netherlands Organization for Scientific Research (NWO), which is partly funded by the Dutch Ministry of Economic Affairs.

Financial support for the publication of this thesis was kindly provided by:

Bontius Stichting

ASCI research school

Library of Leiden University

Contents

1 Introduction	1
2 Computer-aided evaluation of inflammatory changes over time on MRI of the spine in patients with suspected axial spondyloarthritis: a feasibility study	11
3 Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis	31
4 Automatic quantification of tenosynovitis on MRI of the wrist in patients with early arthritis	53
5 Identifying MRI-detected inflammatory features specific for rheumatoid arthritis: two-fold feature reduction maintains predictive accuracy in clinically suspect arthralgia patients	71
6 Summary and general discussion	97
7 Samenvatting en algemene discussie	103
List of publications	111
Acknowledgements	113
Curriculum Vitae	115

1

Introduction

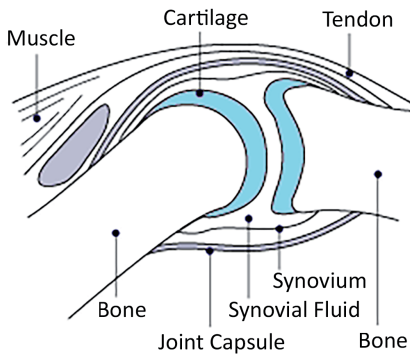
Inflammatory arthritis

Inflammatory arthritis comprises a group of diseases in which the immune system attacks the body's own tissues. The precise cause of these diseases is not yet fully understood. However, combinations of genetic and environmental risk factors have been identified [1,2]. Inflammation can occur as a result of the body producing antigens that trigger an autoimmune response, or as a result of increased production of pro-inflammatory cytokines mistakenly signaling the innate immune system to attack healthy tissues. Two prevalent types of inflammatory arthritis are rheumatoid arthritis and spondyloarthritis [3,4].

Rheumatoid arthritis

Rheumatoid arthritis (RA) primarily manifests itself as inflammation of the synovial joints (Figure 1), especially in the hands, wrists, and feet. Synovial joints are the most common type of joint in the human body, allowing for movement and comprising of two bones covered with articular cartilage and separated by a lubricating fluid called the synovial fluid. The bone surfaces and the fluid are encapsulated by the synovial membrane (also known as synovium), which provides nutrients for the cartilage and produces the synovial fluid. Early inflammation often affects the synovial membrane and the bone marrow, ultimately leading to cartilage loss, bone erosions, and joint deformity if left untreated.

Normal Joint



Joint Affected by Rheumatoid Arthritis

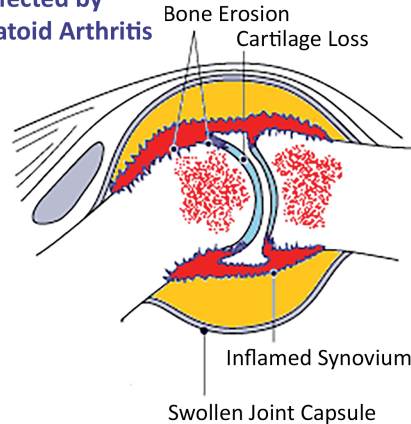
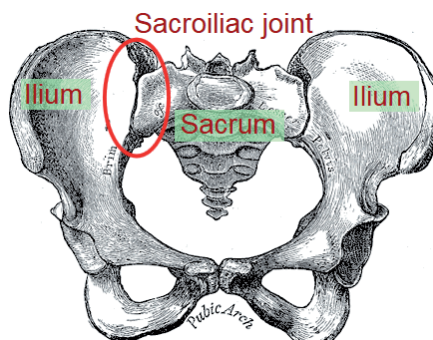


Figure 1. Depiction of changes observed in a synovial joint affected by rheumatoid arthritis. Early inflammation often affects the synovial membrane (synovium) and the bone marrow, ultimately leading to cartilage loss, bone erosions, and joint deformity. (Adapted from Wikimedia [5])

Spondyloarthritis

Spondyloarthritis (SpA) represents an inter-related group of conditions of which ankylosing spondylitis is considered the prototype disease [6], characterized by inflammation in the sacroiliac (SI) joints (Figure 2a) and the vertebrae (Figure 2b). These anatomically axial manifestations give rise to the term axial SpA. Early signs of inflammation often occur in the bone marrow. Long-term inflammation can lead to bone erosion followed by formation of bony bridges that result in fusion of bones in the SI joints and adjacent vertebrae in the spine, severely impairing mobility.



(a)

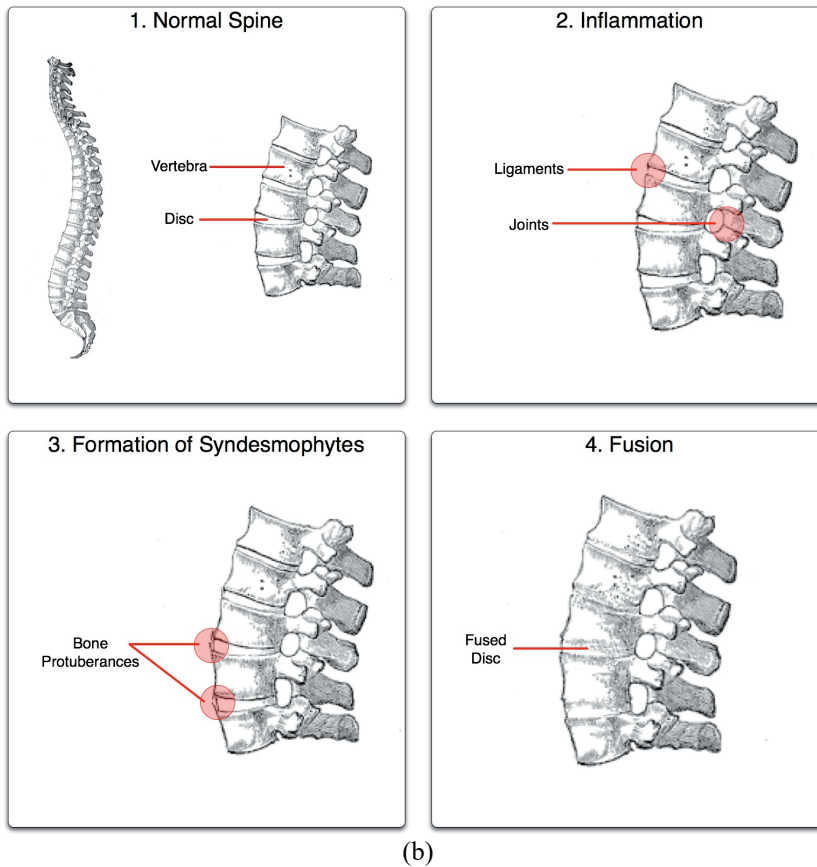


Figure 2. Pathology of axial spondyloarthritis. Inflammation in the sacroiliac joints (a) can lead to fusion of the sacrum and the ilium bones of the pelvis. Inflammation in the vertebrae of the spine (b) can lead to formation of bony bridges called syndesmophytes, resulting in fusion of adjacent vertebrae (b). (Source: Wikimedia [7,8])

Diagnosis and treatment

Clinical diagnosis of RA and SpA typically involves a combination of tests, such as physical examination by a rheumatologist, assessment of symptom history, X-ray imaging, and blood tests. Traditionally, the first line of treatment has consisted of physiotherapy, painkillers, and non-steroidal anti-inflammatory drugs (NSAIDs). Exercise strengthens muscles around joints and helps maintain mobility, while painkillers and NSAIDs reduce pain for a limited time period. More recently, advances in disease-modifying anti-rheumatic drugs (DMARDs) and biological DMARDs have allowed for long-term reduction of inflammation and joint damage

and even a possibility of drug-free sustained remission [9,10]. However, research findings point to the importance of early diagnosis, as treatment in the early stages of the disease increases chances of better outcome and improved quality of life for patients [9,11]. Therefore, much effort is presently being devoted to early diagnosis of RA and SpA. To this end, the diagnostic potential of imaging modalities sensitive to local inflammation is of great interest.

Magnetic resonance imaging

Imaging plays an important role in diagnosis and monitoring of inflammatory arthritis. However, most clinical practices rely on X-ray imaging [12,13], which is limited to depicting structural changes that occur at later disease stages. Over the past two decades, extensive research has been conducted on the use of magnetic resonance imaging (MRI) as means of detecting inflammation in early disease stages before clinical arthritis becomes evident. MRI is sensitive to local inflammation [14], allowing for detailed joint-level assessment of inflammatory changes such as bone marrow edema (feature of inflammation of the bone marrow, also known as osteitis), synovitis (inflammation of the synovial membrane), and tenosynovitis (inflammation of the synovial lining of the sheath surrounding tendons).

In axial SpA patients, the main inflammatory feature of interest is bone marrow edema (BME), since it plays an important role in early diagnosis [15]. It can be visualized using a T2-weighted sequence with fat-saturation or a short tau inversion recovery (STIR) sequence. These acquisition sequences suppress fat signal, forcing healthy bone marrow to appear dark, while bringing out BME as regions of high intensity (Figure 3a).

In RA patients, BME is also an important inflammatory feature, since it is a strong predictor of erosive progression [16]. In addition to that, synovitis and tenosynovitis are frequently observed in patients with early disease [17]. Furthermore, tenosynovitis has been found to be predictive of progression from

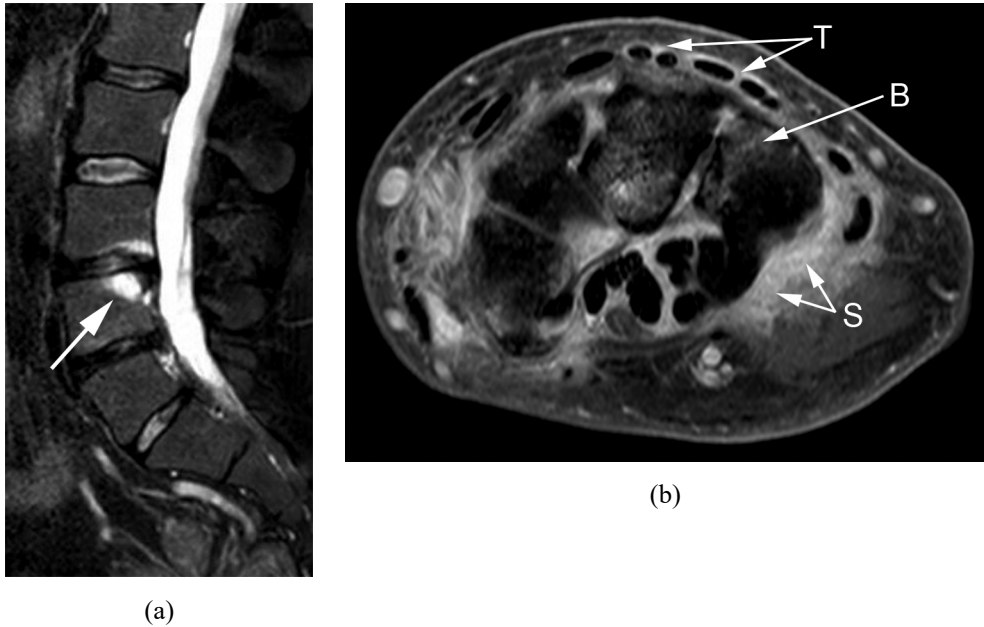


Figure 3. MRI-detected inflammatory features seen in axial SpA and RA. Axial SpA (a): STIR sagittal MRI of the lower spine, fat suppression forces healthy bone marrow to appear dark, while bone marrow edema in the vertebra appears as a region of high intensity (arrow). RA (b): T1-Gd axial MRI of the wrist, combination of fat suppression and post-contrast enhancement reveals bone marrow edema (B arrow), synovitis (S arrows), and tenosynovitis (T arrows) as regions of high intensity.

arthralgia to clinical arthritis [18,19]. This is highly relevant for early diagnosis of RA because arthralgia is the earliest phase at which symptoms of joint pain may prompt a patient to seek medical attention. As in the case of SpA, either a T2-weighted fat-saturated sequence or a STIR sequence can be applied to visualize BME in RA patients. However, these sequences do not allow for accurate evaluation of synovitis and tenosynovitis [20]. On the other hand, a T1-weighted fat-saturated sequence acquired after intravenous injection of a gadolinium contrast agent (T1-Gd) enables the visualization of all three inflammatory features [21] (Figure 3b).

Visual scoring and its limitations

At present, the most common approach to assessing inflammation on MRI of patients with RA and axial SpA is through visual scoring. Several scoring systems

have been proposed and validated over the past two decades [22–25]. The scoring is done semi-quantitatively, in the sense that readers visually approximate the volume of inflammation and assign an integer grade corresponding to that volume. In RA patients, for example, BME is scored on a 0–3 scale: 0, normal; 1, 1–33% of bone edematous; 2, 34–66%; 3, 67–100%. In axial SpA patients, one approach is to evaluate BME per vertebral unit (region between the mid-points of two adjacent vertebrae) on a 0–3 scale: 0, normal; 1, < 25% vertebral unit edematous; 2, 25–50%; 3, > 50%.

One common challenge of current scoring frameworks is that visual assessment is a laborious, time-consuming task, often involving a long list of anatomical locations viewed in multiple imaging planes, slices, and acquisition sequences – all of which requires the availability of trained, experienced readers. Visual scoring is also inherently subject to the simultaneous contrast effect [26] of the human visual system, which causes readers to perceive the same image intensity differently depending on the surrounding background intensities. This can introduce intra- and inter-reader variability in the perceived extent of inflammation. Furthermore, in a setting where follow-up and baseline scans are compared side by side, patient posture differences between scanning sessions complicate the comparison and do not allow for a simple voxel-wise overlay of images.

Computer-aided techniques may help overcome these limitations. Automating the evaluation of inflammation with quantitative measurements derived directly from the image data can standardize interpretation and alleviate the time burden and cost associated with visual scoring. Application of image registration techniques can offer new interactive ways of comparative visualization of baseline and follow-up scans. Ultimately, computer-aided evaluation would allow clinical researchers to dedicate more resources to analysis of the dynamics and pathology of the disease and may help make MRI screening more widely available as part of early identification of inflammatory arthritis.

Outline of this thesis

The main goal of this thesis is to develop computer-aided methods for assessment of MRI-detected inflammation with the aim of aiding early diagnosis of inflammatory arthritis. In particular, we address the tasks of comparative visualization, automatic quantification, and feature selection, as described in the following chapters:

Chapter 2 presents an interactive scoring tool for evaluation of inflammatory changes over time in patients with axial SpA. We use locally-rigid image registration to fuse baseline and follow-up MR scans of the spine into a single color-encoded image, allowing for direct visualization and assessment of inflammatory changes.

Chapter 3 investigates the feasibility of automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis. We develop an atlas-based framework that segments the carpal bones of the wrist joint and measures the presence of signal associated with bone marrow edema within the bones. Correlation between quantitative measurements and visual scores is assessed in a large cohort of early arthritis patients.

Chapter 4 investigates the feasibility of automatic quantification of tenosynovitis by extending and further developing the framework of **chapter 3** to measure tenosynovial inflammation around the extensor and flexor tendons of the wrist.

Chapter 5 sets out to identify MRI-detected inflammatory features specific to RA by comparing the difference in frequency of joint-level inflammation in RA patients and symptom-free volunteers. The identified subset of features is then used to predict progression from clinically suspect arthralgia to clinical arthritis.

Chapter 6 summarizes the findings of this thesis and discusses possible directions of future work.

REFERENCES

1. Scott IC, Steer S, Lewis CM, Cope AP. Precipitating and perpetuating factors of rheumatoid arthritis immunopathology – linking the triad of genetic predisposition, environmental risk factors and autoimmunity to disease pathogenesis. *Best Pract. Res. Clin. Rheumatol.* 2011;25:447–68.
2. Maksymowych WP, Brown MA. Genetics of ankylosing spondylitis and rheumatoid arthritis: where are we at currently, and how do they compare? *Clin. Exp. Rheumatol.* 27:S20-5.
3. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *Lancet.* Elsevier; 2010;376:1094–108.
4. Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, et al. Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. *Arthritis Rheum.* 1998;41:58–67.
5. NIH. Illustration of joint affected by rheumatoid arthritis [Internet]. 2007. Available from: https://commons.wikimedia.org/wiki/File:Rheumatoid_arthritis_joint.gif
6. Braun J, Sieper J. Ankylosing spondylitis. *Lancet.* 2007;369:1379–90.
7. Häggström M. Illustration of the sacroiliac joints [Internet]. 2011. Available from: https://commons.wikimedia.org/wiki/File:Sacroiliac_joint.svg
8. Senseiwa. Illustration of the ankylosing process [Internet]. 2007. Available from: https://commons.wikimedia.org/wiki/File:Ankylosing_process.jpg
9. Ajeganova S, Huizinga T. Sustained remission in rheumatoid arthritis: latest evidence and clinical considerations. *Ther. Adv. Musculoskelet. Dis.* SAGE Publications; 2017;9:249–62.
10. Landewé R, Sieper J, Mease P, Inman RD, Lambert RG, Deodhar A, et al. Efficacy and safety of continuing versus withdrawing adalimumab therapy in maintaining remission in patients with non-radiographic axial spondyloarthritis (ABILITY-3): a multicentre, randomised, double-blind study. *Lancet.* 2018;
11. López-Medina C, Dougados M, Collantes-Estévez E, Moltó A. Adherence to recommendations for the use of anti-tumour necrosis factor and its impact over 5 years of follow-up in axial spondyloarthritis. *Rheumatology.* 2018;57:880–90.
12. Sofka CM. Tracking Rheumatic Disease Through Imaging. *Rheum. Dis. Clin. North Am.* 2013;39:633–44.
13. Grigoryan M, Roemer FW, Mohr A, Genant HK. Imaging in spondyloarthropathies.

Curr. Rheumatol. Rep. 2004;6:102–9.

14. Krabben A, Stomp W, Huizinga TWJ, van der Heijde D, Bloem JL, Reijnierse M, et al. Concordance between inflammation at physical examination and on MRI in patients with early arthritis. *Ann. Rheum. Dis.* BMJ Publishing Group Ltd; 2015;74:506–12.

15. Rudwaleit M, Jurik AG, Hermann K-GA, Landewe R, van der Heijde D, Baraliakos X, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. *Ann. Rheum. Dis.* 2009;68:1520–7.

16. Hetland ML, Ejbjerg B, Hørslev-Petersen K, Jacobsen S, Vestergaard A, Jurik AG, et al. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann. Rheum. Dis.* 2009;68:384–90.

17. Nieuwenhuis WP, Krabben A, Stomp W, Huizinga TWJ, van der Heijde D, Bloem JL, et al. Evaluation of magnetic resonance imaging-detected tenosynovitis in the hand and wrist in early arthritis. *Arthritis Rheumatol.* (Hoboken, N.J.). 2015;67:869–76.

18. van Steenbergen HW, Mangnus L, Reijnierse M, Huizinga TWJ, van der Helm-van Mil AHM. Clinical factors, anticitrullinated peptide antibodies and MRI-detected subclinical inflammation in relation to progression from clinically suspect arthralgia to arthritis. *Ann. Rheum. Dis.* 2016;75:1824–30.

19. Kleyer A, Krieter M, Oliveira I, Faustini F, Simon D, Kaemmerer N, et al. High prevalence of tenosynovial inflammation before onset of rheumatoid arthritis and its link to progression to RA-A combined MRI/CT study. *Semin. Arthritis Rheum.* Elsevier; 2016;46:143–50.

20. Stomp W, Krabben A, van der Heijde D, Huizinga TWJ, Bloem JL, Østergaard M, et al. Aiming for a simpler early arthritis MRI protocol: can Gd contrast administration be eliminated? *Eur. Radiol.* 2015;25:1520–7.

21. Stomp W, Krabben A, van der Heijde D, Huizinga TWJ, Bloem JL, van der Helm-van Mil AHM, et al. Aiming for a shorter rheumatoid arthritis MRI protocol: can contrast-enhanced MRI replace T2 for the detection of bone marrow oedema? *Eur. Radiol.* 2014;24:2614–22.

22. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. *Ann. Rheum. Dis.* 2005;64:i3–7.

23. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Kvien TK. Introduction of a novel magnetic resonance imaging tenosynovitis score for rheumatoid arthritis: reliability in a multireader longitudinal study. *Ann. Rheum. Dis.* 2007;66:1216–20.
24. Haibel H, Rudwaleit M, Brandt HC, Grozdanovic Z, Listing J, Kupper H, et al. Adalimumab reduces spinal symptoms in active ankylosing spondylitis: clinical and magnetic resonance imaging results of a fifty-two-week open-label trial. *Arthritis Rheum.* 2006;54:678–81.
25. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Krishnananthan R, Stone M, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. *Arthritis Rheum.* 2005;53:502–9.
26. Heinemann EG. Simultaneous brightness induction as a function of inducing- and test-field luminances. *J. Exp. Psychol.* 1955;50:89–96.

2

Computer-aided evaluation of inflammatory changes over time on MRI of the spine in patients with suspected axial spondyloarthritis: a feasibility study

This chapter was adapted from:

E. Aizenberg, R. van den Berg, Z. Ez-Zaitouni, D. van der Heijde, M. Reijnierse, O. Dzyubachyk, B.P.F. Lelieveldt, “Computer-aided evaluation of inflammatory changes over time on MRI of the spine in patients with suspected axial spondyloarthritis: a feasibility study,” *BMC Medical Imaging*, 17:55, 2017.

Abstract

Purpose: Evaluating inflammatory changes over time on MR images of the spine in patients with suspected axial Spondyloarthritis (axSpA) can be a labor-intensive task, requiring readers to manually search for and perceptually align a set of vertebrae between two scans. The purpose of this study was to assess the feasibility of computer-aided (CA) evaluation of such inflammatory changes in a framework where scans from two time points are fused into a single color-encoded image integrated into an interactive scoring tool.

Methods: For 30 patients from the SPondyloArthritis Caught Early (SPACE) cohort (back pain ≥ 3 months, ≤ 2 years, onset < 45 years), baseline and follow-up MR scans acquired 9–12 months apart were fused into a single color-encoded image through locally-rigid image registration to evaluate inflammatory changes in 23 vertebral units (VUs). Scoring was performed by two expert readers on a (-2, 2) scale using an interactive scoring tool. For comparison of direction of change (increase/decrease) indicated by an existing reference, Berlin method scores ((-3, 3) scale) of the same MR scans from a different ongoing study were used. The distributions of VU-level differences between CA readers and between the CA and Berlin methods (sign of change scores) across patients were analyzed descriptively. Patient-level agreement between CA readers was assessed by intraclass correlation coefficient (ICC).

Results: Five patients were excluded from evaluation due to failed vertebrae segmentation. Patient-level inter-reader agreement ICC was 0.56 (95% CI: 0.22 to 0.78). Mean VU-level inter-reader differences across 25 patients ranged (-0.04, 0.12) with SD range (0, 0.45). Across all VUs, inter-reader differences ranged (-1, 1) in 573/575 VUs (99.7%). Mean VU-level inter-method differences across patients ranged (-0.04, 0.08) with SD range (0, 0.61). Across all VUs, inter-method differences ranged (-1, 1) in 572/575 VUs (99.5%).

Conclusion: Fusion of MR scans of the spine from two time points into a single color-encoded image allows for direct visualization and measurement of inflammatory changes over time in patients with suspected axSpA.

Introduction

Evaluating inflammatory changes over time on magnetic resonance (MR) images of the spine in patients with suspected axial Spondyloarthritis (axSpA) can be a labor-intensive task. Depending on the rheumatologic scoring method that is used, readers are often required to assess a set of vertebral units (VUs) in several slices [1,2], manually searching for and perceptually aligning the vertebrae between two scans. It would be of great benefit to have a computer-aided (CA) method capable of automatically localizing and labeling the VUs and spatially aligning scans from two time points, so voxel-wise intensity differences could be visualized in a single image.

CA methods involving alignment between multiple images for voxel-wise analysis have been extensively applied in the fields of neuroimaging and radiation therapy. Examples include voxel-based morphometry for comparison of local concentration of gray matter between subjects [3], analysis of multi-subject diffusion data for studying brain connectivity [4], and adaptive radiotherapy [5]. These studies have demonstrated that CA alignment of medical images can aid clinicians with automated biomarker quantification and treatment replanning based on anatomical changes that occur over time.

Spatial alignment of scans from two time points compensates for patient posture differences between scanning sessions and allows to overlay the two images for visual assessment of changes over time. This is done by computing a spatial coordinate mapping between corresponding locations in the two scans, a process known as image registration [6]. Generally, this mapping involves a geometrically non-rigid correspondence between voxels. This may cause physically implausible deformations in rigid anatomical structures, such as bones. An efficient solution to this problem was proposed by Dzyubachyk *et al.* [7] and applied to comparative visualization of whole-body MR scans in patients with multiple myeloma lesions. The highlight of this approach is that, following a global alignment of two time points, a locally rigid (rotation and translation only)

alignment is derived for selected regions of interest (ROIs) within bones. This ensures that bone rigidity is preserved in the final alignment.

In the work presented here, we applied the framework of Dzyubachyk *et al.* [7] to comparative visualization of MR images of the spine in patients with suspected axSpA. The aim of our study was to assess the feasibility of CA evaluation of axSpA inflammatory changes in the spine. This included fusion of scans from two time points into a single color-encoded image vividly distinguishing areas of increase versus decrease in inflammation over time, automatic labeling of VUs, and an interactive scoring module whose entry fields are activated/deactivated in synchronization with the VU selected by the reader in the image.

Methods

Patients

A total of 30 patients from the SPondyloArthritis Caught Early (SPACE) cohort were included in this feasibility study. The SPACE cohort has been described extensively before [8]. In short, the SPACE cohort is an ongoing cohort started in January 2009, including patients aged 16 years and older with chronic back pain (duration ≥ 3 months, ≤ 2 years, onset < 45 years). All patients underwent a diagnostic work-up at baseline, consisting of history taking, physical examination, laboratory tests, and imaging (MR imaging (MRI) and plain radiographs). Patients fulfilling the Assessment of SpondyloArthritis (ASAS) axSpA criteria [9,10] and patients with possible axSpA were included for follow-up visits after 3 and 12 months (including MRI). Possible axSpA was defined as the presence of at least one specific SpA-feature with a high positive likelihood ratio (LR+ above 6) or at least two less specific SpA-features (LR+ below 6), but not fulfilling the ASAS axSpA criteria [10,11].

MRI sequences

Patients underwent MRI of the complete spine in two stages (upper and lower spine) on a 1.5T MR system (Philips Medical Systems, The Netherlands). The acquired sequences were Short Tau Inversion Recovery (STIR) with repetition time (TR) 2500 ms, echo time (TE) 60 ms, inversion time 165 ms, acquisition matrix 304×300, echo train length (ETL) 25, number of averages 3 and T1-weighted Turbo Spin-Echo (TR 550/TE 10, acquisition matrix 512×305, ETL 5, number of averages 3). Imaging was performed in the sagittal plane with a field of view of 380×380 mm, slice thickness of 4 mm, and a slice gap of 0.4 mm.

Vertebrae localization/segmentation/labeling

For each patient, 23 VUs were automatically localized, segmented, and labeled. A VU is defined as the region between the mid-points of two adjacent vertebral bodies. For example, VU1 consists of the lower endplate of vertebra C2 and the upper endplate of vertebra C3. Hence, VU levels 1–23 cover 24 vertebral bodies (C2–S1). Localization and segmentation were carried out using atlas-based segmentation [12]. The atlas set consisted of 11 patients from the SPACE cohort (no overlap with patients included in evaluation). For each atlas patient, 24 vertebral bodies (C2 to S1) were manually outlined in the slice closest to the mid-sagittal plane and the two adjacent slices (a total of three slices). The procedure was carried out separately for upper and lower spine images, producing a total of two manually segmented images per atlas patient. We chose to approximate each vertebral body with a simple polygonal region within the vertebral borders, taking the cortex as an anatomical boundary. This choice was motivated by the fact that for successful locally rigid alignment of two time points it is preferable to have a ROI estimate that under-segments the bone, rather than an estimate that spills over into inherently non-rigid neighboring soft tissue.

The first phase of atlas-based segmentation consisted of image registration between each of the 11 atlas patients and the target patient being segmented. Image registration was performed using the Elastix software package [13,14]. After

spatially mapping vertebrae ROIs from every atlas image onto the target image, a majority vote was applied across all mappings to determine whether a voxel was part of the background or of one of the vertebrae.

Labeling of vertebrae voxels in the upper spine image was done sequentially from top to bottom, over connected components, with the top-most connected component receiving the label “C2,” the following “C3,” etc. Similarly, labeling in the lower spine image was done sequentially from bottom to top, with the bottom-most connected component receiving the label “S1,” the following “L5,” etc. We used a 26-connected neighborhood definition for connectivity in 3D. Connected components less than 20 voxels in size were considered to be noise and were removed.

Locally rigid inter-time point alignment

In what follows, let us consider a pair of MR scans of a single patient and, without loss of generality, refer to one of the scans as “Time Point 1 (TP1)” and the second scan as “Time Point 2 (TP2).” According to the framework proposed by Dzyubachyk *et al.* [7], locally rigid alignment of two images is derived from a global non-rigid alignment of this image pair. We used the Elastix software package [13,14] to globally align TP2 to TP1. The registration yielded a deformation field specifying for each physical position in TP1 the corresponding physical position in TP2. Next, for each VU, the landmark transform [15] was used to estimate a locally rigid alignment between the VU region in TP1 (specified by the atlas-based segmentation result) and the corresponding physical region in TP2 (specified by the deformation field) [7]. This ensured that spatial correspondence between voxels within the VU in TP1 and TP2 was restricted to translation and rotation, preserving bone rigidity.

It is important to note that the described method can be equivalently applied in the reverse direction, by globally aligning TP1 to TP2, and subsequently using VU segmentations from TP2. Thus, in order to align two scans in a locally rigid manner, it is sufficient to segment and label vertebrae in one of the two scans.

Color-encoded fusion of time points

After locally aligning two time points on the VU level, differences in intensity (e.g. inflammation) between corresponding voxels were visualized through color-encoded fusion of the two scans. First, intensity values of TP1 were color-mapped to orange color space (RGB triple {255,128,0}), and intensity values of TP2 were color-mapped to light blue color space (RGB triple {0,127,255}). The fusion image was then obtained by voxel-wise superposition of the two color-mapped images [7]. Since orange and light blue are complementary colors, areas where no changes occurred between the two time points (TP2 intensity = TP1 intensity) are displayed in shades of gray. On the other hand, an increase in inflammation over time (TP2 intensity > TP1 intensity) is displayed in shades of light blue (Figure 1). In the opposite case, a decrease in inflammation over time (TP2 intensity < TP1 intensity) is displayed in shades of orange. In addition to its complementary nature, the choice of orange and light blue is motivated by the fact that these two colors can be perceived even by readers with color vision deficiency [7]. No intensity standardization was applied to original images prior to color-encoded fusion.

Evaluation of inflammatory changes

Two experienced readers (RvdB and ZEZ) independently evaluated inflammatory changes between MR scans of the spine (STIR only), acquired 9–12 months apart, directly from the color-encoded fusion image. The choice of using only STIR images for CA scoring was motivated by our focus on inflammatory lesions and the fact that automatic alignment of T1 images to STIR images requires additional image registration steps, which would introduce additional sources of error. The readers were blinded to the original images and their time order, as well as patient and clinical characteristics. Each VU was assigned a score ranging from -2 (dramatic decrease of inflammation), via 0 (no change), to +2 (dramatic increase of inflammation), reflecting net change in the degree of inflammation within the VU. Navigation through the images and evaluation were carried out using an interactive software tool that we implemented in MeVisLab 2.7.1 (MeVis Medical Solutions,

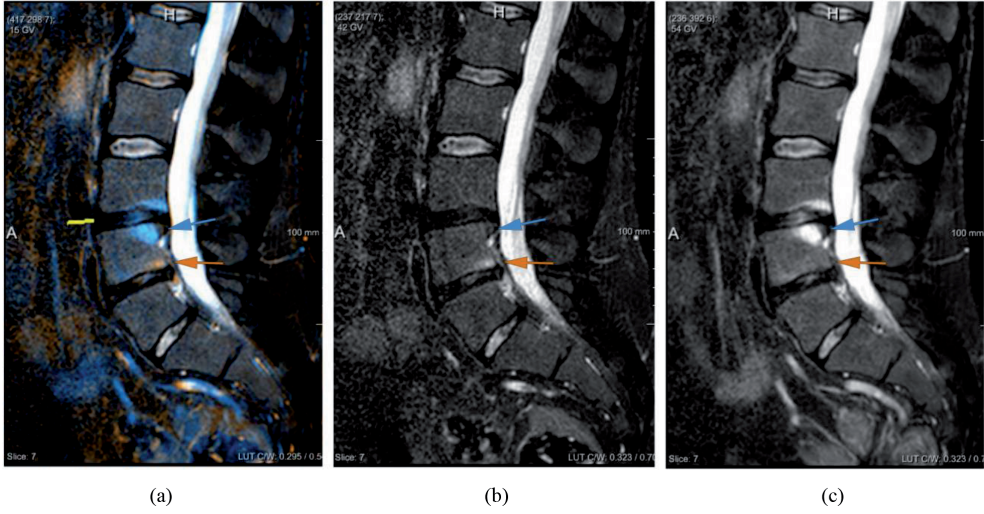


Figure 1. Color-encoded fusion of two MR scans of the same subject acquired at two different time points. Inflammation increase (blue arrow) in VU21 and decrease (orange arrow) in VU22 in the second time point (c) compared to first time point (b) are displayed in blue and orange, respectively, in the color-encoded fusion image (a). In this example, the locally rigid alignment is applied to VU21, indicated by the yellow line in (a).

Germany) [16]. The tool consists of two windows: the comparative visualization module (Figure 2) and the scoring module (Figure 3).

For comparison of direction of change (increase/decrease) indicated by an existing reference, Berlin method [1] scores of the same pair of MR scans from a different ongoing SPACE cohort study at our institution were used. The MR scans (STIR and T1) were independently evaluated by two experts (MdH and PACB) according to the Berlin method [1], yielding status scores for each of the time points. Each VU was assigned a score ranging from 0 to 3 reflecting the fraction of bone volume affected by bone marrow edema: 0, normal; 1, < 25% VU edematous; 2, 25–50%; 3, > 50%. The readers were blinded to the time order of the images, as well as patient and clinical characteristics. Changes in inflammation over time were calculated as differences in status scores after de-blinding the time order of the MR scans.

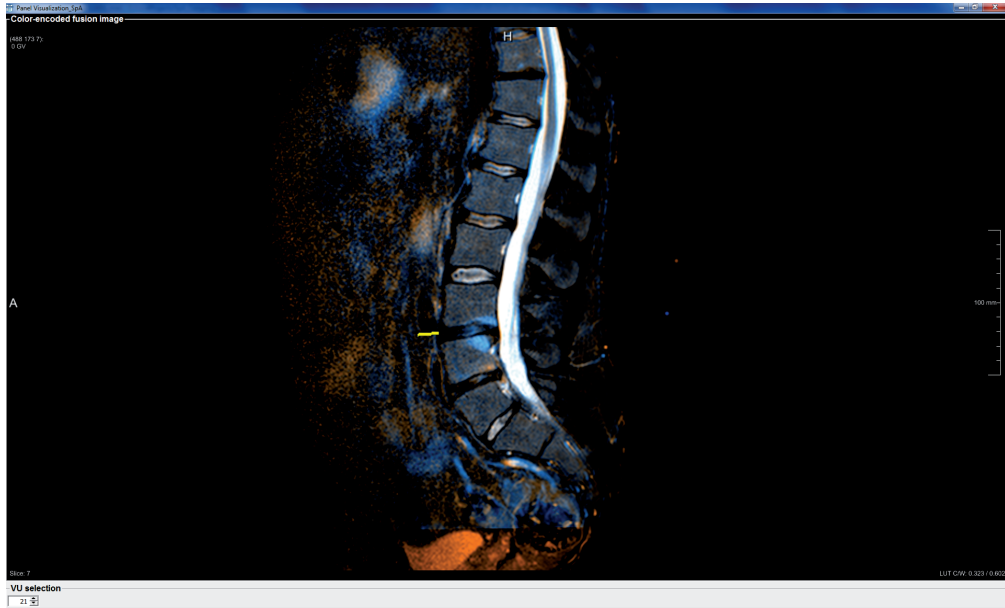


Figure 2. Comparative visualization module. The module displays the color-encoded fusion image and allows the user to specify the VU of interest in the VU selection field at the bottom left of the window, which will trigger locally rigid alignment of two time points for that VU. A visual indication for the position of the VU in the image is provided to aid navigation (yellow line next to VU 21).

Statistical analysis

For each of the 23 VU levels, the distributions of VU-level inter-reader and inter-method differences across patients were analyzed descriptively. For inter-reader differences, the VU-level difference was computed between change scores assigned to the VU by the two CA readers. For inter-method differences, the focus was on the direction of change indicated by each method, and therefore, VU-level difference was computed between the sign of the CA change score (mean of two readers) and the sign of the Berlin change score (mean of two readers), where the sign function takes the value -1 in case of negative change, +1 in case of positive change, and 0 in case of no change.

Agreement between CA readers was assessed on the patient level (change summed across VUs of each patient) by computing intraclass correlation coefficient (ICC, two-way mixed, single measures, absolute agreement definition).

The statistics were computed using MATLAB R2015b (The MathWorks, Inc., USA) and IBM SPSS Statistics 23 (IBM Corporation, USA).

The screenshot displays a software interface titled "Panel ScoringModule_SpA". It contains a vertical list of 23 panels, each representing a Vertebral Unit (VU) from VU 12 to VU 23. Each panel has a header with the VU number and a sub-header "Inflammation present" followed by a checkbox. Below the checkbox are five radio buttons labeled -2, -1, 0, +1, and +2. Panel VU 21 is currently active, highlighted with a green background, and its "-2" radio button is selected. The other panels are deactivated, with their checkboxes disabled and radio buttons greyed out. At the bottom of the interface, there is a button labeled "Save scores to file".

Figure 3. Scoring module. The module acts as an interactive scoring sheet, consisting of 23 panels representing the VUs. Every panel contains a group of option buttons (only one of the options can be selected) through which the reader assigns a change score to the VU, as well as a checkbox to indicate the presence of inflammation in cases of no net change. Only one VU panel is active at any given moment. The choice of VU in the comparative visualization module activates the corresponding panel in the scoring module, while deactivating panels of the remaining 22 VUs. This ensures that one VU is not mistaken for another while filling out the interactive scoring sheet.

Results

In 18/30 patients, atlas-based segmentation provided satisfactory segmentation and correct labeling of all 23 VUs in at least one of the time points (as explained above, in order to align two scans in a locally rigid manner, it is sufficient to segment and label vertebrae in one of the two scans). In seven patients, failure to segment the lowest vertebra in the upper spine image and/or the highest vertebra in the lower spine image, resulted in lack of segmentation for one VU (frequently corresponding to the levels Th9–Th11). The segmentations for these VUs were added by manual correction. Five patients were excluded from further evaluation due to inaccurate alignment with atlas images that led to missing vertebrae segmentations and incorrect labeling of VUs. Thus, a total of 25 patients (and

hence 575 VUs) were evaluated. Baseline patient characteristics and descriptive statistics of Berlin and CA scores at baseline and over time are presented in Table 1. As demonstrated by baseline characteristics, it should be pointed out that most patients had low levels of inflammation.

Table 1. Baseline patient characteristics and descriptive statistics of Berlin and CA scores of the 25 patients evaluated in the study

Baseline patient characteristics		
Characteristic	Patients (<i>n</i> = 25)	
Age at inclusion in years, mean (SD)	31.7 (8.3)	
Male, n (%)	12 (48)	
Duration of back pain in months, mean (SD)	14.4 (8.0)	
IBP, n (%)	19 (76)	
HLA-B27 positivity, n (%)	15 (60)	
Elevated CRP, n (%)	6 (24)	
Sacroiliitis on MRI (ASAS definition), n (%)	8 (32)	
Sacroiliitis on radiograph, n (%)	3 (12)	
Positive MRI (ASAS definition), n (%)	2 (8)	
ASAS classification positive, n (%)	14 (56)	
Berlin and CA scores descriptive statistics		
Variable	Berlin method (reader 1 / reader 2)	CA method (reader 1 / reader 2)
VU-level score at baseline, median (range)	0 (0,1) / 0 (0,1)	NA
VU-level score at follow-up, median (range)	0 (0,1) / 0 (0,1)	NA
Patient-level score at baseline, median (range)	0 (0,5) / 0 (0,3)	NA
Patient-level score at follow-up, median (range)	1 (0,5) / 0 (0,4)	NA
Change in VU-level score, median (range)	0 (-1,1) / 0 (-1,1)	0 (-1,2) / 0 (-2,2)
Change in patient-level score, median (range)	0 (-2,3) / 0 (-1,2)	0 (-3,3) / 0 (-11,5)

Inter-reader differences between CA readers

VU-level differences between CA readers' change scores are shown in Figure 4a. Mean VU-level differences across patients ranged from -0.04 to 0.12 with standard deviation (SD) range (0, 0.45). Most differences were observed in the lower thoracic spine and the lumbar spine. In 21/23 VU levels, differences ranged between -1 and 1 across all patients. In 2/23 VU levels, a difference of 2 was observed in one patient. In total, across all patients, VU-level differences ranged (-1, 1) in 573/575 VUs (99.7%). On the patient level, the ICC between the two CA readers was 0.56 (95% confidence interval (CI): 0.22 to 0.78), indicating moderate agreement between readers.

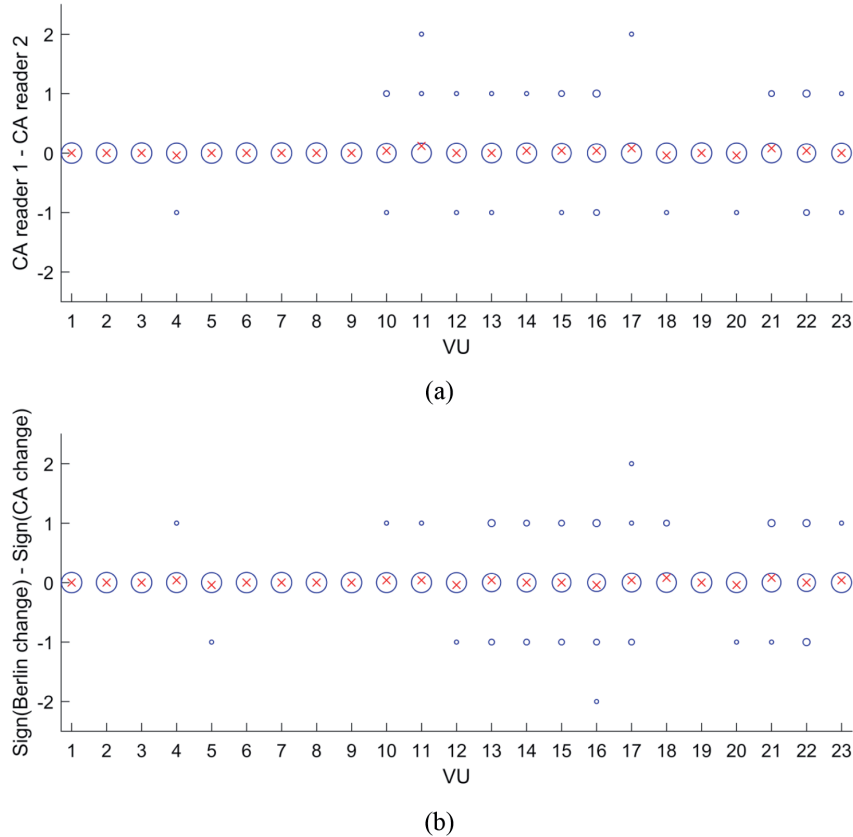


Figure 4. VU-level difference between CA readers' change scores (a) and VU-level difference between sign of CA and Berlin change scores (b). Exact difference values are shown in blue (size of bubble data points is proportional to the occurrence of the difference value across 25 patients). Mean VU-level differences across 25 patients are shown in red.

Inter-method differences between CA and Berlin methods

VU-level differences between the direction of change indicated by the CA and Berlin methods are shown in Figure 4b. Mean VU-level differences across patients ranged from -0.04 to 0.08 with SD range (0, 0.61). Most differences were observed in the lower thoracic spine and the lumbar spine. In 21/23 VU levels, differences ranged between -1 and 1 across all patients. In 1/23 VU levels, a difference of 2 (positive Berlin change, negative CA change) was observed in one patient. In 1/23 VU levels, a difference of -2 (negative Berlin change, positive CA change) was observed in one patient. In total, across all patients, VU-level differences ranged (-1, 1) in 572/575 VUs (99.5%). Differences of precisely -1 or 1 (change detected only by one of the two methods) were observed in 40/575 VUs, and among those in 33/40 VUs the change was detected by the CA method while Berlin score indicated zero change. Figure 5 shows examples of VU-level differences between the two methods.

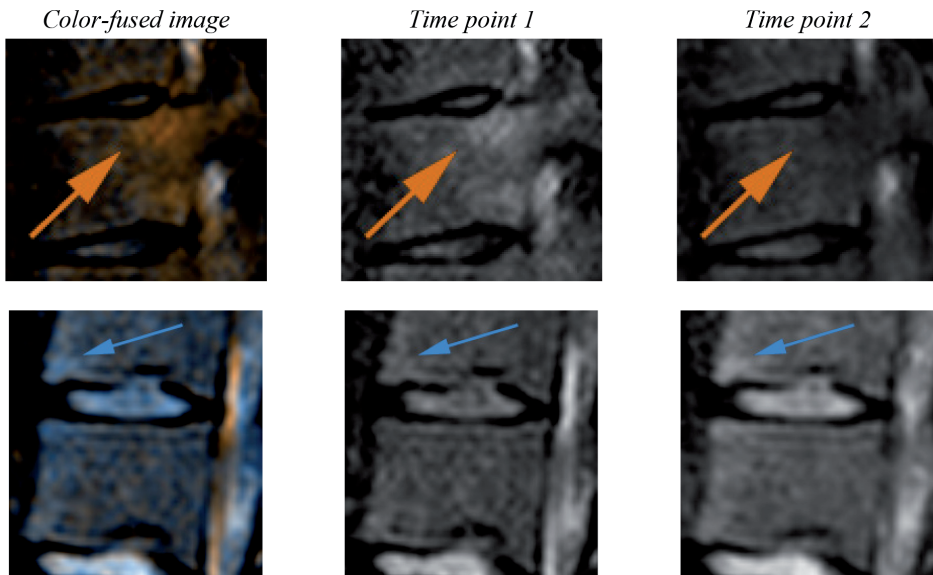


Figure 5. Examples of VU-level inter-method differences. Top row: lesion area (orange arrow) received a CA change score of -1, but a Berlin change score of 0, because of being considered a degenerative lesion (status scores = 0). Bottom row: lesion area (blue arrow) received a CA change score of 1, but a Berlin change score of 0, because of zero Berlin status scores.

Discussion

In this study, we assessed the feasibility of CA evaluation of inflammatory changes on MR scans of the spine in patients with suspected axSpA. Readers agreed that a key advantage of CA evaluation is that fusion of two scans acquired at different time points into a single color-encoded image allows for direct visualization and measurement of inflammatory changes, as opposed to derivation of change scores from status scores that measure presence and extent of lesions separately for each time point. The results indicate that in nearly all VUs of all patients, VU-level differences between CA readers and between the CA and Berlin methods were bounded between -1 and 1, ensuring that scores do not offer opposing opinions on the direction of inflammatory change (increase versus decrease). The fact that most differences occurred in the lower thoracic spine and the lumbar spine is consistent with the observation that most inflammatory activity in the spine of early disease patients takes place in these regions [17,18]. The majority of non-zero differences between the CA and Berlin scores were observed when change was detected by the CA method while zero change was indicated by the Berlin method. These quantitative results support our qualitative observation that small gradual changes in an existing lesion are often not reflected in Berlin status scores, but can be readily visualized and measured by the CA method.

The moderate inter-reader patient-level agreement and difference in the range of readers' scores suggest that the CA grading scale may be defined too loosely with respect to affected bone volume, making the score more susceptible to subjective interpretation of the degree of change. Readers pointed out that a challenging aspect of the CA method is estimation of net inflammatory change in VUs with multiple inflammatory lesions. For example, one such VU exhibited increase in one quadrant, while exhibiting decrease in another quadrant. The two readers had different opinions as to which change was stronger, resulting in opposing scores and thus a mean change of zero. One way to overcome such discrepancies would be to score change separately for each of the four quadrants,

akin to scoring in the Spondyloarthritis Research Consortium of Canada (SPARCC) method [2].

For the purpose of this feasibility study, we made a deliberate decision to measure change based only on the color-fused image, while blinding readers to the original images. However, the readers noted that in daily practice it would be helpful to have the original images (STIR and T1) available next to the color-fused image, as this would contribute to a more informative scoring decision. The color-fused image could then be used as a map that directs the reader's attention to locations of potential inflammatory changes, while the original images would be used to make the final judgement about the type and degree of observed change. The reader would still benefit from locally rigid alignment between the two time points while assessing original images, since the two scans will be aligned such that the VU of interest has identical viewing points in both images. Another feature that would enhance user experience is stitching of upper and lower spine images into a single image. This would offer a more natural workflow, without the need to load two separate images for every patient. A simultaneous view of the complete spine would also facilitate a more holistic assessment of disease activity.

This study has several important limitations. The SPACE cohort consists only of early disease patients with low levels of inflammation, making it harder to study inflammatory changes, since changes were infrequent. Another limitation is that it was not possible to provide patient-level inter-method agreement statistics, as the scoring methodology and scale range are different, and this would result in uninterpretable results. However, the CA method was not designed with the aim of replicating the Berlin method, but rather as an independent scoring framework. It is of interest to compare the responsiveness of the two methods by quantifying sensitivity to change in a population with treatment and placebo patient groups, as was previously done for other scoring methods [19]. Assessing responsiveness after an effective intervention could provide information on differences in the psychometric characteristics of the two scoring methods. This could be addressed in a follow-up study. The definition of the CA change score should also provide

clear guidelines for the case of degenerative lesions to avoid discrepancies with existing methods that do not score these lesions (Figure 5). The lack of intensity standardization prior to color-encoded fusion is a potential source of error. However, we should note that standardization is also not applied in the long-established procedure of Berlin method scoring. We sought to explore the feasibility of color-encoded fusion without additional image post-processing steps that are not present in the Berlin method workflow. Future studies should indeed investigate the effect of intensity standardization on change scores of both methods. An additional limitation is that CA scoring was performed only using STIR images. This differs from the common clinical approach of confirming inflammatory lesions observed in STIR images as low intensity areas in T1 images. Inclusion of T1 images may improve the robustness of the scoring method. Furthermore, it might allow visualization of changes in structural lesions, such as fatty lesions. Finally, it should be recognized that since 5/30 patients had to be excluded due to failed segmentation and 7/30 segmentations had to be manually adjusted, the method is not yet sufficiently robust to be used in practice. We should stress, however, that this study did not attempt to solve the problem of vertebrae segmentation in MRI. Our goal was to explore the prospect of CA assessment of patients with suspected axSpA and thereby provide yet another stimulus for development of robust vertebrae segmentation methods for MRI.

Although this study does not focus on the topic of vertebrae segmentation, we can note potential directions for improving the atlas-based segmentation framework used in this study. To ensure the applicability of this segmentation framework to a variety of MRI acquisition protocols and scanners, it would be helpful to construct an atlas consisting of sub-atlases of MR images acquired under similar echo/repetition times and magnetic field strengths. Then, prior to segmenting a target image, the system would automatically identify the most appropriate sub-atlas based on acquisition parameters recorded in the image's DICOM data. Additional improvement in segmentation might be achieved by operating with stitched images of the spine, as opposed to separate upper/lower

spine images. We have observed that in some cases segmentation was successful in one part of the upper/lower pair but failed in the other. Therefore, we hypothesize that the more easily matched spine region can “pull” the second spine region into its correct position in the target image when stitched.

Conclusion

This feasibility study has demonstrated that fusion of MR scans of the spine from two time points into a single color-encoded image allows for direct visualization and measurement of inflammatory changes over time in patients with suspected axSpA. A future study, with similar design to that of Lukas *et al.* [19], should assess the performance of the CA method in patients with a wide range of activity at baseline and follow-up, quantifying inter-reader reliability, sensitivity to change, and time needed to score each set of MR images. This would also provide a comprehensive comparison of the CA method to the Berlin and SPARCC methods.

Acknowledgements

We would like to thank Manouk de Hooge and Pauline A. C. Bakker for the Berlin method scoring of the patients in this study. We also thank Freek de Bruin for the insightful discussions that helped advance this study.

REFERENCES

1. Haibel H, Rudwaleit M, Brandt HC, Grozdanovic Z, Listing J, Kupper H, et al. Adalimumab reduces spinal symptoms in active ankylosing spondylitis: clinical and magnetic resonance imaging results of a fifty-two-week open-label trial. *Arthritis Rheum.* 2006;54:678–81.
2. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Krishnananthan R, Stone M, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. *Arthritis Rheum.* 2005;53:502–9.
3. Ashburner J, Friston KJ. Voxel-Based Morphometry—The Methods. *Neuroimage.* 2000;11:805–21.
4. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *Neuroimage.* 2006;31:1487–505.
5. Wang H, Dong L, O’Daniel J, Mohan R, Garden AS, Ang KK, et al. Validation of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy. *Phys. Med. Biol. IOP Publishing;* 2005;50:2887–905.
6. Rueckert D, Schnabel JA. *Medical Image Registration.* Springer Berlin Heidelberg; 2010. p. 131–54.
7. Dzyubachyk O, Blaas J, Botha CP, Staring M, Reijnierse M, Bloem JL, et al. Comparative exploration of whole-body MR through locally rigid transforms. *Int. J. Comput. Assist. Radiol. Surg.* 2013;8:635–47.
8. van den Berg R, de Hooze M, van Gaalen F, Reijnierse M, Huizinga T, van der Heijde D. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. *Rheumatology (Oxford).* 2013;52:1492–9.
9. Rudwaleit M, Landewé R, van der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann. Rheum. Dis.* 2009;68:770–6.
10. Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann. Rheum. Dis.*

BMJ Publishing Group Ltd and European League Against Rheumatism; 2009;68:777–83.

11. Rudwaleit M, Heijde D van der, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. *Ann. Rheum. Dis.* BMJ Publishing Group Ltd and European League Against Rheumatism; 2004;63:535–43.

12. Rohlfing T, Brandt R, Menzel R, Russakoff DB, Maurer CRJ. Quo vadis, atlas-based segmentation? *Handb. Med. Image Anal.* - Vol. III Regist. Model. 2005. p. 435–486.

13. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging.* 2010;29:196–205.

14. Aizenberg E. Elastix parameters for atlas-based segmentation of vertebral bodies [Internet]. 2016. Available from: <http://elastix.bigr.nl/wiki/index.php/Par0042>

15. Horn BKP. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A.* 1987;4:629.

16. MeVisLab [Internet]. Available from: <http://www.mevislabs.de>

17. Althoff CE, Sieper J, Song I-H, Haibel H, Weiß A, Diekhoff T, et al. Active inflammation and structural change in early active axial spondyloarthritis as detected by whole-body MRI. *Ann. Rheum. Dis.* BMJ Publishing Group Ltd and European League Against Rheumatism; 2013;72:967–73.

18. Lorenzin M, Ortolan A, Frallonardo P, Vio S, Lacognata C, Oliviero F, et al. Spine and sacroiliac joints on magnetic resonance imaging in patients with early axial spondyloarthritis: prevalence of lesions and association with clinical and disease activity indices from the Italian group of the SPACE study. *Reumatismo.* 2016;68:72.

19. Lukas C, Braun J, van der Heijde D, Hermann K-GA, Rudwaleit M, Østergaard M, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. *J. Rheumatol.* 2007;34:862–70.

3

Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis

This chapter was adapted from:

E. Aizenberg, E.A.H. Roex, W.P. Nieuwenhuis, L. Mangnus, A.H.M. van der Helm-van Mil, M. Reijnerse, J.L. Bloem, B.P.F. Lelieveldt, B.C. Stoel, “Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: a feasibility study,” *Magnetic Resonance in Medicine*, vol. 79(2), pp. 1127–1134, 2018.

Abstract

Purpose: To investigate the feasibility of automatic quantification of bone marrow edema (BME) on MRI of the wrist in patients with early arthritis.

Methods: For 485 early arthritis patients (clinically confirmed arthritis of ≥ 1 joint, symptoms for < 2 years), MR scans of the wrist were processed in three automatic stages. First, super-resolution reconstruction was applied to fuse coronal and axial scans into a single high-resolution three-dimensional image. Next, the carpal bones were located and delineated using atlas-based segmentation. Finally, the extent of BME within each bone was quantified by identifying image intensity values characteristic of BME by fuzzy clustering and measuring the fraction of voxels with these characteristic intensities within each bone. Correlation with visual BME scores was assessed through Pearson correlation coefficient.

Results: Pearson correlation between quantitative and visual BME scores across 485 patients was $r = 0.83$, $P < 0.001$.

Conclusion: Quantitative measurement of BME on MRI of the wrist has potential to provide a feasible alternative to visual scoring. Complete automation requires automatic detection and compensation of acquisition artifacts.

Introduction

The presence of bone marrow edema-like abnormalities (BME) has been shown to be a strong predictor of radiographic progression in rheumatoid arthritis (RA) patients [1–4], and is therefore an important biomarker in early arthritis. Evaluation of BME is done on MRI, where it is visually scored based on the Outcome Measures in Rheumatology RA-MRI Scoring (RAMRIS) system [5,6]. This scoring method requires a trained reader to visually estimate the volume of BME. Such estimates are challenging and time-consuming because of the need to assess multiple imaging planes and slices and are inherently undermined by the simultaneous contrast effect [7–9] of the human visual system, which causes the reader to perceive the same image intensity value differently depending on surrounding background intensities.

An automatic and quantitative approach to evaluating BME on MR scans could overcome the limitations of visual scoring by offering high-precision measurements derived directly from three-dimensional (3D) image data. It could alleviate the time burden of training and manual scoring for clinical researchers and could facilitate the use of MRI in drug evaluation studies, where employing a trained team of readers is costly.

Several previous studies on BME quantification in the wrist joint [10–12] relied on a semi-automatic method proposed by Li *et al.* [13]. However, this technique requires an expert to manually delineate non-edema and edema regions of interest within every bone that needs to be evaluated. These studies were also limited to a small sample size of fewer than 20 subjects. One related study focuses on fully automatic BME quantification in the knee joint [14], but it is not directly clear how to extend the bone segmentation method [15] to a joint with more than two bones, as is the case in the wrist.

In the work presented here, we developed an automatic framework for measuring the fraction of bone volume affected by BME in the eight carpal bones of the wrist joint. In contrast to previous methods, we used atlas-based segmentation to automatically locate and delineate the carpal bones. Our aim was

to investigate the feasibility of BME quantification through such atlas-based approach and assess the correlation between quantitative measurements and visual BME scores in a large cohort of early arthritis patients.

Methods

Patients

A total of 573 early arthritis patients from the Leiden Early Arthritis Clinic (EAC) cohort [16] (mean age, 54.7 years; age range, 18.1–87.9 years) were studied: 354 female (mean age, 53.0 years; age range, 18.7–85.3 years) and 219 male (mean age, 57.5 years; age range, 18.1–87.9 years) patients. Inclusion required clinically confirmed arthritis by physical examination in ≥ 1 joint and symptom duration of < 2 years. MR scans were obtained for the wrist joint of the most painful side (or the dominant side in cases of equally severe symptoms on both sides). The study was approved by the institutional medical ethics committee and all participants provided written informed consent.

MRI sequences

The wrist joint was scanned with an ONI MSK Extreme 1.5T extremity MR scanner (GE, Wisconsin, USA) with a 100 mm coil. Before contrast agent injection, T1-weighted fast spin-echo (FSE) sequence (T1) was acquired in the coronal plane with repetition time (TR) of 650 ms, echo time (TE) of 17 ms, acquisition matrix 388×288 , echo train length (ETL) 2. After intravenous injection of Gd-chelate (gadoteric acid, Guerbet, Paris, France, standard dose of 0.1 mmol/kg), T1-weighted FSE sequence with frequency-selective fat saturation (T1-Gd) was obtained in the coronal plane (TR 650/TE 17, acquisition matrix 364×224 , ETL 2) and the axial plane (TR 570/TE 7, acquisition matrix 320×192 , ETL 2). Coronal sequences were acquired with a slice thickness of 2 mm and a slice gap of 0.2 mm. Axial sequences were acquired with a slice thickness of 3 mm and a slice gap of 0.3 mm. The use of a T1-Gd sequence instead of a T2-weighted fat-saturated sequence is a validated modification that has been shown to perform

equally well in the depiction of BME and allows for a faster scanning protocol [17,18], which in turn reduces patient discomfort. Safety risk was minimized to the degree possible by the use of a macrocyclic contrast agent [19,20].

Visual scoring of BME

BME was assessed in line with the definitions proposed by RAMRIS [5] with validated modification of substituting T2-weighted fat-saturated sequence with T1-Gd sequence [17,18]. BME was independently scored by two trained readers who were blinded to clinical data on a 0–3 scale based on the estimated fraction of affected bone volume: 0, no BME; 1, 1–33% of bone edematous; 2, 34–66%; 3, 67–100%. The within-reader intra-class correlation coefficients (ICCs) for the total inflammation score were 0.98 and 0.93; the between-reader ICC was 0.95. The mean BME score of the two readers was considered.

Patients for which at least one reader marked one or more bones as unscorable (typically due to fat suppression issues) were excluded ($n = 11$). Patients whose T1-Gd images suffered from incomplete fat suppression, but still considered scorable by readers based on T1 images showing low signal intensity in the matching areas with BME on T1-Gd were retained.

Quantitative image analysis framework

Our automatic framework consisted of three stages. First, super-resolution reconstruction was applied to fuse coronal and axial T1-Gd scans into a single high-resolution 3D image. Next, the carpal bones were located and delineated using atlas-based segmentation. Finally, the extent of BME within each bone was quantified by identifying image intensity values characteristic of BME by fuzzy clustering and measuring the fraction of voxels with these characteristic intensities within each bone. Note that since the super-resolution reconstruction step requires a coronal and axial scan of the same sequence as input, this stage, and therefore the entire framework, could only be applied to T1-Gd scans. Therefore, pre-contrast T1

images, which were acquired only in the coronal plane, were not used in the quantitative image analysis framework.

Super-resolution reconstruction

When readers evaluate BME visually, they make use of two complementary scans: one acquired in the coronal plane and the second in the axial plane. This is due to the fact that slice thickness in each of the scans (2 mm in coronal; 3 mm in axial) is much larger than the in-plane spacing between voxels (~ 0.2 mm). Therefore, one scan compensates for anatomical detail lost in the other scan, allowing the reader to perceptually form a more complete assessment of the anatomy. Naturally, this raises the question how to simulate such perceptual fusion of two images on the computer, in order to obtain a single 3D image with isotropic voxels and high resolution in all three viewing planes. This type of problem, reconstruction of a high-resolution image of an object from multiple low-resolution images of the same object, is commonly referred to as super-resolution reconstruction (SRR).

A variety of SRR methods have been proposed for MRI [21–24]. In this study, we applied the method developed by Poot *et al.* [24]. This algorithm belongs to the family of spatial domain SRR methods, which construct a linear model of the image acquisition system and reconstruct the high-resolution image by solving a system of linear equations. This system is often underdetermined, as in our case, and is solved by applying regularization. We used Laplacian regularization with parameter $\lambda = 0.05$. This value was optimized in an experiment by two expert radiologists (MR and JLB) to provide satisfactory balance between image noise/artifacts and visual clarity of BME, synovial tissue, cartilage, and fluid around tendons. Prior to applying SRR, the axial scan was spatially aligned to the coronal scan using the Elastix software package [25,26], axial image intensity was linearly matched to the coronal image intensity, and the field of view of both images was restricted to the overlapping physical space between the two scans. Figure 1 shows an example of applying SRR to a pair of coronal and axial scans.

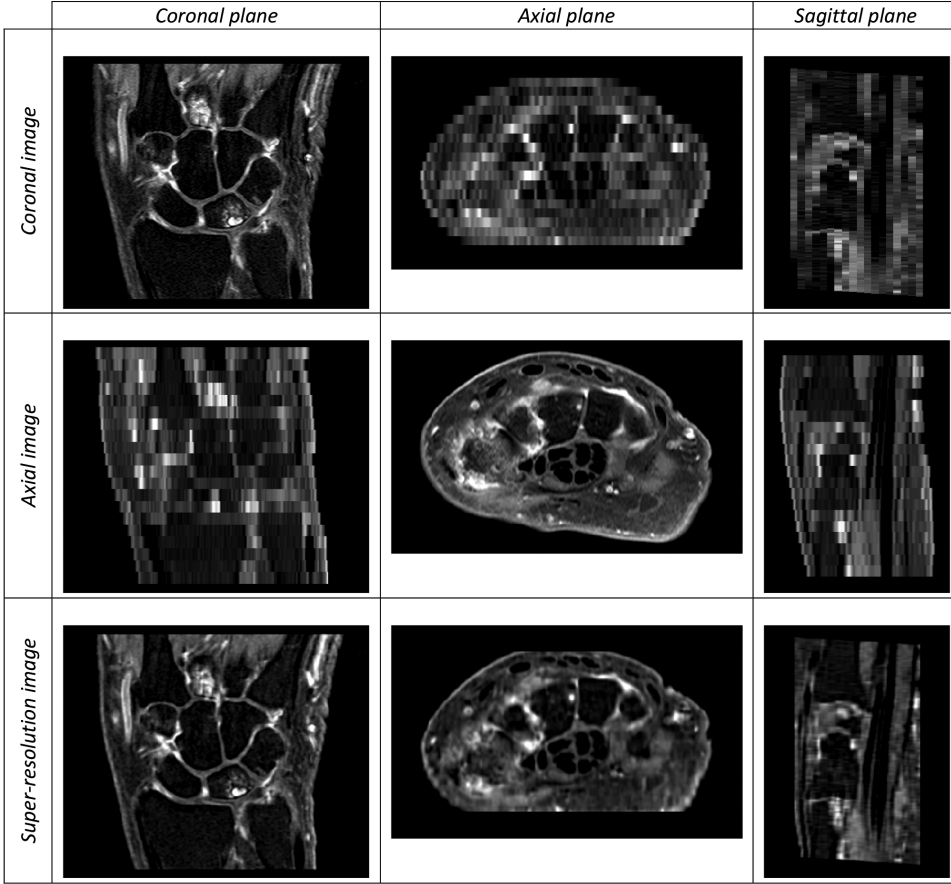


Figure 1. Coronal, axial, and super-resolution images (top to bottom rows, respectively) and their coronal, axial, and sagittal viewing planes (left to right columns, respectively). The original scans exhibit high resolution only in one plane, while the super-resolution image exhibits high resolution in all three planes.

Segmentation of carpal bones

The carpal bones were located and delineated using atlas-based segmentation (ABS) [27]. The atlas consisted of 13 early arthritis patients. For each atlas patient, the carpal bones were manually segmented in the coronal and axial T1-Gd images, yielding two segmentation images. The voxels of these manual segmentation images were assigned an integer bone label value ranging from 1 to 8 in locations corresponding to one of the eight carpal bones, or otherwise the value 0 in locations outside the bones. Then, separately for each bone, the two manual segmentation images were fused using SRR. Voxels with values above 78% of the

bone label value were assigned the bone label value, and the remaining voxels were zeroed to discard noise. The resulting eight images were superimposed to obtain the complete segmentation image in high-resolution space.

The first phase of the ABS routine consisted of image registration between each of the 13 atlas images and the target image being segmented. Image registration (using Elastix [25]) was done in two stages [28]: first, a similarity mapping to account for global translation, rotation, and scaling, followed by a B-spline mapping to account for local deformations. After spatially mapping carpal bone segmentations from every atlas image onto the target image, a majority vote was applied across all mappings, determining whether a voxel was labeled as background or as one of the carpal bones.

It should be noted that all atlas images contained the right wrist joint. For segmentation of the left wrist, atlas images were horizontally mirrored prior to registration. In order to avoid biased measurements, patients that were part of the ABS atlas were excluded from optimization and validation phases.

Assessment of segmentation accuracy

To assess the accuracy of ABS, a leave-one-out cross-validation was performed. In each of the 13 runs, 12 out of 13 atlas images would constitute the atlas set, and the remaining image would be used as the target image to be segmented. The result was validated against manual segmentation of the coronal T1-Gd image. Segmentation accuracy was evaluated by computing precision and recall rates for each carpal bone. Here, precision rate refers to the fraction of voxels segmented by ABS that overlap with the manual bone segmentation, while recall rate refers to the fraction of voxels within the manual bone segmentation that were correctly segmented by ABS.

BME quantification

BME is characterized by high signal intensity on T1-Gd images due to contrast enhancement and the suppressed normal fatty bone marrow. The precise intensity values vary per acquisition, depending on the strength of contrast enhancement and

fat suppression. The variation of these values is further broadened by inherent magnetic field inhomogeneities of the MR scanner. To account for these acquisition-specific intensity ranges of edematous vs. non-edematous bone marrow, fuzzy C-means clustering [29,30] was applied to the intensity values of all voxels in each image, assuming two clusters. This yields two probability map images (one per cluster) where each voxel contains the probability of that voxel belonging to the respective cluster. Let C_2 be the cluster whose center value is the higher of the two computed cluster centers. As Figure 2 illustrates, high probabilities (bright voxels) within the C_2 probability map correspond to locations of high fluid content, such as BME and synovium.

For each carpal bone, the fraction of bone affected by BME was estimated as the fraction of voxels (out of the total number of voxels within the bone's segmentation) whose probability of belonging to C_2 was higher than the threshold value T_{C_2} (numeric value optimized below). The resulting quantitative BME measurement (BME-QM) takes any fractional values between 0 and 1.

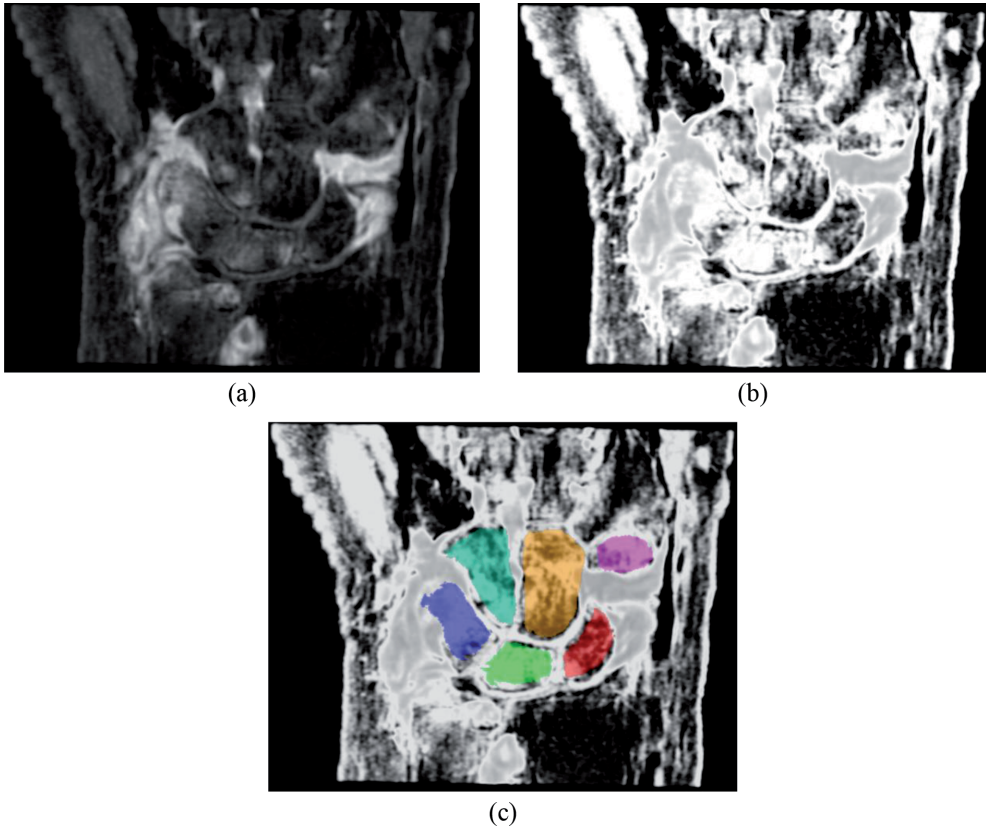
Optimization

In order to optimize the T_{C_2} threshold parameter based on correlation with visual BME scores, a training set of patients was defined. The number of patients with low-moderate BME in our cohort is much larger than the number of patients with severe BME. Therefore, random sampling of the cohort does not guarantee inclusion of patients with severe BME in the population sample. To ensure that patients with high degree of BME were represented in the training set, we categorized a set of 468 patients by the maximum visual BME score (R_{max}) across the carpal bones. Four sampling categories were defined corresponding to four intervals within R_{max} range. Table 1 lists the defined categories and the number of patients that fall into each category. Next, 15 patients were randomly selected from each category to form a training set of 60 patients.

Table 1. Training set sampling categories

Patient category index	R_{max} interval	Number of patients
0	$R_{max} = 0$	189
1	$0 < R_{max} \leq 1$	208
2	$1 < R_{max} \leq 2$	42
3	$2 < R_{max} \leq 3$	29

Note: Random sampling across all categories would form a training set that mainly consists of patients with $R_{max} \leq 1$. In contrast, randomly selecting 15 patients from category 3, for example, guarantees that the training set will include 15 patients in which at least one bone received a visual BME score greater than 2. Thus, random sampling from individual categories helps ensure T_{C2} is optimized with respect to the entire range of the visual BME score.

**Figure 2.** SRR image of the wrist (a), its C2 probability map image (b), and C2 image with carpal bone segmentation overlay from ABS (c).

To minimize the influence of acquisition artifacts and segmentation errors on threshold optimization, three patients whose MR scans suffered from incomplete fat suppression and one patient for which ABS failed were excluded from the obtained training set. This brought the final training set size to 56 patients. The optimal value of T_{C2} was found by maximizing the Pearson correlation coefficient r between the sum of visual BME scores across all carpal bones and the sum of BME-QM across all carpal bones.

Validation

After optimizing and locking the value of T_{C2} , the method was validated on 502 patients that were not part of the training set.

Statistical analysis

The Pearson correlation coefficient r between the sum of visual BME scores across all carpal bones and the sum of BME-QM across all carpal bones was evaluated. P -values below 0.05 were indicative of statistical significance. MR scans that suffered from incomplete fat suppression were noted and excluded from the correlation computation. Scans with other acquisition artifacts, such as noise patterns and incomplete field of view were excluded from the analysis. Patients in which one or more bones were not segmented by ABS yielded undefined values for BME-QM. Since undefined values cannot be included in the correlation computation, these patients were excluded from statistical analysis. The statistics were computed using MATLAB R2015b (MathWorks, Inc.).

Results

Assessment of segmentation accuracy

The mean bone-level recall and precision rates of ABS with respect to manual segmentations across 13 patients are shown in Figure 3. Recall rates were lowest in the pisiform (mean of 0.58 ± 0.09 SD) and highest in the capitate (mean of 0.82 ± 0.03 SD). Precision rates were high in all bones, with mean values ranging from 0.92 to 0.96 and SD values ranging from 0.02 to 0.05.

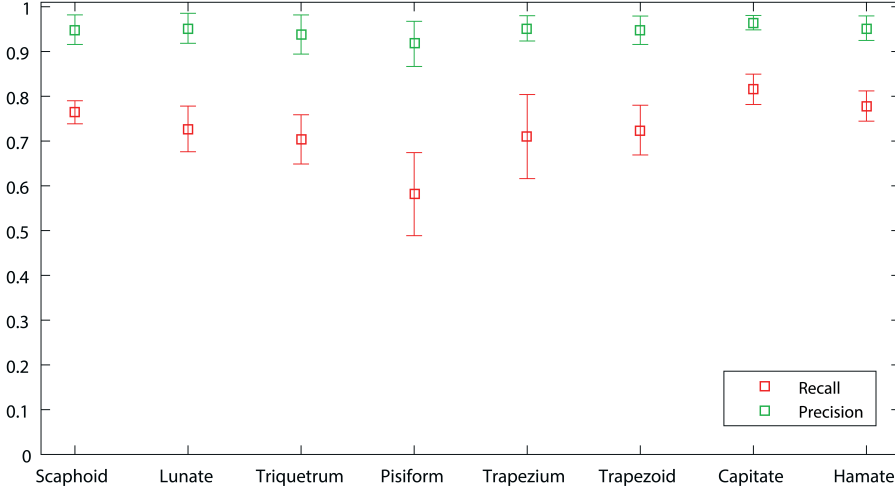


Figure 3. Mean (\pm SD) bone-level recall and precision rates of ABS with respect to manual segmentations across 13 patients.

Optimization

The maximum Pearson correlation ($r = 0.86$, $P < 0.001$), over 56 training set patients, between the sum of visual BME scores across all carpal bones and the sum of BME-QM across all carpal bones was achieved at threshold value $T_{C2} = 0.83$ (Figure 4). The scatter plot of the data is shown in Figure 5.

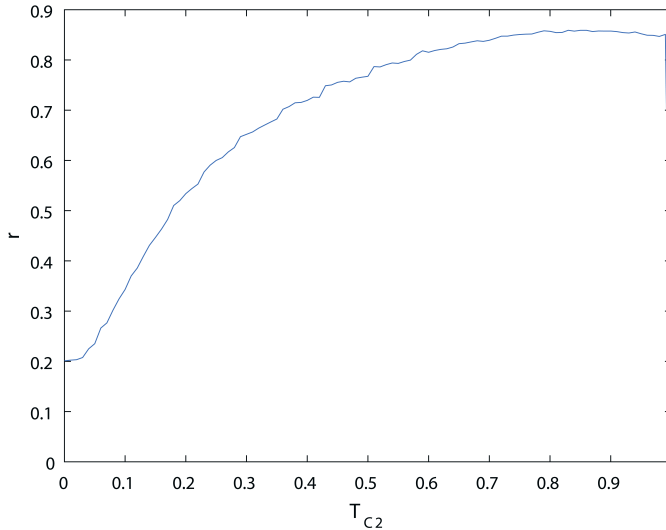


Figure 4. Pearson correlation coefficient r , over 56 training set patients, between the sum of visual BME scores across all carpal bones and the sum of BME-QM across all carpal bones, as a function of T_{C2} .

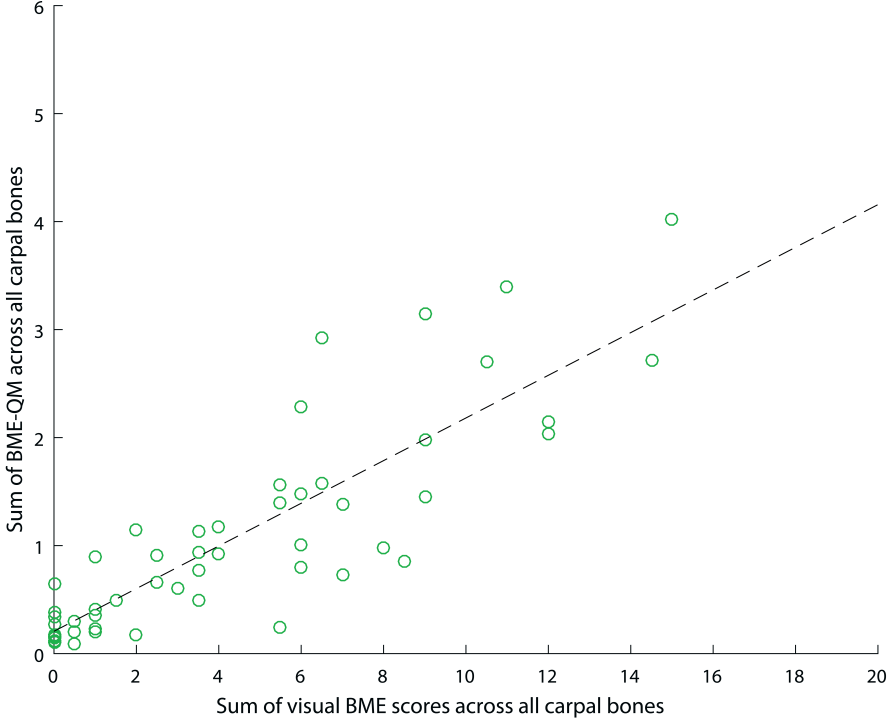


Figure 5. Scatter plot of sum of BME-QM across all carpal bones vs. sum of visual BME scores across all carpal bones for 56 training set patients. Each data point represents a single patient. $r = 0.86$, $P < 0.001$, $T_{C2} = 0.83$. Dashed black line represents linear regression fit.

Validation

Out of 502 patients, BME-QM was undefined in six patients due to failed segmentation. Three patients were excluded due to noise artifacts ($n = 2$) and incomplete field of view ($n = 1$) in their images. MR scans of eight patients suffered from incomplete fat suppression. For the remaining 485 patients, the Pearson correlation between the sum of visual BME scores across all carpal bones and the sum of BME-QM across all carpal bones was $r = 0.83$, $P < 0.001$. The scatter plot of the data is shown in Figure 6. Most patients formed clusters of steadily increasing BME-QM values, as the visual score value increased. Some outliers from this general trend were clearly visible for visual score value of 0 and BME-QM values between 1 and 2. These high quantitative values were due to

inaccurate segmentation of the carpal bones. Several patients whose images suffered from incomplete fat suppression produced BME-QM values that were largely deviating from the observed regression fit.

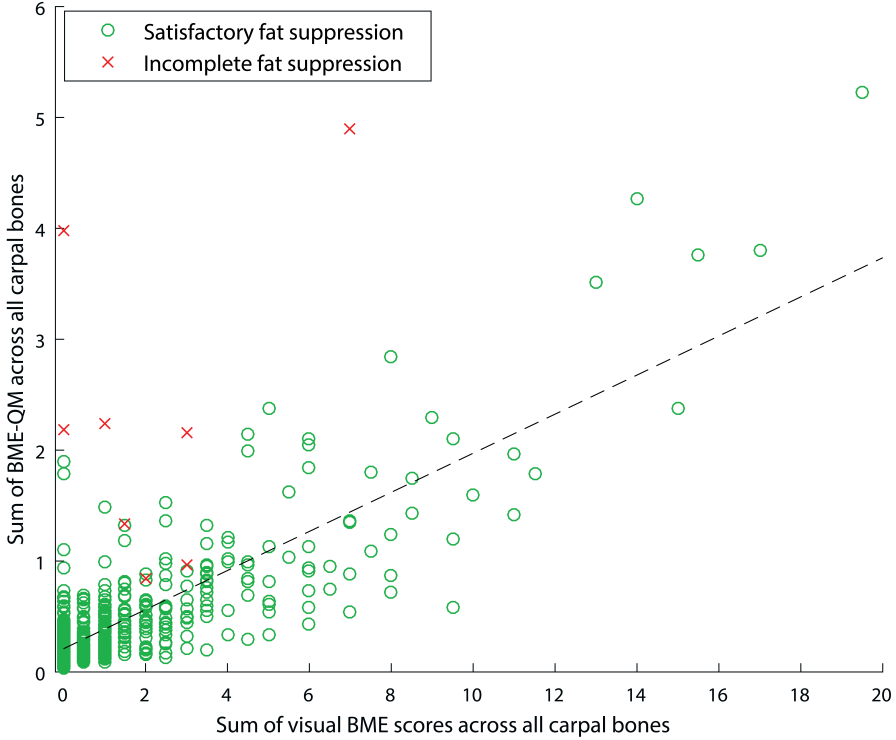


Figure 6. Scatter plot of sum of BME-QM across all carpal bones vs. sum of visual BME scores across all carpal bones for 493 validation set patients. Each data point represents a single patient. Linear regression fit (dashed black line) and Pearson correlation r were computed over 485 patients whose MR scans did not suffer from incomplete fat suppression (circular data points): $r = 0.83$, $P < 0.001$, $T_{c2} = 0.83$.

Discussion

In this study, we investigated the feasibility of automatic quantification of BME on MRI of the wrist in patients with early arthritis through an atlas-based approach. We chose to focus on the carpal bones, since they provide a complex multi-object scenario for exploring the feasibility of an atlas-based quantification framework. The advantage of this framework is that it can be straightforwardly expanded to other areas of the wrist and other joints by adding these areas of interest to the atlas. Validation results across 485 early arthritis patients indicated good

correlation between BME-QM and visual BME scores. It should be noted that perfect correlation is inherently not achievable because of the coarse grading scale of the visual score and the fine grading scale of BME-QM.

Our training strategy helped ensure that during validation BME-QM correlated well across the entire range of the visual BME score. The fact that the correlation curve in Figure 4 is relatively flat for T_{C2} values between 0.75 and 0.9 suggests that there is a range of T_{C2} values in this interval that result in good agreement between quantitative and visual scores. Furthermore, since BME-QM measures the fraction of voxels with C2 probability above T_{C2} , this seems to indicate that locations considered as BME in visual scoring often result in C2 probability values around 0.9. We also examined the effect of a smaller training set on T_{C2} optimization (data not shown), with five patients randomly selected from each R_{max} category forming a training set of 20 patients. We observed a similarly stable high correlation for T_{C2} values between 0.75 and 0.9, suggesting that the optimization step is not overly sensitive to training set size, as long as patients from all categories of BME severity (R_{max}) are represented in the training data.

The time required to execute the BME-QM framework for one patient on an Intel® Xeon® E5-1620 v3 CPU was ~58 min (SRR, ~20 min; ABS, ~35 min; BME quantification ~3 min). ABS is the most time-consuming step, but it can be accelerated 10-fold by running image registrations between all atlas images and the target image in parallel. Since registrations are independent of each other, this can be easily achieved given sufficient computing power. However, in large cohort studies, where evaluation of image data is often carried out days or weeks after the image is acquired, such acceleration may be irrelevant; an automatic framework can be executed immediately after image acquisition in an integrated fashion, thus ensuring quantitative results are available by the time a research project enters the evaluation phase.

ABS provided satisfactory segmentation for the vast majority of patients. In practice, failed segmentation cases will require manual adjustment by an expert in order for BME-QM to be computed. Over-segmentation of bones or shifted

segmentations that include synovium voxels increase the value of BME-QM due to contrast enhancement in the synovium. It is preferable to slightly under-segment the bone to ensure the exclusion of synovium while retaining most of the bone marrow within the segmentation. That said, significant under-segmentation may lead to an upward bias in BME-QM. Quantitative assessment of ABS accuracy in 13 patients revealed the tendency of ABS to under-segment bones (mid-range recall rates and high precision rates). Therefore, the current framework may raise false alarms when bone volume is under-estimated in the presence of moderate BME. The mid-range recall rates also suggest unwanted variability in BME-QM due to incomplete bone segmentation. The fact that the lowest recall rates were observed in the pisiform while the highest in the capitate, is likely due to the fact that the pisiform is the smallest of the carpal bones while the capitate is the largest of the carpal bones. An additional challenge during registration is the varied intensity and pattern of BME across patients. It is therefore advisable to avoid using very fine grid spacing during the B-spline registration step, since alignment between images on a coarser scale should be less sensitive to these local variations. Another potential pitfall is inclusion of erosions in the segmentation result. Erosions may contain high intensities that will mistakenly contribute to the value of BME-QM. To address these possible pitfalls and improve bone-level recall rates, an automatic refinement step should follow ABS in the future. In addition, to ensure robustness of the atlas to variations in MRI acquisition protocols and scanners at different sites, it may be necessary to form a larger atlas set consisting of sub-atlases of wrist scans acquired under different echo/repetition times and magnetic field strengths. The most suitable sub-atlas can then be automatically identified based on the acquisition parameters of a specific target image.

Incomplete fat suppression during acquisition of MR scans has an adverse effect on the accuracy of BME-QM. Bone marrow fat signal that is not properly suppressed results in high intensities that are mistaken for edema voxels by the clustering algorithm. Fat suppression quality requirements for BME-QM are higher compared to visual scoring. This is due to the availability of pre-contrast image

data in visual scoring and pattern recognition during visual assessment of increased signal intensity secondary to insufficient fat suppression. Although fat suppression issues are relatively rare, they must be identified prior to applying BME-QM to reduce false positives. The possibility of identifying and compensating fat suppression issues automatically should be investigated. In addition, more robust fat suppression techniques that are less sensitive to bulk susceptibility, such as Dixon techniques, may be beneficial when BME-QM is used.

A limitation of the current study is that the quality of carpal bone segmentation in training and validation set patients was judged subjectively. Quantitative assessment of segmentation accuracy was not possible, since no ground truth, manual segmentations were available for these patients. Quantification of segmentation accuracy would allow to supplement the BME measurement with a confidence measure. Another limitation is that pre-contrast image data could not be included in the framework, since pre-contrast T1 scans were acquired only in the coronal plane, while SRR requires a coronal and axial scan of the same sequence as input. Therefore, a straightforward voxel-to-voxel comparison between SRR T1-Gd images and pre-contrast T1 images was not possible. Inclusion of pre-contrast data would allow to explore a subtraction methodology as means of quantifying BME and could also facilitate the detection of fat suppression issues.

Recently, another framework aimed at automatically quantifying RA-related biomarkers, called quantitative RAMRIS (RAMRIQ), was proposed by Bowes *et al.* [31,32] and employed in a treatment effects study by Conaghan *et al.* [33]. These studies focus on measuring change over time, demonstrating higher sensitivity of quantitative measurements compared to RAMRIS. In contrast, we focused on validation of quantitative measurements at a single time point. In the future, it would be interesting to employ BME-QM for measuring change over time and evaluate its sensitivity.

Conclusion

We conclude that BME-QM has potential to provide a feasible alternative to visual scoring of BME on MRI of the wrist in patients with early arthritis. Complete automation requires further refinement of carpal bone segmentation and automatic detection and compensation of acquisition artifacts. Future work should also add more locations of interest relevant to RA to the atlas and extend this framework to other types of inflammation, such as synovitis and tenosynovitis. These developments can save time and manual effort for clinical researchers and help assess the value of MRI both for diagnosing RA and monitoring its treatment.

Acknowledgements

We would like to sincerely thank Dirk H. J. Poot for providing us with the implementation of his super-resolution algorithm and advice on applying the method to our image data. We also appreciate the input of Elize C. Newsum in visual scoring of the MR scans.

REFERENCES

1. McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, et al. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. *Arthritis Rheum.* 2003;48:1814–27.
2. Hetland ML, Ejbjerg B, Hørslev-Petersen K, Jacobsen S, Vestergaard A, Jurik AG, et al. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann. Rheum. Dis.* 2009;68:384–90.
3. Bøyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. MRI in early rheumatoid arthritis: synovitis and bone marrow oedema are independent predictors of subsequent radiographic progression. *Ann. Rheum. Dis.* 2011;70:428–33.
4. Nieuwenhuis WP, van Steenberghe HW, Stomp W, Stijnen T, Huizinga TWJ, Bloem JL, et al. The Course of Bone Marrow Edema in Early Undifferentiated Arthritis and Rheumatoid Arthritis: A Longitudinal Magnetic Resonance Imaging Study at Bone Level. *Arthritis Rheumatol.* (Hoboken, N.J.). 2016;68:1080–8.
5. Østergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. *J. Rheumatol.* 2003;30:1385–6.
6. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. *Ann. Rheum. Dis.* 2005;64:i3-7.
7. Diamond AL. Foveal simultaneous brightness contrast as a function of inducing, and test-field luminances. *J. Exp. Psychol.* 1953;45:304–14.
8. Heinemann EG. Simultaneous brightness induction as a function of inducing- and test-field luminances. *J. Exp. Psychol.* 1955;50:89–96.
9. Leibowitz H, Mote FA, Thurlow WR. Simultaneous contrast as a function of separation between test and inducing fields. *J. Exp. Psychol.* 1953;46:453–6.
10. Li X, Yu A, Virayavanich W, Noworolski SM, Link TM, Imboden J. Quantitative characterization of bone marrow edema pattern in rheumatoid arthritis using 3 Tesla MRI. *J. Magn. Reson. Imaging.* 2012;35:211–7.
11. Teruel JR, Burghardt AJ, Rivoire J, Srikhum W, Noworolski SM, Link TM, et al. Bone

structure and perfusion quantification of bone marrow edema pattern in the wrist of patients with rheumatoid arthritis: a multimodality study. *J. Rheumatol.* 2014;41:1766–73.

12. Yang H, Rivoire J, Hoppe M, Srikhun W, Imboden J, Link TM, et al. Computer-aided and manual quantifications of MRI synovitis, bone marrow edema-like lesions, erosion and cartilage loss in rheumatoid arthritis of the wrist. *Skeletal Radiol.* 2015;44:539–47.

13. Li X, Ma BC, Bolbos RI, Stahl R, Lozano J, Zuo J, et al. Quantitative assessment of bone marrow edema-like lesion and overlying cartilage in knees with osteoarthritis and anterior cruciate ligament tear using MR imaging and spectroscopic imaging at 3 Tesla. *J. Magn. Reson. Imaging.* 2008;28:453–61.

14. Dodin P, Abram F, Pelletier J-P, Martel-Pelletier J. A fully automated system for quantification of knee bone marrow lesions using MRI and the osteoarthritis initiative cohort. *J. Biomed. Graph. Comput.* 2012;3:51.

15. Dodin P, Martel-Pelletier J, Pelletier J-P, Abram F. A fully automated human knee 3D MRI bone segmentation using the ray casting technique. *Med. Biol. Eng. Comput.* 2011;49:1413–24.

16. de Rooy DPC, van der Linden MPM, Knevel R, Huizinga TWJ, van der Helm-van Mil AHM. Predicting arthritis outcomes--what can be learned from the Leiden Early Arthritis Clinic? *Rheumatology (Oxford).* 2011;50:93–100.

17. Stomp W, Krabben A, van der Heijde D, Huizinga TWJ, Bloem JL, van der Helm-van Mil AHM, et al. Aiming for a shorter rheumatoid arthritis MRI protocol: can contrast-enhanced MRI replace T2 for the detection of bone marrow oedema? *Eur. Radiol.* 2014;24:2614–22.

18. Sudoł-Szopińska I, Jurik AG, Eshed I, Lennart J, Grainger A, Østergaard M, et al. Recommendations of the ESSR Arthritis Subcommittee for the Use of Magnetic Resonance Imaging in Musculoskeletal Rheumatic Diseases. *Semin. Musculoskelet. Radiol.* 2015;19:396–411.

19. Frenzel T, Apte C, Jost G, Schöckel L, Lohrke J, Pietsch H. Quantification and Assessment of the Chemical Form of Residual Gadolinium in the Brain After Repeated Administration of Gadolinium-Based Contrast Agents. *Invest. Radiol.* 2017;1–9.

20. Radbruch A, Haase R, Kickingeder P, Bäumer P, Bickelhaupt S, Paech D, et al. Pediatric Brain: No Increased Signal Intensity in the Dentate Nucleus on Unenhanced T1-weighted MR Images after Consecutive Exposure to a Macrocyclic Gadolinium-based Contrast Agent. *Radiology. Radiological Society of North America;* 2017;1–9.

21. Greenspan H, Oz G, Kiryati N, Peled S. MRI inter-slice reconstruction using super-resolution. *Magn. Reson. Imaging.* 2002;20:437–46.
22. Shilling RZ, Robbie TQ, Bailloleul T, Mewes K, Mersereau RM, Brummer ME. A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI. *IEEE Trans. Med. Imaging.* 2009;28:633–44.
23. Woo J, Murano EZ, Stone M, Prince JL. Reconstruction of high-resolution tongue volumes from MRI. *IEEE Trans. Biomed. Eng.* 2012;59:3511–24.
24. Poot DHJ, Van Meir V, Sijbers J. General and efficient super-resolution method for multi-slice MRI. *Med. Image Comput. Comput. Assist. Interv.* 2010;13:615–22.
25. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging.* 2010;29:196–205.
26. Aizenberg E. Elastix parameters for alignment between coronal and axial images of the wrist [Internet]. 2016. Available from: <http://elastix.bigr.nl/wiki/index.php/Par0040>
27. Rohlfing T, Brandt R, Menzel R, Russakoff DB, Maurer CRJ. Quo vadis, atlas-based segmentation? *Handb. Med. Image Anal. - Vol. III Regist. Model.* 2005. p. 435–486.
28. Aizenberg E. Elastix parameters for atlas-based segmentation of carpal bones [Internet]. 2016. Available from: <http://elastix.bigr.nl/wiki/index.php/Par0041>
29. Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Kluwer Academic Publishers; 1981;
30. Amiri M. Yashil's Fuzzy C-Means Clustering MATLAB Toolbox Ver. 1.0 [Internet]. 2003. Available from: http://ce.sharif.edu/~m_amiri/project/yfcmc/index.htm
31. Bowes MA, Guillard G, Gill E, Vincent GR, Hensor E, Freeston JE, et al. Novel Quantification of MRI Provides a More Sensitive Outcome Measure Than RAMRIS. 2014 ACR/ARHP Annu. Meet. 2014.
32. Bowes MA, Guillard G, Vincent GR, Freeston JE, Vital EM, Emery P, et al. Quantitative MRI Measurement of Tenosynovitis Demonstrates Differing Responses of Synovitis and Tenosynovitis after RA Treatment. 2015 ACR/ARHP Annu. Meet. 2015.
33. Conaghan PG, Østergaard M, Bowes MA, Wu C, Fuerst T, van der Heijde D, et al. Comparing the effects of tofacitinib, methotrexate and the combination, on bone marrow oedema, synovitis and bone erosion in methotrexate-naïve, early active rheumatoid arthritis: results of an exploratory randomised MRI study incorporating semiquantitati. *Ann. Rheum. Dis.* 2016;

4

Automatic quantification of tenosynovitis on MRI of the wrist in patients with early arthritis

This chapter was adapted from:

E. Aizenberg*, D.P. Shamonin*, M. Reijnierse, A.H.M. van der Helm–van Mil, B.C. Stoel, “Automatic quantification of tenosynovitis on MRI of the wrist in patients with early arthritis: a feasibility study,” *European Radiology*, doi: 10.1007/s00330-018-5807-2, 2018.

* Authors contributed equally.

Abstract

Purpose: Tenosynovitis (inflammation of the synovial lining of the sheath surrounding tendons) is frequently observed on MRI of early arthritis patients. Since visual assessment of tenosynovitis is a laborious task, we investigated the feasibility of automatic quantification of tenosynovitis on MRI of the wrist in a large cohort of early arthritis patients.

Methods: For 563 consecutive early arthritis patients (clinically confirmed arthritis ≥ 1 joint, symptoms < 2 years), MR scans of the wrist were processed in three automatic stages. First, super-resolution reconstruction was applied to fuse coronal and axial scans into a single high-resolution three-dimensional image. Next, 10 extensor/flexor tendon regions were segmented using atlas-based segmentation and marker-based watershed. A measurement region of interest (ROI) was defined around the tendons. Finally, tenosynovitis was quantified by identifying image intensity values associated with tenosynovial inflammation using fuzzy clustering and measuring the fraction of voxels with these characteristic intensities within the measurement ROI. A subset of 60 patients was used for training and the remaining 503 patients for validation. Correlation between quantitative measurements and visual scores was assessed through Pearson correlation coefficient.

Results: Pearson correlation between quantitative measurements and visual scores across 503 patients was $r = 0.90$, $P < 0.001$. False detections due to blood vessels and synovitis present within the measurement ROI contributed to a median offset from zero equivalent to 13.8% of the largest measurement value.

Conclusion: Quantitative measurement of tenosynovitis on MRI of the wrist is feasible and largely consistent with visual scores. Further improvements in segmentation and exclusion of false detections are warranted.

Introduction

Initiation of treatment in the early stages of rheumatoid arthritis (RA) has been associated with higher chances of drug-free sustained remission and improved quality of life [1]. Therefore, it is important to recognize patients who are at risk of progressing to RA as early as possible, either in the symptomatic phase of arthralgia, which precedes clinical arthritis, or in the earliest phases of clinically detectable arthritis. Recent studies suggest that MRI-detected inflammation can aid this task [2–4], especially in combination with serological markers [2]. Among the different types of inflammation observed on MRI of hands and wrists, it has been shown that tenosynovitis (inflammation of the synovial lining of the sheath surrounding tendons) is independently predictive of RA development, both in patients presenting with early arthritis and with arthralgia [2–5]. In addition, changes in MRI-detected tenosynovitis may be of interest in treatment response evaluation.

Assessment of tenosynovitis on MRI is commonly done according to the scoring method of Haavardsholm *et al.* [6], in which a reader examines multiple tendon regions and estimates the thickness of peritendinous effusion or synovial proliferation with contrast enhancement. This is a laborious task, which requires the availability of trained, experienced readers. Automating the evaluation of tenosynovitis could offer standardized, high-precision measurements derived directly from the image data and alleviate the time burden and cost associated with visual scoring. To date, limited research is available on this topic. Bowes *et al.* have published a conference abstract on quantifying change in tenosynovitis over time in 34 RA patients receiving treatment [7], but data on single time point validation of these quantitative measurements with respect to visual scores are not publicly available.

In a recent study, we developed an automatic framework for measuring bone marrow edema (a strong predictor of radiographic progression in RA patients [8]) on MR images of the wrist [9]. In the work presented here, we sought to extend that framework to measure tenosynovitis of the extensor and flexor tendons of the

wrist. Our aim was to investigate the feasibility of tenosynovitis quantification and assess the correlation between quantitative measurements and visual scores in a large cohort of early arthritis patients.

Methods

Patients

A total of 563 early arthritis patients consecutively included in the Leiden Early Arthritis Clinic cohort [10] were studied. Mean age (\pm SD) was 54.9 (\pm 15.4) years, 350 patients (62.2%) were female. Inclusion required clinically confirmed arthritis by physical examination in ≥ 1 joints and symptom duration < 2 years. The cohort study was approved by the medical ethics committee of Leiden University Medical Center (Leiden, The Netherlands). All participants provided written informed consent.

MRI scanning and visual scoring

The wrist joint of the most painful side (or the dominant side in cases of equally severe symptoms on both sides) was scanned with a 1.5T extremity MR scanner (GE Healthcare, Waukesha, WI, USA) using a 100 mm coil, with contrast enhancement and frequency-selective fat saturation (T1-Gd). Table 1 summarizes the acquisition parameters. In line with the definitions proposed by Haavardsholm *et al.* [6], tenosynovitis was evaluated in six extensor compartments and four flexor regions within the wrist joint (Figure 1). Visual scoring was independently performed by two trained readers blinded to clinical data. For each anatomical region, the readers provided a grade on a 0–3 scale based on the estimated maximum width of peritendinous effusion or synovial proliferation with contrast enhancement, as follows: grade 0, normal; grade 1, < 2 mm; grade 2, ≥ 2 mm and < 5 mm; grade 3, ≥ 5 mm. The scoring region was bounded by the distal radius/ulna proximally and the hook of the hamate distally. The intra-reader intra-class correlation coefficients (ICCs) of the two readers for the total tenosynovitis score (sum across all tendon regions), based on 40 MRIs scored twice, were 0.99

and 0.83. The inter-reader ICC for the total tenosynovitis score, based on all 563 MRIs, was 0.87. In what follows, the mean score of the two readers was always considered.

Table 1. MRI sequences

	Coronal scan	Axial scan
Repetition time (ms)	650	570
Echo time (ms)	17	7
Acquisition matrix	364×224	320×192
Echo train length	2	2
Slice thickness (mm)	2	3
Slice gap (mm)	0.2	0.3

Described are the acquisition parameters of T1-weighted fast spin-echo sequences with frequency-selective fat saturation obtained after intravenous injection of Gd-chelate (gadoteric acid, Guerbet, Paris, France, standard dose of 0.1 mmol/kg).

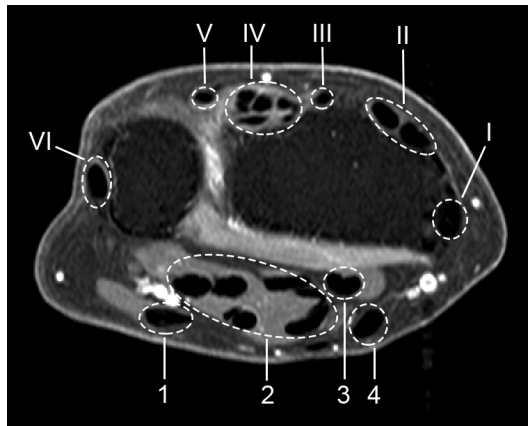


Figure 1. Tendon regions (compartments) scored for tenosynovitis, shown on axial MR image of the wrist (T1, post-gadolinium, fat-saturated). The six defined extensor compartments contain: abductor pollicis longus, extensor pollicis brevis (I); extensor carpi radialis longus, extensor carpi radialis brevis (II); extensor pollicis longus (III); extensor digitorum communis, extensor indicis proprius (IV); extensor digiti quinti proprius (V); extensor carpi ulnaris (VI). The four flexor regions contain: flexor carpi ulnaris (1); ulnar bursa, including flexor digitorum profundus and superficialis tendon quartets (2); flexor pollicis longus (tendon) in radial bursa (3); flexor carpi radialis (4). Note: the flexor carpi ulnaris does not have a tenosynovial sheath; nevertheless, inflammation around this tendon is also observed, and therefore enhancement of tissue surrounding this tendon is scored [11].

Quantitative image analysis framework

Super-resolution reconstruction

The coronal and axial MR scans compensate each other in terms of anatomical detail, since the slice thickness in each of the scans (2 mm in coronal; 3 mm in axial) is much larger than the in-plane spacing between voxels (~ 0.2 mm). In order for a quantitative framework to make use of all available image data in a compact and efficient manner, it is desirable to fuse the two scans into a single 3D image using super-resolution reconstruction (SRR). The application of SRR to MR images of the wrist has been detailed in our previous work [9]. We applied the method of Poot *et al.* [12] with Laplacian regularization ($\lambda = 0.05$).

Measurement region of interest

The computation of the ROI required automatically segmenting the tendons, carpal bones, distal radius/ulna, and the image region bounded by skin. The bones and initial landmarks for the tendon regions were obtained using atlas-based segmentation [13]. The atlas consisted of 13 early arthritis patients (separate dataset, excluding patients evaluated visually and quantitatively in this study). For each atlas patient, the tendon regions and bones were manually segmented in the axial T1-Gd images and then extended to SRR space by nearest neighbor interpolation. After spatially mapping every atlas image onto the target image using the Elastix toolbox [14–16], a majority vote was applied across all mappings, determining whether a voxel would be labeled as one of the tendons, bones, or neither. It should be noted that all atlas images contained the right wrist joint. For segmentation of the left wrist, atlas images were horizontally mirrored prior to registration.

Having obtained initial landmarks for the tendon regions, the tendons were segmented by a similar approach to Chen *et al.* [17] using marker-based watershed segmentation [18–20], followed by removal of segmented regions whose intensity was > 75 (tendons are characterized by low image intensities on T1-Gd images) or

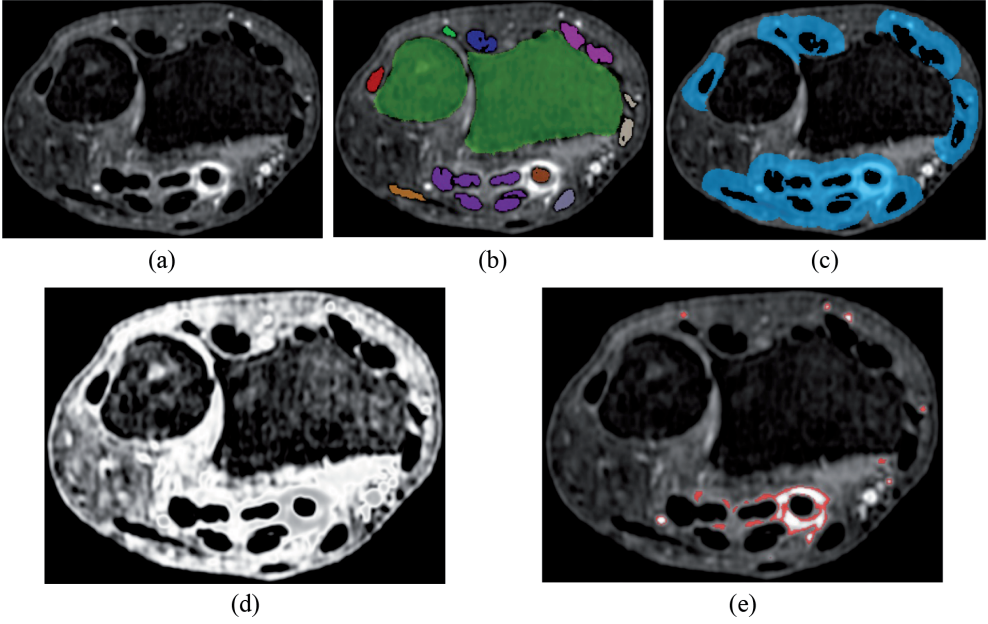


Figure 2. SRR image of the wrist (a), segmented tendon regions and bones (b), the resulting measurement ROI (c) in which tenosynovitis is quantified ($D = 3$ mm), C2 probability map (d), ROI locations included in the quantitative measurement marked in red (e). Image depicted in the figure received a total visual score of 1: grade 1 tenosynovitis in flexor region 3. Some of the voxels identified in neighboring flexor region 2 corresponded to a low grade enhancement, but were not picked up in visual scoring. Several blood vessels located within the ROI were also included, introducing a number of false detections that counted towards the quantitative measurement.

whose volume was < 0.01 ml. An example of the resulting segmentations is shown in Figure 2(b).

In order to segment the image region bounded by skin, the entire image extent of the hand was approximated. First, the background was segmented by performing region growing with seeds placed at the four corners of each image slice. Then, the resulting binary image was inverted and the largest connected component was retained.

Finally, for each segmented tendon, a distance transform was performed and voxels within a fixed distance (D) of the tendons were included in the measurement ROI as long as these voxels were not part of other labeled structures. The distal radius/ulna boundary of the ROI was determined by identifying the axial slice where the two bones were closest to each other. The hook of the hamate boundary

was determined by searching for the axial slice with the largest number of segmented hamate voxels. An example of the resulting measurement ROI is shown in Figure 2(c). As detailed in the optimization section, the value of distance parameter D was obtained by maximizing correlation with visual scores on a training set of patients.

Assessment of tendon segmentation accuracy

To assess the accuracy of tendon segmentation, a leave-one-out cross-validation was performed. In each of the 13 runs, 12 out of 13 atlas images would constitute the atlas set, and the remaining image would be used as the target image to be segmented. The result was validated against manual segmentation of the axial image. Segmentation accuracy was evaluated by computing precision and recall rates for each of the 10 tendon regions. Here, precision rate refers to the fraction of voxels segmented by the algorithm that overlap with the manual segmentation, while recall rate refers to the fraction of voxels within the manual segmentation that were correctly segmented by the algorithm.

Tenosynovitis quantification

Tenosynovitis is characterized by high signal intensity on T1-Gd (fat-suppressed) images due to contrast enhancement. Intensity values vary per acquisition, depending on the relative strength of contrast enhancement, the homogeneity of the fat suppression, and the inherent magnetic field inhomogeneities of the MR scanner. To account for these acquisition-specific intensity ranges of tenosynovitis, fuzzy C-means clustering [21,22] was applied to the intensity values of all voxels in each image, assuming two clusters. This yields two probability map images, where each voxel contains the probability of that location belonging to the respective cluster. Let C2 be the cluster whose center value is the higher of the two computed cluster centers. As Figure 2(d) illustrates, high probabilities (bright voxels) within the C2 probability map correspond to locations of healthy synovial tissue. Since our focus is on regions of inflammation, where image intensity is expected to be higher compared to healthy synovium, voxels whose intensity was

lower than the value of C2 cluster center were removed, resulting in a one-sided C2 probability map.

Tenosynovitis was then quantified by computing the fraction of voxels within the measurement ROI whose one-sided C2 probability values p_{C2} were bounded by $T_L \leq p_{C2} < T_H$. As detailed below, the numeric values of the lower and upper thresholds (T_L, T_H) were optimized on a training set of patients to maximize correlation with visual scores.

Optimization

In order to optimize the (T_L, T_H) thresholds and distance parameter D based on correlation with visual scores, a training set of patients was defined. The number of patients with low tenosynovitis (grades 0 and 1) in our early arthritis cohort was much larger than the number of patients with moderate-severe tenosynovitis (grades 2 and 3). Therefore, a random sampling of the cohort would not guarantee inclusion of patients with severe tenosynovitis in the training sample. In order to produce a more balanced training set representing the full range of tenosynovitis severity, we used a similar sampling approach as in our study on bone marrow edema [9]. We categorized 563 patients by the maximum visual score (V_{max}) across the scored tendon regions. Three sampling categories were defined corresponding to three severity intervals within V_{max} range (0–3): $V_{max} = 0$; $0 < V_{max} \leq 1$; $1 < V_{max} \leq 3$. Table 2 lists the defined categories and the number of patients that fall into each category. Next, 20 patients were randomly selected from each category to form a training set of 60 patients. The optimal distance and threshold values were found by computing the quantitative measurement for $D = 1, 2, 3, 4, 5, 6$ mm and all possible combinations (step size 0.01) of (T_L, T_H) and determining which set of parameters maximized the Pearson correlation coefficient r between the total visual score of tenosynovitis (sum across all tendon regions) and the total quantitative tenosynovitis measurement.

Table 2. Training set sampling categories

Severity category index	V_{max} interval	Number of patients
0	$V_{max} = 0$	200
1	$0 < V_{max} \leq 1$	261
2	$1 < V_{max} \leq 3$	102

Note: Random sampling across all categories would form a training set that mainly consists of patients with $V_{max} \leq 1$. In contrast, randomly selecting 20 patients from category 2, for example, guarantees that the training set will include 20 patients in which at least one tendon region received a visual score greater than 1. Thus, random sampling from each severity category helps ensure D , T_L , and T_H are optimized with respect to the entire range of tenosynovitis severity.

Validation

After optimizing and locking the values of D , T_L , and T_H , the method was validated by computing the quantitative tenosynovitis measurement for the 503 patients that were not part of the training set and evaluating the Pearson correlation coefficient between the total visual score and the total quantitative measurement.

Statistical analysis

When assessing the Pearson correlation coefficient between visual scores and quantitative measurements, P -values below 0.05 were considered to be statistically significant. Statistics were computed using MATLAB R2015b (The MathWorks Inc, Natick, MA, USA).

Results

Assessment of tendon segmentation accuracy

The median and interquartile range (IQR) of recall and precision rates of tendon region segmentation across 13 atlas images are shown in Figure 3. Flexor regions exhibited high precision rates (median values ranging from 0.92 to 0.97) and moderate-high recall rates (median values ranging from 0.85 to 0.90). The rates were generally lower for extensor regions and exhibited more variability (median precision ranging from 0.78 to 0.94 and median recall ranging from 0.32 to 0.85).

The lowest recall (including 3 failed segmentations) was observed for extensor region III.

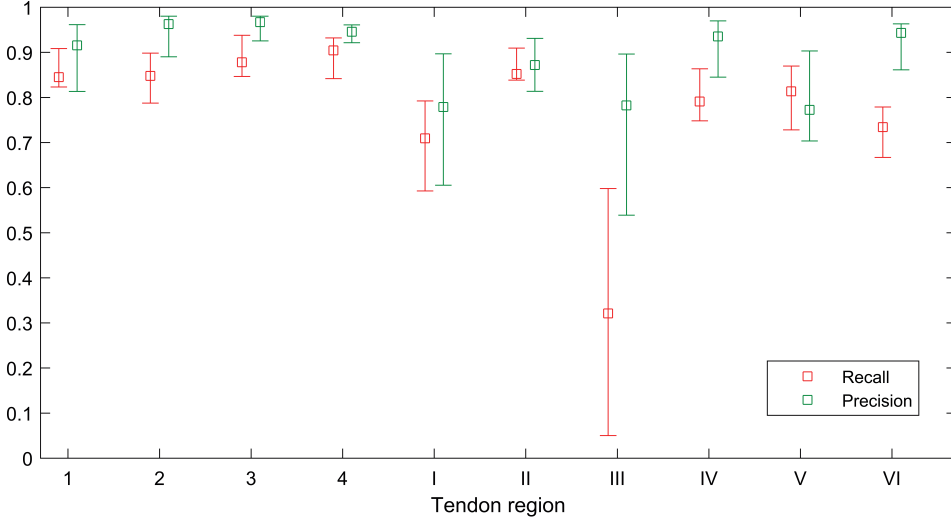


Figure 3. Median and interquartile range of recall and precision rates of tendon region segmentation across 13 atlas images.

Optimization

The highest Pearson correlation value ($r = 0.93$, $P < 0.001$) between the total visual score of tenosynovitis and the total quantitative measurement over 60 training set patients was observed with distance parameter $D = 3$ mm and threshold values $T_L = 0.82$, $T_H = 0.94$. As illustrated by the scatter plot in Figure 4, increasing levels of tenosynovitis severity were fairly consistently matched with increasing values of the quantitative measurement. The measurements of patients with total visual score 0 had a median offset from zero of 0.04 (IQR 0.03–0.05), constituting 14.8% of the largest observed value of 0.27 for the most severely affected patients. Figure 2(e) shows an example of measurement ROI locations that were counted towards the quantitative measurement.

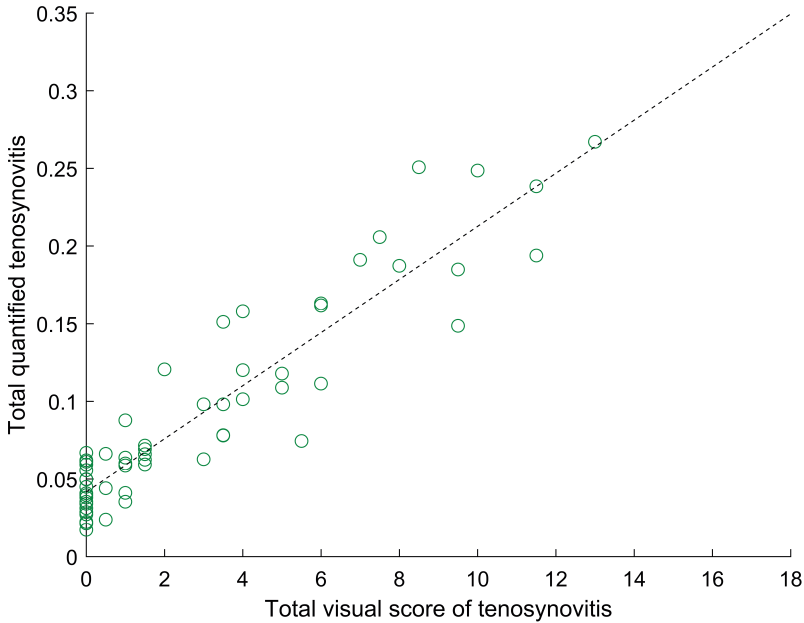


Figure 4. Scatter plot of total quantitative measurements of tenosynovitis versus total visual scores for 60 training set patients. Each data point represents a single patient. Pearson correlation $r = 0.93$, $P < 0.001$, ($D = 3$ mm, $T_L = 0.82$, $T_H = 0.94$). Dashed black line represents linear regression fit.

Validation

Having obtained the optimized parameter values, the quantitative measurement was computed for 503 patients, and correlation was assessed. The resulting Pearson correlation coefficient was $r = 0.90$, $P < 0.001$. The scatter plot in Figure 5 shows that majority of patients exhibited a consistent trend of increasing quantitative measurements with increasing levels of tenosynovitis severity. The measurements of patients with total visual score 0 had a median offset from zero of 0.04 (IQR 0.03–0.05) (same as in training), constituting 13.8% of the largest observed value of 0.29. Visual inspection of results indicated that blood vessels and synovitis present within the measurement ROI were often mistakenly counted as tenosynovitis by the quantitative measurement, increasing its numeric value. The strongly outlying case of a patient with visual score 0 and a quantitative

measurement of 0.15 was caused by a failed tendon segmentation due to an unusually low intensity distribution of healthy synovium.

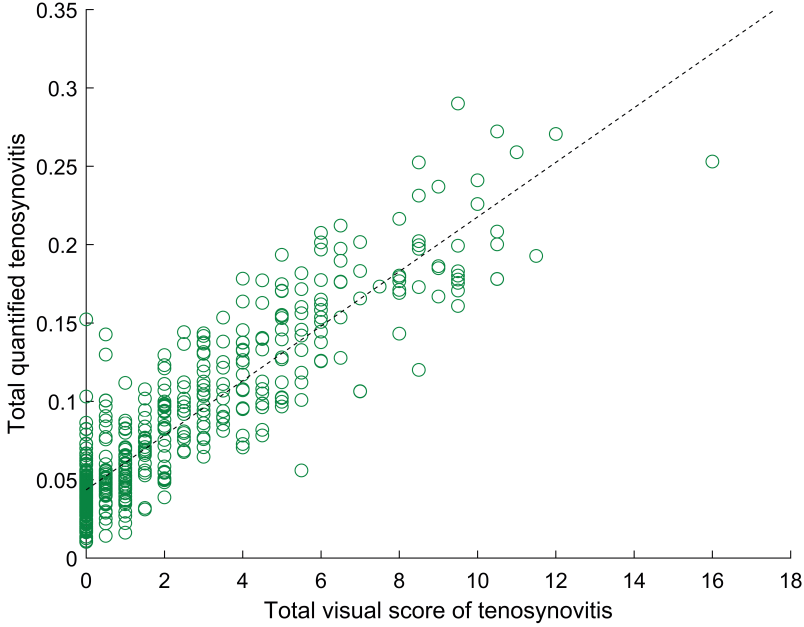


Figure 5. Scatter plot of total quantitative measurements of tenosynovitis versus total visual scores for 503 validation set patients. Each data point represents a single patient. Pearson correlation $r = 0.90$, $P < 0.001$, ($D = 3$ mm, $T_L = 0.82$, $T_H = 0.94$). Dashed black line represents linear regression fit.

Discussion

In this study, we investigated the feasibility of automatic quantification of tenosynovitis on MRI of the wrist in a large cohort of early arthritis patients. The presented method extended our previously developed atlas-based framework [9] to the extensor and flexor tendons of the wrist, providing the landmarks necessary for tendon segmentation and definition of the ROI in which tenosynovitis was measured. The results exhibited strong correlation between quantitative measurements and visual scores. Quantitative measurements should not be viewed as a replication of visual scoring and therefore this study assessed consistency and correlation, rather than absolute agreement. The observed correlation is especially encouraging considering that there is an inherent degree of variability within visual

scores due to the interval-based definition of the visual grades. These findings indicate that automatic quantification of tenosynovitis on MRI of early arthritis patients is feasible, and that quantitative measurements are largely consistent with visual scoring. However, this study also brings out multiple challenges pertinent to the quantification task, such as moderate segmentation performance and sources of false detections. As detailed in the following discussion, these are important issues that will need to be addressed on the path to a robust quantification framework.

Interestingly, the overall moderate tendon segmentation recall rates did not seem to have a strong adverse effect on correlation between quantitative measurements and visual scores. This can be explained by the fact that even if a tendon is partially segmented, the measurement ROI around the segmentation is still likely to include the tendon's synovial lining. Although the ROI will then also include voxels inside the tendon, on T1-Gd images tendons are characterized by low image intensities which do not contribute towards the inflammation measurement; one exception is enhancement due to concomitant tendinitis. It should be recognized, however, that in this study we measured the total inflammation across all evaluated tendon regions, which may have reduced sensitivity to errors made on the individual region level. This is particularly relevant when considering the low recall rates for extensor region III. The 3/13 failed cross-validation cases indicate that reliable quantification of inflammation around this tendon was not always feasible. A likely reason for this is that extensor region III is the smallest of the 10 tendon regions and exhibits higher curvature, making the placement of atlas-based landmarks more challenging. One type of segmentation error that had no effect on total measurements was mislabeling of one tendon region as another; however, future studies must thoroughly assess mislabeling errors if evaluation of tenosynovitis on individual region level is of interest. More generally, it should be noted that inaccuracies in tendon segmentation do affect the total number of voxels included in the measurement ROI and thereby introduce some variability in the quantitative measurement. Therefore, improving segmentation accuracy is an important direction of future

work both for measurement precision and accurate evaluation of tenosynovitis on the individual region level.

As illustrated by Figure 2(e), locations counted towards the quantitative measurement did not always include all voxels within the inflammation, but most voxels along the boundary of the inflammation were typically included. One possible reason for this is that the threshold parameters (T_L, T_H) were optimized with respect to scores that reflect the maximum thickness of peritendinous effusion or synovial proliferation in each tendon region. Maximum thickness is not equivalent to total volume, and therefore it is plausible that some voxels within the inflammation were not included in the measurement. Figure 2(e) also illustrates that one drawback of the current method is that blood vessels introduce false detections that contribute towards the quantitative measurement. This observation explains one of the factors behind the consistent offset from zero both during training and validation. Future improvements should include detection of blood vessels and their exclusion from the measurement ROI.

Visual inspection of quantification results indicated that synovitis present within the measurement ROI (for example, between carpal bones and tendons) was mistakenly counted as tenosynovitis by the quantitative measurement. In visual scoring, trained readers employ their expertise and pattern recognition to classify the observed inflammation as either synovitis or tenosynovitis. The presented method did not include such classification, and therefore it is not surprising that it counted all inflammation detected within the measurement ROI as tenosynovitis. This is another contributing factor to the offset observed in training and validation. Since synovitis is often present in joints in the vicinity of tendons affected by tenosynovitis [11], a more specific definition of the measurement ROI is warranted.

Conclusion

In conclusion, the presented method provides a reference on the path to automatic quantification of tenosynovitis on MRI and lays out possible directions for future

improvements. The common presence of tenosynovitis in RA and its association with RA development in arthralgia and early arthritis patients motivate the development of quantitative measurement techniques. These advances would aid clinical researchers by standardizing interpretation and allowing them to dedicate more resources to analysis rather than visual scoring, facilitating both research and potential clinical implementation.

Acknowledgements

We would like to sincerely thank Dirk H. J. Poot for providing us with the implementation of his super-resolution algorithm.

REFERENCES

1. Ajeganova S, Huizinga T. Sustained remission in rheumatoid arthritis: latest evidence and clinical considerations. *Ther. Adv. Musculoskelet. Dis.* SAGE Publications; 2017;9:249–62.
2. van Steenbergen HW, Mangnus L, Reijnders M, Huizinga TWJ, van der Helm-van Mil AHM. Clinical factors, anticitrullinated peptide antibodies and MRI-detected subclinical inflammation in relation to progression from clinically suspect arthralgia to arthritis. *Ann. Rheum. Dis.* 2016;75:1824–30.
3. Kleyer A, Krieter M, Oliveira I, Faustini F, Simon D, Kaemmerer N, et al. High prevalence of tenosynovial inflammation before onset of rheumatoid arthritis and its link to progression to RA-A combined MRI/CT study. *Semin. Arthritis Rheum.* Elsevier; 2016;46:143–50.
4. Nieuwenhuis WP, van Steenbergen HW, Mangnus L, Newsum EC, Bloem JL, Huizinga TWJ, et al. Evaluation of the diagnostic accuracy of hand and foot MRI for early Rheumatoid Arthritis. *Rheumatology.* Oxford University Press; 2017;56:1367–77.
5. Aizenberg E, ten Brinck RM, Reijnders M, van der Helm-van Mil AHM, Stoel BC. Identifying MRI-detected inflammatory features specific for rheumatoid arthritis: two-fold feature reduction maintains predictive accuracy in clinically suspect arthralgia patients. *Semin. Arthritis Rheum.* 2018;
6. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Kvien TK. Introduction of a novel magnetic resonance imaging tenosynovitis score for rheumatoid arthritis: reliability in a multireader longitudinal study. *Ann. Rheum. Dis.* 2007;66:1216–20.
7. Bowes MA, Guillard G, Vincent GR, Freeston JE, Vital EM, Emery P, et al. Quantitative MRI Measurement of Tenosynovitis Demonstrates Differing Responses of Synovitis and Tenosynovitis after RA Treatment. 2015 ACR/ARHP Annu. Meet. 2015.
8. Hetland ML, Ejbjerg B, Hørslev-Petersen K, Jacobsen S, Vestergaard A, Jurik AG, et al. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann. Rheum. Dis.* 2009;68:384–90.
9. Aizenberg E, Roex EAH, Nieuwenhuis WP, Mangnus L, van der Helm-van Mil AHM, Reijnders M, et al. Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: A feasibility study. *Magn. Reson. Med.* 2018;79:1127–34.
10. de Rooy DPC, van der Linden MPM, Knevel R, Huizinga TWJ, van der Helm-van Mil

AHM. Predicting arthritis outcomes--what can be learned from the Leiden Early Arthritis Clinic? *Rheumatology* (Oxford). 2011;50:93–100.

11. Nieuwenhuis WP, Krabben A, Stomp W, Huizinga TWJ, van der Heijde D, Bloem JL, et al. Evaluation of magnetic resonance imaging-detected tenosynovitis in the hand and wrist in early arthritis. *Arthritis Rheumatol.* (Hoboken, N.J.). 2015;67:869–76.

12. Poot DHJ, Van Meir V, Sijbers J. General and efficient super-resolution method for multi-slice MRI. *Med. Image Comput. Comput. Assist. Interv.* 2010;13:615–22.

13. Rohlfing T, Brandt R, Menzel R, Russakoff DB, Maurer CRJ. Quo vadis, atlas-based segmentation? *Handb. Med. Image Anal. - Vol. III Regist. Model.* 2005. p. 435–486.

14. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging.* 2010;29:196–205.

15. Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M, et al. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front. Neuroinform.* 2013;7:50.

16. Aizenberg E, Shamonin DP. Elastix parameters for atlas-based segmentation of bones and tendons on MR images of the wrist [Internet]. 2018. Available from:

<http://elastix.bigr.nl/wiki/index.php/Par0051>

17. Chen H-C, Wang Y-Y, Lin C-H, Wang C-K, Jou I-M, Su F-C, et al. A Knowledge-Based Approach for Carpal Tunnel Segmentation from Magnetic Resonance Images. *J. Digit. Imaging.* Springer-Verlag; 2013;26:510–20.

18. Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* 1991;13:583–98.

19. Vincent L. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Trans. Image Process.* 1993;2:176–201.

20. Xu S, Liu H, Song E. Marker-Controlled Watershed for Lesion Segmentation in Mammograms. *J. Digit. Imaging.* Springer-Verlag; 2011;24:754–63.

21. Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Kluwer Academic Publishers; 1981;

22. Amiri M. Yashil's Fuzzy C-Means Clustering MATLAB Toolbox Ver. 1.0 [Internet]. 2003. Available from: http://ce.sharif.edu/~m_amiri/project/yfcmc/index.htm

5

Identifying MRI-detected inflammatory features specific for rheumatoid arthritis: two-fold feature reduction maintains predictive accuracy in clinically suspect arthralgia patients

This chapter was adapted from:

E. Aizenberg^{*}, R.M. ten Brinck^{*}, M. Reijnerse, A.H.M. van der Helm–van Mil, B.C. Stoel, “Identifying MRI-detected inflammatory features specific for rheumatoid arthritis: two-fold feature reduction maintains predictive accuracy in clinically suspect arthralgia patients,” *Seminars in Arthritis and Rheumatism*, doi: 10.1016/j.semarthrit.2018.04.005, 2018.

^{*} Authors contributed equally.

Abstract

Purpose: MRI-detected inflammation is considered of diagnostic value for rheumatoid arthritis (RA), but its evaluation involves a time-consuming scoring of 61 joint-level features. It is not clear, however, which of these features are specific for RA and whether evaluating a subset of specific features is sufficient to differentiate RA patients. This study aimed to identify a subset of RA-specific features in a case-control setting and validate them in a longitudinal cohort of arthralgia patients.

Methods: The difference in frequency of MRI-detected inflammation (bone marrow edema, synovitis, tenosynovitis) between 199 RA patients and 193 controls was studied in 61 features across the wrist, metacarpophalangeal, and metatarsophalangeal joints. A subset of RA-specific features was obtained by applying a cutoff on the frequency difference while maximizing discriminative performance. For validation, this subset was used to predict arthritis development in 225 clinically suspect arthralgia (CSA) patients. Diagnostic performance was compared to a reference method that uses the complete set of 61 features normalized for inflammation levels in age-matched controls.

Results: Subset of 30 features, mainly (teno)synovitis, was obtained from the case-control setting. Validation in CSA patients yielded an area of 0.69 (95% CI: 0.59–0.78) under the ROC curve and a positive predictive value (PPV) of 31%, compared to 0.68 (95% CI: 0.60–0.77) and 29% PPV of the reference method with 61 features.

Conclusion: Subset of 30 MRI-detected inflammatory features, dominated by (teno)synovitis, offers a considerable reduction of scoring efforts without compromising accuracy for prediction of arthritis development in CSA patients.

Introduction

MRI-detected inflammation has been shown to predict erosive progression in early rheumatoid arthritis (RA) [1] and contribute to prediction of arthritis development in patients presenting with clinically suspect arthralgia (CSA) [2,3]. However, evaluating MR scans for bone marrow edema (BME), synovitis, and tenosynovitis across the wrist, metacarpophalangeal (MCP), and metatarsophalangeal (MTP) joints commonly amounts to a time-consuming scoring of 61 joint-level features in line with the RA MRI scoring system (RAMRIS) [4]. Yet, it is not clear which of these features are specific for RA and whether evaluating a subset of specific joint-level features would provide a similar or improved diagnostic performance when predicting progression from CSA to RA. Recent studies by Van Steenbergen *et al.* [3], Kleyer *et al.* [5], and Mangnus *et al.* [6] suggest that while certain anatomical locations and types of inflammation exhibit stronger association with arthritis development, others are also prevalent among symptom-free persons.

Identification of RA-specific features could both simplify the use of MRI in practice and advance the understanding of arthritis pathogenesis. Patients with CSA are a population of special interest in this context. CSA is a symptomatic phase preceding clinical arthritis, and therefore, it provides opportunity to clinically recognize patients who are at risk of progression to RA. The study of Van Steenbergen *et al.* [3] in 150 CSA patients found that 20% of these patients developed clinically detectable arthritis within two years of being recognized as having CSA by the treating rheumatologist. Furthermore, identifying patients at risk of progression to RA in the pre-arthritis phase would allow to study whether earlier treatment can increase chances of improved outcome [7].

Considering on the one hand the high sensitivity of MRI in measuring local inflammation in patients who already progressed from CSA to clinical arthritis [8], but on the other hand the presence of some inflammatory features in symptom-free persons [6], we suspect that comparing the frequency of inflammation across established RA patients and symptom-free persons may help identify features that would be most predictive of progression from CSA to clinical arthritis. This study

aimed to 1) determine the difference in frequency of joint-level inflammation between RA patients (cases) and symptom-free persons (controls), 2) identify a subset of features that, on the one hand, are specific for RA based on the difference in case-control frequency of inflammation, and on the other hand maximize discriminative ability compared to the complete set of features, and 3) validate the identified subset of features for prediction of progression from CSA to clinical arthritis within a 2-year follow-up period in a longitudinal cohort of CSA patients.

Methods

Subjects

Three groups of individuals from previously reported cohorts were studied, as detailed below: patients with established RA, symptom-free persons, and patients with CSA. All cohort studies were approved by the medical ethics committee of Leiden University Medical Center (Leiden, The Netherlands). All participants provided written informed consent.

Cases: rheumatoid arthritis patients from the Leiden Early Arthritis Clinic cohort

The Leiden Early Arthritis Clinic (EAC) cohort [9] is a longitudinal inception cohort that includes patients with arthritis clinically confirmed by physical examination and symptom duration of less than two years. The cohort was initiated in 1993 at Leiden University Medical Center (Leiden, The Netherlands). Baseline MRI was added to the study protocol in August 2010. Consecutive patients that presented with RA meeting the 1987 American College of Rheumatology (ACR) criteria [10] at 1-year follow-up, between August 2010 and October 2014, were studied ($n = 199$) and are subsequently referred to as cases.

Controls: symptom-free volunteers

Symptom-free volunteers from a previously reported study [6] served as controls ($n = 193$). Volunteers were recruited via advertisements in local newspapers and websites and had no history of inflammatory rheumatic diseases, no

musculoskeletal symptoms during the month preceding the study, and no evidence of arthritis at physical examination.

Clinically suspect arthralgia patients from the Leiden CSA cohort

The CSA cohort [11] is a population-based inception cohort that started in 2012 at Leiden University Medical Center (Leiden, The Netherlands) with the aim of studying the symptomatic phase of RA that precedes clinical arthritis. Inclusion required the presence of arthralgia of the small joints for less than a year that was at increased risk of progressing to RA according to the patient's rheumatologist's clinical expertise. General practitioners in our region rarely perform autoantibody testing before referral [12]; hence, rheumatologists included patients based on the clinical presentation [11]. This approach to identifying CSA was proven accurate in clinical practice [13], but contains a certain degree of subjectivity. To harmonize inclusion in future studies, an EULAR taskforce recently developed a definition of arthralgia suspicious for progression to rheumatoid arthritis [2]. This definition is based on 7 parameters: symptom duration < 1 year, symptoms located in MCP joints, morning stiffness duration ≥ 60 min, most severe symptoms in early morning, presence of first-degree relative with RA, difficulty with making a fist, and positive squeeze test of MCP joints. The EULAR taskforce did not provide a single recommended cutoff point for the number of positive parameters that define CSA, but it was noted that a high sensitivity (> 90%) with respect to patients identified as CSA was obtained if ≥ 3 out of 7 parameters were present.

Following admission to the cohort, patients' baseline assessment included the Health Assessment Questionnaire (HAQ), a 66-swollen joint count (SJC) and 68-tender joint count (TJC), blood samples (including C-reactive protein (CRP), IgM rheumatoid factor (RF), ACPA (anti-cyclic citrullinated peptide 2, Eurodiagnostica, The Netherlands)), and acquisition of MRI. Treatment with disease-modifying anti-rheumatic drugs (DMARDs) was not allowed. Non-steroidal anti-inflammatory drugs (NSAIDs) were allowed, but stopped 24 hours

prior to MRI in order to prevent the suppression of subclinical inflammation at the moment of MR imaging.

Patients included between April 2012 and March 2015 with available baseline MRI data were studied ($n = 225$). Among these patients, 162 (72%) exhibited presence of ≥ 3 of the CSA parameters defined by EULAR [2]. Follow-up ended when clinical arthritis had developed or else after two years. Positive outcome was defined as arthritis development within two years of baseline MRI, identified at joint examination by an experienced rheumatologist. Out of the 225 studied patients, 41 (18.2%) patients progressed to clinical arthritis within the 2-year follow-up period.

MRI scanning and scoring

For all patients in the three cohorts, contrast-enhanced MRI was performed in the wrist, MCP(2–5), and MTP(1–5) joints of the most painful side (or the dominant side in case of equally severe symptoms on both sides). The joints were scanned with a 1.5T extremity MR scanner (GE Healthcare, Waukesha, WI, USA) using a 100 mm coil for the hand and a 145 mm coil for the foot. In the hand, a T1-weighted fast spin-echo (FSE) sequence was acquired before contrast injection in the coronal plane (repetition time (TR) of 575 ms, echo time (TE) of 11.2 ms, acquisition matrix 388×288, echo train length (ETL) 2). After intravenous injection of gadolinium contrast (gadoteric acid, Guerbet, Paris, France, standard dose of 0.1 mmol/kg), a T1-weighted FSE sequence with frequency-selective fat saturation (T1-Gd) was acquired in the coronal plane (TR/TE 700/9.7 ms, acquisition matrix 364×224, ETL 2) and the axial plane (wrist: TR/TE 540/7.7 ms, acquisition matrix 320×192, ETL 2; MCP joints: TR/TE 570/7.7 ms, acquisition matrix 320×192, ETL 2). The obtained sequences for the forefoot were a T1-Gd sequence in the axial plane (TR/TE 700/9.5 ms, acquisition matrix 364×224, ETL 2) and the coronal plane (perpendicular to the axis of the metatarsals) (TR/TE 540/7.5 ms, acquisition matrix 320×192, ETL 2). Coronal sequences of the hand had 18 slices with a slice thickness of 2 mm and a slice gap of 0.2 mm. Coronal sequences of the

foot had 20 slices with a slice thickness of 3 mm and a slice gap of 0.3 mm. All axial sequences had a slice thickness of 3 mm and a slice gap of 0.3 mm with 20 slices for the wrist, 16 for the MCP joints, and 14 for the foot. Further information about the MRI protocol and some exceptions are described in the Supplementary Material.

Bone marrow edema (BME) and synovitis were scored in line with the definitions proposed by the RAMRIS method [4]. The BME score was based on the fraction of affected bone volume: 0, no BME; 1, 1–33% of bone edematous; 2, 34–66%; 3, 67–100%. Histopathology studies of lesions defined as BME by RAMRIS have shown that these lesions contain lymphocytic infiltrates; therefore, the imaging feature BME in RA has been also called osteitis [14–16]. The synovitis score was based on the volume of enhanced tissue in the synovial compartment: 0, none; 1, mild; 2, moderate; 3, severe. Since the carpometacarpal (CMC)-1 joint (base metacarpal-1 and trapezium) does not communicate with the intercarpal joint, and it is a prediction site for arthrosis, it was excluded.

Tenosynovitis in the wrist and MCP joints was scored in line with Haavardsholm *et al.* [17]. The score was based on the estimated maximum width of peritendinous effusion or synovial proliferation with contrast enhancement: 0, normal; 1, < 2 mm; 2, \geq 2 mm and < 5 mm; 3, \geq 5 mm.

In total, 61 features were evaluated: 31 bones for BME (distal radius, distal ulna, 7 carpal bones, base metacarpal(2–5), proximal/distal MCP(2–5), proximal/distal MTP(1–5)), 12 joints for synovitis (intercarpal, radiocarpal, distal radioulnar, MCP(2–5), MTP(1–5) joints), and 18 tendon regions (compartments) (Figure 1) for tenosynovitis (6 extensor compartments and 4 flexor regions in the wrist, 4 flexor and 4 extensor tendons at the MCP level). Note that although extensor tendons at the MCP level and the flexor carpi ulnaris at the wrist do not have a tenosynovial sheath, inflammation around these tendons is also observed [18], and therefore enhancement of tissue surrounding these tendons is scored.

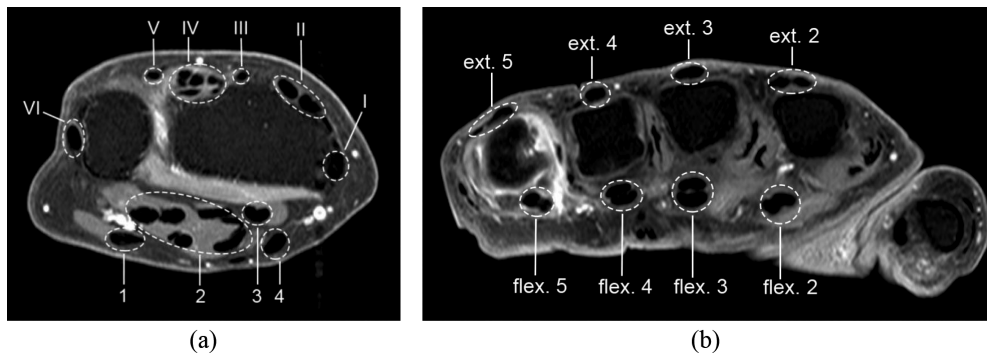


Figure 1. Tendon regions (compartments) scored for tenosynovitis in the wrist (a) and the MCP joints (b), shown on axial MR images (T1, post-gadolinium, fat-saturated). In the wrist, the six defined extensor compartments contain: abductor pollicis longus, extensor pollicis brevis (I); extensor carpi radialis longus, extensor carpi radialis brevis (II); extensor pollicis longus (III); extensor digitorum communis, extensor indicis proprius (IV); extensor digiti quinti proprius (V); extensor carpi ulnaris (VI). The four flexor regions in the wrist contain: flexor carpi ulnaris (1); ulnar bursa, including flexor digitorum profundus and superficialis tendon quartets (2); flexor pollicis longus (tendon) in radial bursa (3); flexor carpi radialis (4). In the MCP joints, the four extensor regions (ext. 2–5) contain the extensor tendons of the fingers, and the four flexor regions (flex. 2–5) contain the paired flexor tendons, corresponding to MCP joints 2–5. Note: extensor tendons at the MCP level and the flexor carpi ulnaris at the wrist do not have a tenosynovial sheath; nevertheless, inflammation around these tendons is also observed, and therefore enhancement of tissue surrounding these tendons is scored [18].

Scoring was performed by a total of four independent experienced readers (two per each cohort) blinded to clinical data. The readers were physicians and active as researchers in the field of rheumatology, more specifically RA research. They received training for several months under the supervision of an experienced reader in order to learn the RAMRIS scoring system and had to achieve intraclass correlation coefficients (ICC) of 0.90 or higher on an MRI training set before they were allowed to score for research purposes. The inter-reader and intra-reader ICCs are reported in Supplementary Table A.1. In what follows, the mean score across readers was always considered.

Difference in joint-level frequency of inflammation between cases and controls

For each of the 61 inflammatory features, the frequency of presence of MRI-detected inflammation was computed separately across cases and controls. Presence of MRI inflammation in a given feature was defined as a visual score greater than 0. Next, the feature-wise frequency values obtained for controls were subtracted from the frequency values obtained for cases. The resulting values are referred to as control-adjusted frequency of inflammation. High values of control-adjusted frequency would reveal features that are specific for RA, while low values would indicate features that are either non-specific or have low prevalence of inflammation in RA patients.

Feature identification and prediction of outcome in the case-control setting

With the knowledge of difference in frequency of MRI-detected inflammation between cases and controls for each of the 61 features, we sought to identify a subset of features that would be specific for RA and, at the same time, would maximize discriminative ability compared to the complete set of features. Subsets of features of different specificity can be explored by varying a cutoff (threshold) value on the control-adjusted frequency and retaining features whose control-adjusted frequency is above that cutoff value. Lower cutoff values would produce larger subsets containing more non-specific features, while higher cutoff values would produce smaller subsets with more specific features. The discriminative ability of each such subset can be assessed by computing the total inflammation score across the retained features for every patient and measuring the area under the receiver operating characteristic curve (ROC). Consequently, our goal was to find the smallest subset that would yield an area under the curve (AUC) that was closest (or higher) to the AUC yielded by the complete set of 61 features. Thus, our feature identification method consisted of the following four stages:

- 1) Vary the value of control-adjusted frequency cutoff with a step size of 0.05; for each cutoff value, form a subset of features whose control-adjusted frequency of inflammation is above the cutoff value.
- 2) For every case and control subject, compute the total inflammation score across the obtained subset of features. Here, raw scores are considered, representing the severity of inflammation for each feature.
- 3) Assign positive outcome (RA) if the total inflammation score is greater than the value of a total inflammation threshold T_{infl} . Construct an ROC curve by varying the value of T_{infl} and compute the area under the curve.
- 4) Determine the smallest subset of features that yielded an AUC that is closest (or higher) to the AUC of the complete set of 61 features.

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were computed for the obtained subset of features at the ROC point closest to (0,1).

Validation in CSA patients based on the subset of features obtained from the case-control setting

The ultimate stage of the study was to validate the subset of inflammatory features obtained from the case-control setting for prediction of arthritis development in CSA patients. An underlying assumption made here is that features yielding good predictive performance on the case-control population would also yield good diagnostic performance on the CSA population, where positive outcome was defined as progression from CSA to clinical arthritis within two years of baseline MRI.

For every patient, the total inflammation score across the identified subset of features was computed. Once again, here the raw scores are considered, representing the severity of inflammation for each feature. Positive outcome was assigned if the total inflammation score was greater than the value of a total inflammation threshold T_{infl} . An ROC curve was constructed by varying the value

of T_{infl} . Diagnostic performance was quantified by AUC, sensitivity, specificity, PPV, and NPV. The latter four measures were computed for the T_{infl} value obtained from the case-control setting. For comparison of diagnostic performance, the method of Van Steenberg *et al.* [3] was applied to the same data and its AUC, sensitivity, specificity, PPV, and NPV were computed. In brief, the method assigns positive outcome if the inflammation score (i.e. severity of inflammation) of at least one of the 61 features was observed in less than 5% of age-matched controls. Since readers are blinded to patient age when evaluating the MR scans, all 61 features must be scored, so that outcome can be assigned after de-blinding of age and referencing with respect to age-matched controls.

Finally, recognizing that the total inflammation threshold obtained from the case-control setting might be too high for CSA patients, since inflammation levels are generally less severe in early disease patients, the test characteristics were computed again for the point on the CSA ROC curve that was closest to (0,1). This was performed as a sub-analysis to further explore the ROC curve produced by the identified subset of features. It should be clearly pointed out that this sub-analysis was subject to overfitting, because in this case the total inflammation threshold was optimized using validation data.

Results

Clinical characteristics

Baseline characteristics of studied subjects from the three cohorts are shown in Table 1. Among cases (127 females and 72 males), the mean age (\pm SD) was 56.1 (\pm 14.4) years, and among controls (136 females and 57 males) the mean age (\pm SD) was 49.8 (\pm 15.8) years. The mean age (\pm SD) of CSA patients (174 females and 51 males) was 44.2 (\pm 13.0) years. At baseline, 28 out of 225 (12.4%) CSA patients tested positive for anti-citrullinated peptide antibodies (ACPA). Within the 2-year follow-up period, 41 (18.2%) CSA patients progressed to clinical arthritis. Among these 41 patients, 17 patients (41.5%) were ACPA-positive at baseline.

Table 1. Baseline characteristics of subjects in the three cohorts

	RA patients (<i>n</i> = 199)	Symptom-free persons (<i>n</i> = 193)	CSA patients (<i>n</i> = 225)
Age in years, mean (SD)	56.1 (14.4)	49.8 (15.8)	44.2 (13.0)
Female, <i>n</i> (%)	127 (63.8)	136 (70.5)	174 (77.3)
BMI in kg/m ² *, mean (SD)	26.6 (4.3)	24.8 (3.9)	27.0 (4.9)
Elevated CRP, <i>n</i> (%)	129 (64.8)	<i>Not assessed</i>	49 (21.8)
HAQ score *, median (IQR)	1.0 (0.63–1.50)	<i>Not assessed</i>	0.50 (0.25–0.88)
IgM-RF positive, <i>n</i> (%)	121 (60.8)	<i>Not assessed</i>	46 (20.4)
ACPA positive, <i>n</i> (%)	108 (54.3)	<i>Not assessed</i>	28 (12.4)
TJC *, median (IQR)	5 (4–7)	0	6 (3–10)
SJC, median (IQR)	6 (3–10)	0	0

Legend:

ACPA = anti-citrullinated peptide antibody; BMI = body mass index; CRP = C-reactive protein; CSA = clinically suspect arthralgia; HAQ = Health Assessment Questionnaire; IgM-RF = immunoglobulin M rheumatoid factor; IQR = interquartile range; RA = rheumatoid arthritis; SD = standard deviation; SJC = swollen joints count; TJC = tender joint count.

* Missing data were as follows: BMI in CSA calculated for 224 patients, TJC in CSA calculated for 222 patients, HAQ in the RA patients calculated for 187 patients, TJC in the RA patients calculated for 192 patients, SJC in the RA patients calculated for 192 patients.

Difference in joint-level frequency of inflammation between cases and controls

The feature-wise frequency of MRI-detected inflammation in cases and controls is shown in Figure 2. After subtraction of control frequencies, several notable deviations from case values were observed. In particular, the frequency value reduced significantly for BME in the lunate and synovitis in the distal radioulnar, radiocarpal, and intercarpal joints. In contrast, control-adjusted frequency values remained close to case values and simultaneously high on the absolute scale for tenosynovitis in wrist flexor regions 2–4.

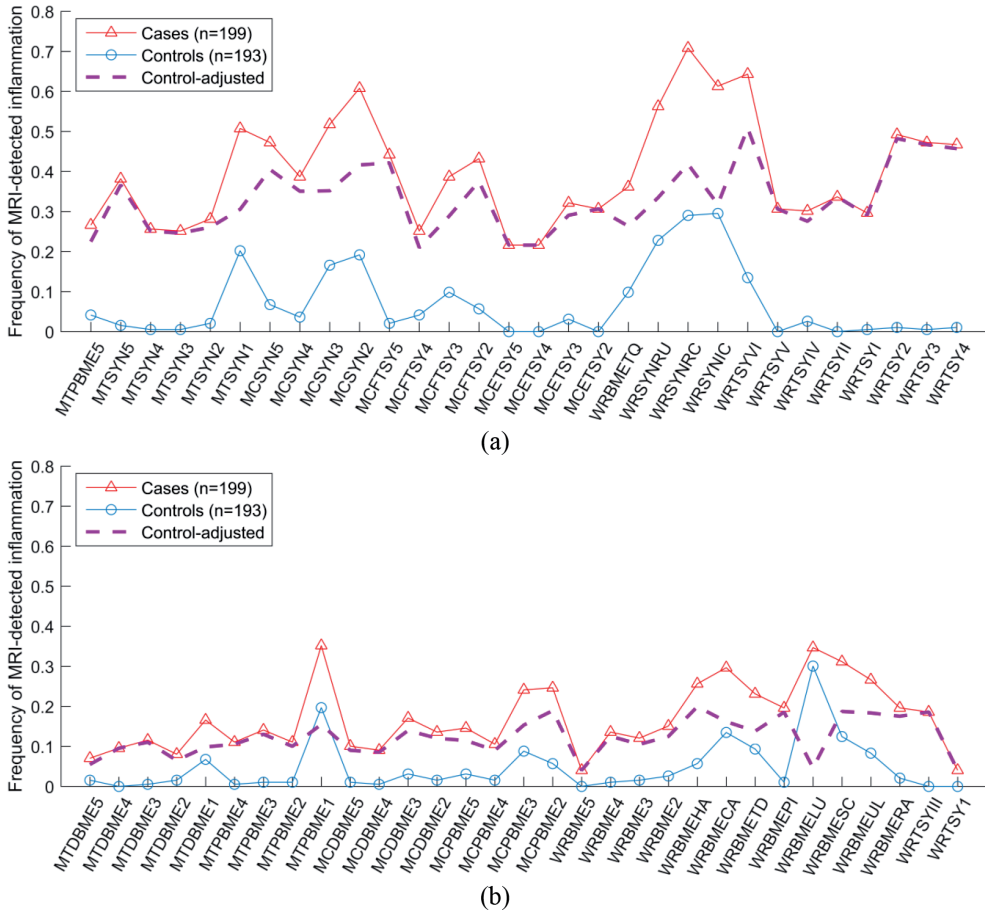


Figure 2. Frequency of MRI-detected inflammation across cases and controls in 61 inflammatory features, shown separately for the identified subset of 30 features (a) and 31 features that were not part of the identified subset (b). Control-adjusted frequency computed as feature-wise difference between cases and controls.

Feature abbreviations:

BME = bone marrow edema
SYN = synovitis
TSY = tenosynovitis

MT = metatarsal
 MTD = MT distal
 MTP = MT proximal
 MC = metacarpal
 MCD = MC distal
 MCP = MC proximal
 MCF = MC flexor
 MCE = MC extensor
 WR = wrist
 WR*(I-VI) = WR extensor
 compartments I-VI
 WR*(1-4) = WR flexor regions
 1-4

HA = hamate
CA = capitate
TD = trapezoid
PI = pisiform
TQ = triquetrum
LU = lunate
SC = scaphoid
UL = ulna
RA = radius
RU = distal radioulnar joint

RC = radiocarpal joint
IC = intercarpal joints

Feature identification and prediction of outcome in the case-control setting

Figure 3(a–b) displays the AUC for prediction of outcome (RA) in the case-control setting under different feature subsets, produced by varying the cutoff value on control-adjusted frequency with a step size of 0.05. The complete set of 61 features (cutoff value = 0) yielded an AUC of 0.91 (95% confidence interval (CI): 0.89 to 0.94). The smallest subset of features that yielded a similar (and higher) AUC of 0.93 (95% CI: 0.90 to 0.95), which was effectively comparable to that of the complete set, was observed at cutoff value 0.2 and consisted of 30 features (listed in Table 2). Most identified features were locations of tenosynovitis and synovitis, in addition to two BME locations (MTP5 and the triquetrum). Among features that were left out, 29/31 (94%) were locations of BME and 2/31 were tenosynovitis of wrist flexor region 1 and extensor compartment III. The total inflammation threshold corresponding to the ROC point closest to (0,1) was $T_{\text{Infl}} = 4.5$, with a sensitivity of 79%, specificity of 92%, PPV of 91%, and NPV of 81%.

Validation in CSA patients based on the subset of features obtained from the case-control setting

Applying the subset of 30 features obtained from the case-control setting to prediction of arthritis development in CSA patients yielded an AUC of 0.69 (95% CI: 0.59 to 0.78). The ROC curve is shown in Figure 3(c) together with the diagnostic test characteristics plotted as a function of the total inflammation threshold T_{Infl} in Figure 3(d). The threshold value derived from the case-control setting ($T_{\text{Infl}} = 4.5$) produced a sensitivity of 37%, specificity of 82%, PPV of 31%, and NPV of 85%. The method of Van Steenberg *et al.* [3] (61 features with age-referencing) was applied to the same data, yielding an AUC of 0.68 (95% CI: 0.60 to 0.77), sensitivity of 80%, specificity of 56%, PPV of 29%, and NPV of 93%. The diagnostic test characteristics of both methods are summarized in Table 3.

As a sub-analysis, to further explore the ROC curve produced by the subset of 30 features, the test characteristics were computed again for the point on the CSA ROC curve that was closest to (0,1), which corresponded to $T_{\text{Infl}} = 2.5$. This

Table 2. Subset of 30 inflammatory features obtained from the case-control setting

Feature	Location	Control-adjusted frequency
BME	MTP5	0.22
	Triquetrum	0.26
Synovitis	MTP5	0.37
	MTP4	0.25
	MTP3	0.25
	MTP2	0.26
	MTP1	0.31
	MCP5	0.41
	MCP4	0.35
	MCP3	0.35
	MCP2	0.42
	Distal radioulnar joint	0.33
	Radiocarpal joint	0.42
	Inter-carpal joints	0.32
Tenosynovitis	MCP5 flexor	0.42
	MCP4 flexor	0.21
	MCP3 flexor	0.29
	MCP2 flexor	0.38
	MCP5 extensor	0.22
	MCP4 extensor	0.22
	MCP3 extensor	0.29
	MCP2 extensor	0.31
	Wrist extensor compartment:	
	VI	0.51
	V	0.31
	IV	0.28
	II	0.34
	I	0.29
	Wrist flexor region:	
	2	0.48
	3	0.47
	4	0.46
Control-adjusted frequency computed as feature-wise difference in frequency of MRI-detected inflammation between cases and controls.		

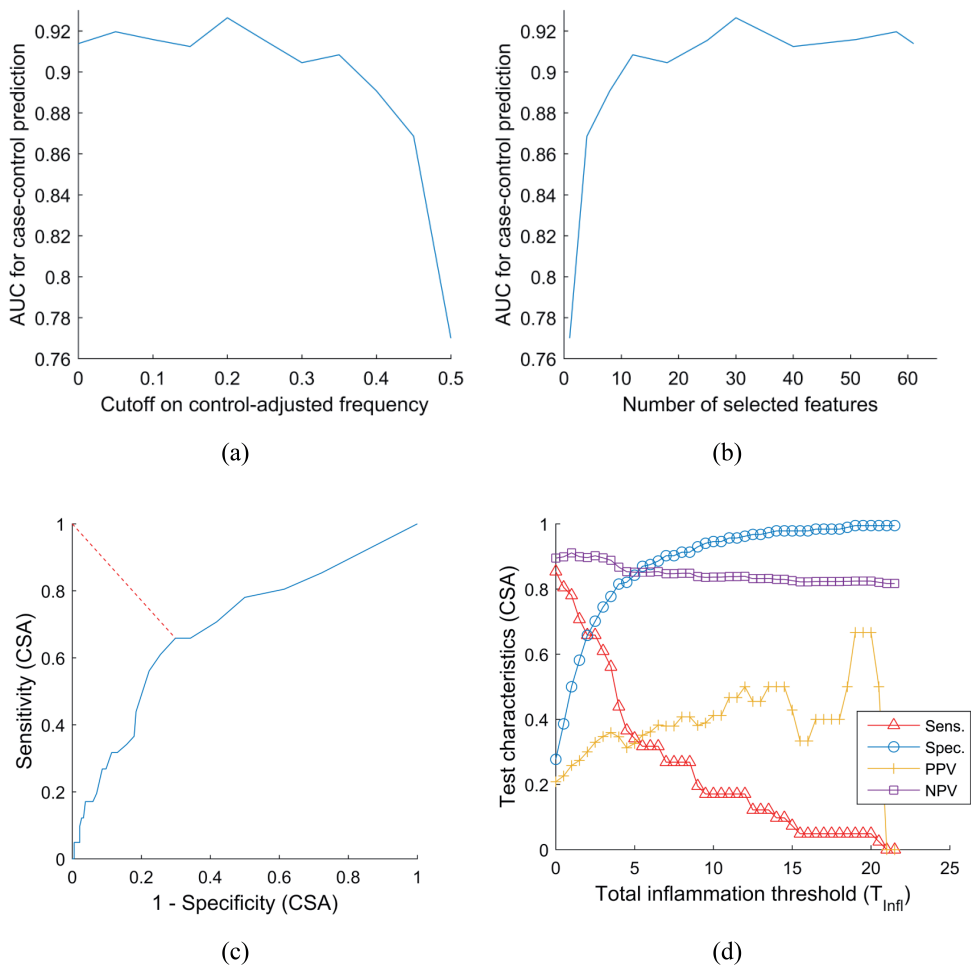


Figure 3. Feature identification and validation. Identification: by varying the cutoff value on control-adjusted frequency of inflammation (a), feature subsets of different sizes are produced (b), with corresponding AUC values for prediction of outcome (RA) in the case-control setting. The smallest subset of features yielding an AUC comparable to the complete set was observed at cutoff value 0.2 and consisted of 30 features. Validation: ROC curve (c) and diagnostic test characteristics (d) for prediction of arthritis development in CSA patients, based on the subset of features obtained from the case-control setting.

configuration produced a sensitivity of 66%, specificity of 70%, PPV of 33%, and NPV of 90%. As recognized above, this sub-analysis was subject to overfitting, because in this case the total inflammation threshold was optimized using validation data.

Table 3. Diagnostic test characteristics for prediction of arthritis development in CSA patients

	Sens. (95% CI)	Spec. (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC (95% CI)
30 features subset ($T_{\text{Inf}}=4.5$)	37% (22%-51%)	82% (77%-88%)	31% (18%-44%)	85% (80%-91%)	0.69 (0.59-0.78)
Van Steenberg <i>et al.</i> 61 features with age-referencing	80% (68%-93%)	56% (49%-63%)	29% (21%-37%)	93% (88%-98%)	0.68 (0.60-0.77)
<i>Sub-analysis:</i> 30 features subset ($T_{\text{Inf}}=2.5$)	66% (51%-80%)	70% (63%-77%)	33% (23%-43%)	90% (85%-95%)	0.69 (0.59-0.78)

Presented are the diagnostic test characteristics for prediction of arthritis development within two years of baseline MRI in 225 patients with clinically suspect arthralgia. Sens. = sensitivity; Spec. = specificity; PPV = positive predictive value; NPV = negative predictive value; AUC = area under the curve; T_{Inf} = total inflammation threshold.

Discussion

This study identified a subset of RA-specific joint-level MRI-detected inflammatory features in a case-control setting and validated them for prediction of progression to clinical arthritis in patients with CSA. The comparable AUCs of the presented method and the reference method of Van Steenberg *et al.* [3] suggest that it is possible to preserve discriminative ability while scoring only half (30/61) of the features that are typically scored, mainly focusing on locations of (teno)synovitis and leaving out the majority of BME locations. Furthermore, the presented method does not require referencing inflammation levels with respect to age-matched controls during the scoring process, meaning that outcome can be assigned even when a reader is blinded to patient age. These findings indicate that the scoring of MR scans can be significantly simplified and encourage further research into the identified inflammatory features in the broader context of arthritis pathogenesis.

We have made an underlying assumption that a subset of features yielding good predictive performance on the case-control population (cases being RA patients) can also yield good diagnostic performance on the CSA population, with progression to arthritis within two years of baseline MRI as the outcome. Our results confirm this assumption. The quality of diagnostic performance in CSA patients should be judged in comparison to the method of Van Steenbergen *et al.* [3], since it exploits the entire set of 61 features. To that end, comparison between AUCs is more informative than comparison between sensitivity/specificity pairs, since the latter depend on the definition of the optimal point on the ROC curve. As Figure 3(d) illustrates, a range of combinations of test characteristic values are achievable depending on the choice of the total inflammation threshold T_{Infl} . The choice of the optimal T_{Infl} value would depend on the objective of the diagnostic test. Lower thresholds provide better trade-off between sensitivity and specificity, but result in moderate PPV. On the other hand, higher thresholds yield higher PPV and specificity, but result in low sensitivity.

In practice, a diagnostic test for progression from CSA to clinical arthritis would combine any MRI-detected inflammatory features with other RA biomarkers, such as ACPA and C-reactive protein. MRI-detected inflammation should be seen as a potential complement to other features, not as a substitute. The discovery of a smaller subset of joint-level features that capture the overall diagnostic capacity of MRI-detected inflammation with respect to arthritis development raises questions about whether the underlying biological processes driving the inflammation at the identified anatomical locations can lead to a better understanding of arthritis pathogenesis and, ultimately, improved early diagnosis and treatment of the disease.

The identified subset of features is dominated by tenosynovitis and synovitis. This bridges earlier findings about the role of these features in early arthritis patients, CSA patients, and symptom-free persons [3,6,18] and extends the findings of Kleyer *et al.* [5] about tenosynovitis and its association with arthritis development. Interestingly, the only BME locations included in the subset were

MTP5 and the triquetrum. The location of MTP5 is known to show the first erosion in RA patients, before an erosion can be identified in the hand or wrist [19]. The BME in the triquetrum is less easily explained. Insertion of intercarpal ligaments might play a role. This study shows that commonly seen subtle BME in the carpal bones and heads of metacarpal bones is not specific for RA patients. Also, BME secondary to arthrosis (e.g. the scaphotrapeziotrapezoidal joint), subchondral cysts, and avascular necrosis of the lunate are common findings that are frequently not secondary to RA. It is important to underline that here we examine features with the purpose of differentiating RA patients from subjects without clinical arthritis. BME remains an important predictor of erosive progression in patients with established RA.

This study was limited to populations in our region, and therefore, further replication studies are needed to confirm the findings in other populations. Another limitation is that in the first 78 patients in the CSA cohort and in 114 patients from the EAC cohort, MRI of the feet was acquired without contrast enhancement and only in the axial plane (relative to the anatomical position). Since no coronal scans of the foot were available for these patients, MTP tenosynovitis features were not scored, and therefore not included in this study. In the absence of post-contrast MRI, scoring was done conservatively, which may have resulted in underestimation of inflammation for MTP synovitis features [20]. In EAC patients, this could have resulted in a lower estimate of the case frequency of inflammation presence. However, since all MTP synovitis features were included in the identified subset, this did not influence final results. On the other hand, in CSA patients, underestimation of severity of MTP synovitis could have resulted in a lower estimate of the method's sensitivity. With regard to feature selection and outcome prediction, the total inflammation score across the subset of identified features assumed equal weighting of all features. However, considering the non-uniformity of control-adjusted frequency across features, it is possible that a non-uniform weighting of features could improve diagnostic performance. This can be explored in future studies. Finally, only patients meeting the 1987 ACR criteria

were selected as cases. The 1987 classification criteria for RA are quite stringent, which has to be considered for generalizability of our results to all patients with inflammatory arthritis.

We believe that the use of MRI in research setting has important strengths, such as reproducibility and generally being well tolerated by our patients, which combined with its sensitivity to inflammation justify the acquisition of MR images in patients with imminent RA and established RA. To assess the potential value of MRI in daily practice for prediction of progression from CSA to RA, further replication studies in other CSA populations are needed. Comparison of inflammation on MRI and ultrasound imaging should also be investigated, as this could have implications for the need of MR imaging (which is both more expensive and laborious than ultrasound). Future studies will also need to look into the added value of acquiring images of both hands and feet. It has been shown that MRI-detected inflammation in the feet is common among early RA patients [21], and that a combined evaluation of hands and feet can help identify patients with continuing disease activity which would have been missed when considering clinical response in hands alone [22]. Larger studies replicating these findings are warranted.

Conclusion

In conclusion, our results indicate that a reduced subset of 30 out of 61 commonly evaluated MRI-detected inflammatory features achieves comparable diagnostic performance in prediction of arthritis development in patients with CSA. This finding suggests a considerable reduction of scoring efforts, facilitating further studies into the diagnostic value of MRI in CSA. In addition, the reduced subset of joint-level features opens new research questions about the processes driving the inflammation at the identified anatomical locations and whether this can help gain better understanding of arthritis pathogenesis.

Supplementary Material

Notes on MRI protocol

In the first 78 patients in the CSA cohort and 114 patients from the EAC cohort, MRI of the forefoot was acquired only in the axial plane (relative to the anatomical position) using a T1-weighted FSE sequence (TR/TE 400/12.5 ms, acquisition matrix 388×256, ETL 2) and a T2-weighted FSE fat-saturated sequence (TR/TE 3300/53 ms, acquisition matrix 300×252, ETL 7). In the remaining 147 patients in the CSA cohort and 85 patients from the EAC cohort, the T1-Gd sequences listed in the main text were acquired in both the axial and coronal planes.

According to the RAMRIS method [4], T2-weighted fat-suppressed sequences, or when this sequence is not available a short tau inversion recovery (STIR) sequence, should be used to assess bone marrow edema (BME). Previously, three studies have demonstrated that a contrast-enhanced T1-weighted fat-suppressed sequence has a strong correlation with T2-weighted fat-suppressed sequences [23–25]. A T2-weighted image shows increased water signal and a contrast-enhanced T1-weighted sequence shows increased water content and the increased perfusion and interstitial leakage. A strong correlation has been shown in arthritis patients but also in patients without inflammatory diseases such as bone bruises, intraosseous ganglions, bone infarcts, and even nonspecific cases [24,25]. We used the contrast-enhanced T1-weighted fat-suppressed sequence as it allowed for a shorter scan time.

Supplementary Table A.1.

Inter-reader and intra-reader intraclass correlation coefficients (ICC) for MRI scoring

Inter-reader ICC for the CSA cohort	Reader 1	Reader 2		
Reader 1	x	0.97		
Reader 2	0.97	X		
Inter-reader ICC for the EAC cohort	Reader 3	Reader 4		
Reader 3	x	0.95		
Reader 4	0.95	X		
Inter-reader ICC for the symptom-free controls	Reader 1	Reader 2		
Reader 1	x	0.96		
Reader 2	0.96	X		
Intra-reader ICC	Reader 1	Reader 2	Reader 3	Reader 4
	0.98	0.99	0.98	0.93

REFERENCES

1. Hetland ML, Ejbjerg B, Hørslev-Petersen K, Jacobsen S, Vestergaard A, Jurik AG, et al. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann. Rheum. Dis.* 2009;68:384–90.
2. van Steenbergen HW, Aletaha D, Beart-van de Voorde LJJ, Brouwer E, Codreanu C, Combe B, et al. EULAR definition of arthralgia suspicious for progression to rheumatoid arthritis. *Ann. Rheum. Dis.* 2017;76:491–6.
3. van Steenbergen HW, Mangnus L, Reijnerse M, Huizinga TWJ, van der Helm-van Mil AHM. Clinical factors, anticitrullinated peptide antibodies and MRI-detected subclinical inflammation in relation to progression from clinically suspect arthralgia to arthritis. *Ann. Rheum. Dis.* 2016;75:1824–30.
4. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. *Ann. Rheum. Dis.* 2005;64:i3–7.
5. Kleyer A, Krieter M, Oliveira I, Faustini F, Simon D, Kaemmerer N, et al. High prevalence of tenosynovial inflammation before onset of rheumatoid arthritis and its link to progression to RA-A combined MRI/CT study. *Semin. Arthritis Rheum.* Elsevier; 2016;46:143–50.
6. Mangnus L, van Steenbergen HW, Reijnerse M, van der Helm-van Mil AHM. Magnetic Resonance Imaging-Detected Features of Inflammation and Erosions in Symptom-Free Persons From the General Population. *Arthritis Rheumatol.* 2016;68:2593–602.
7. van Nies JAB, Krabben A, Schoones JW, Huizinga TWJ, Kloppenburg M, van der Helm-van Mil AHM. What is the evidence for the presence of a therapeutic window of opportunity in rheumatoid arthritis? A systematic literature review. *Ann. Rheum. Dis.* 2014;73:861–70.
8. Krabben A, Stomp W, Huizinga TWJ, van der Heijde D, Bloem JL, Reijnerse M, et al. Concordance between inflammation at physical examination and on MRI in patients with early arthritis. *Ann. Rheum. Dis.* BMJ Publishing Group Ltd; 2015;74:506–12.
9. de Rooy DPC, van der Linden MPM, Knevel R, Huizinga TWJ, van der Helm-van Mil AHM. Predicting arthritis outcomes--what can be learned from the Leiden Early Arthritis Clinic? *Rheumatology (Oxford).* 2011;50:93–100.
10. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The

- American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 1988;31:315–24.
11. van Steenbergen HW, van Nies JAB, Huizinga TWJ, Bloem JL, Reijnierse M, van der Helm-van Mil AHM. Characterising arthralgia in the preclinical phase of rheumatoid arthritis using MRI. *Ann. Rheum. Dis.* 2015;74:1225–32.
12. Newsum EC, de Waal MWM, van Steenbergen HW, Gussekloo J, van der Helm-van Mil AHM. How do general practitioners identify inflammatory arthritis? A cohort analysis of Dutch general practitioner electronic medical records. *Rheumatology (Oxford)*. 2016;55:848–53.
13. van Steenbergen HW, van der Helm-van Mil AHM. Clinical expertise and its accuracy in differentiating arthralgia patients at risk for rheumatoid arthritis from other patients presenting with joint symptoms. *Rheumatology (Oxford)*. 2016;55:1140–1.
14. McQueen FM. Bone marrow edema and osteitis in rheumatoid arthritis: the imaging perspective. *Arthritis Res. Ther.* 2012;14:224.
15. Jimenez-Boj E, Nöbauer-Huhmann I, Hanslik-Schnabel B, Dorotka R, Wanivenhaus A-H, Kainberger F, et al. Bone erosions and bone marrow edema as defined by magnetic resonance imaging reflect true bone marrow inflammation in rheumatoid arthritis. *Arthritis Rheum.* 2007;56:1118–24.
16. Dalbeth N, Smith T, Gray S, Doyle A, Antill P, Lobo M, et al. Cellular characterisation of magnetic resonance imaging bone oedema in rheumatoid arthritis; implications for pathogenesis of erosive disease. *Ann. Rheum. Dis.* 2009;68:279–82.
17. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Kvien TK. Introduction of a novel magnetic resonance imaging tenosynovitis score for rheumatoid arthritis: reliability in a multireader longitudinal study. *Ann. Rheum. Dis.* 2007;66:1216–20.
18. Nieuwenhuis WP, Krabben A, Stomp W, Huizinga TWJ, van der Heijde D, Bloem JL, et al. Evaluation of magnetic resonance imaging-detected tenosynovitis in the hand and wrist in early arthritis. *Arthritis Rheumatol. (Hoboken, N.J.)*. 2015;67:869–76.
19. Hulsmans HMJ, Jacobs JWG, Van Der Heijde DMFM, Van Albada-Kuipers GA, Schenk Y, Bijlsma JWJ. The course of radiologic damage during the first six years of rheumatoid arthritis. *Arthritis Rheum.* 2000;43:1927–40.
20. Stomp W, Krabben A, van der Heijde D, Huizinga TWJ, Bloem JL, Østergaard M, et al. Aiming for a simpler early arthritis MRI protocol: can Gd contrast administration be eliminated? *Eur. Radiol.* 2015;25:1520–7.

21. Boutry N, Lardé A, Lapègue F, Solau-Gervais E, Flipo R-M, Cotten A. Magnetic resonance imaging appearance of the hands and feet in patients with early rheumatoid arthritis. *J. Rheumatol.* 2003;30:671–9.
22. Sewerin P, Buchbender C, Vordenbäumen S, Scherer A, Miese F, Brinks R, et al. Advantages of a combined rheumatoid arthritis magnetic resonance imaging score (RAMRIS) for hand and feet: does the RAMRIS of the hand alone underestimate disease activity and progression? *BMC Musculoskelet. Disord. BioMed Central*; 2014;15:104.
23. Stomp W, Krabben A, van der Heijde D, Huizinga TWJ, Bloem JL, van der Helm-van Mil AHM, et al. Aiming for a shorter rheumatoid arthritis MRI protocol: can contrast-enhanced MRI replace T2 for the detection of bone marrow oedema? *Eur. Radiol.* 2014;24:2614–22.
24. Schmid MR, Hodler J, Vienne P, Binkert CA, Zanetti M. Bone Marrow Abnormalities of Foot and Ankle: STIR versus T1-weighted Contrast-enhanced Fat-suppressed Spin-Echo MR Imaging. *Radiology.* 2002;224:463–9.
25. Mayerhoefer ME, Breitenseher MJ, Kramer J, Aigner N, Norden C, Hofmann S. STIR vs. T1-weighted fat-suppressed gadolinium-enhanced MRI of bone marrow edema of the knee: Computer-assisted quantitative comparison and influence of injected contrast media volume and acquisition parameters. *J. Magn. Reson. Imaging.* 2005;22:788–93.

6

Summary and general discussion

In this thesis, we developed several computer-aided methods for assessment of MRI-detected inflammation in patients with inflammatory arthritis. The described studies focused on the tasks of comparative visualization, automatic quantification, and feature selection, with the underlying aim of aiding early diagnosis of spondyloarthritis (SpA) and rheumatoid arthritis (RA).

Chapter 2 presented an interactive scoring tool for evaluation of inflammatory changes over time in patients with axial SpA. Locally-rigid image registration was applied to compensate for patient posture differences between scanning sessions and fuse baseline and follow-up MR scans into a single color-encoded image. The resulting visualization offered vivid distinction between areas of increase versus decrease in inflammation over time, coupled with automatic labeling of vertebral units (VUs) and an interactive scoring module whose entry fields were activated in synchronization with the VU selected by the reader in the image. Expert readers found that a key advantage of such computer-aided scoring was that it allowed for direct visualization and measurement of inflammatory changes from a single image, as opposed to two separate images. In addition, the synchronization between the image and the scoring module significantly decreased the chance of mistyping errors while filling out the digital scoring form. At the same time, the moderate inter-reader agreement on the degree of inflammatory change pointed to the need of further standardizing interpretation of such color-encoded visualization. To this end, automatic quantification of the degree of change could be the ultimate desirable goal.

Chapter 3 proposed a framework for automatic quantification of bone marrow edema on MRI of the wrist, for early detection of RA. To combine image data from the coronal and axial sequences into a single 3D image, super-resolution reconstruction was applied. The carpal bones were located using atlas-based segmentation and signal associated with bone marrow edema was identified by fuzzy clustering. Correlation between quantitative measurements and visual scores was assessed in a large cohort of early arthritis patients. The resulting measurements were largely consistent with visual scores, indicating that automatic quantification of bone marrow edema on MRI of the wrist is feasible. It was observed, however, that incomplete fat suppression during MRI acquisition can have an adverse effect on measurement accuracy, resulting in false detections. Solving this requires further improvement of the quantification method.

Chapter 4 extended and further developed the method of **Chapter 3** to measure tenosynovitis of the extensor and flexor tendons of the wrist. Atlas-based segmentation was used to locate the bones and place initial landmarks for tendon regions. These initial landmarks were then used as inputs for marker-based watershed segmentation. A measurement region of interest was defined around the tendons. As in **Chapter 3**, signal associated with inflammation was identified using fuzzy clustering, with the modification of obtaining a one-sided probability map. Correlation between quantitative measurements and visual scores was assessed in a large cohort of early arthritis patients. Strong correlation was observed, indicating that automatic quantification of tenosynovitis on MRI of the wrist is feasible. The study also brought out multiple challenges pertinent to the quantification task, such as moderate segmentation performance and sources of false detections. In particular, blood vessels and synovitis present with the measurement region of interest were two strongly contributing factors to a consistent offset in quantitative measurements.

Chapter 5 sought to identify whether the common set of 61 MRI-detected inflammatory features visually graded across the wrist, metacarpophalangeal, and metatarsophalangeal joints can be reduced to a smaller set of features specific for

RA, given the knowledge that some features are also frequently present in symptom-free persons. The difference in frequency of inflammation presence was studied between 199 RA patients and 193 controls. A subset of 30 RA-specific features (mainly locations of tenosynovitis and synovitis) was obtained by applying a cutoff on the frequency difference while maximizing discriminative performance. For validation, this subset was used to predict arthritis development in 225 clinically suspect arthralgia (CSA) patients. The smaller subset demonstrated comparable predictive accuracy to the original set. These results suggest that it is possible to preserve the diagnostic capacity of MRI with regard to prediction of progression from CSA to clinical arthritis while scoring only half of the features that are typically scored. In addition, this leads to new research questions about the processes driving the inflammation at the identified anatomical locations and whether this can help gain better understanding of arthritis pathogenesis.

General discussion

This thesis contributed towards computer-aided assessment of MRI-detected inflammation in patients with inflammatory arthritis. The feasibility studies of **Chapters 2–4** provide reference points for interactive comparative visualization in axial SpA and automatic quantification of inflammation in RA, while **Chapter 5** further elucidates the diagnostic role of individual inflammatory features in prediction of RA development. Results showcase the promise of computer-aided techniques to overcome the limitations of visual scoring discussed in **Chapter 1**. In particular, on the automatic quantification front, the observed correlations between quantitative measurements and visual scores are encouraging considering that the methods were validated in a large cohort of early arthritis patients. Nevertheless, these techniques are not yet sufficiently robust and precise to be used in clinical practice. We identified a number of key challenges that must be addressed on the path to achieving this goal.

First, improving segmentation accuracy is an important direction of future work, as it is an essential component of the presented image processing

frameworks. In comparative visualization of MRI of the spine, the accuracy of locally-rigid registration is directly dependent on the precision of vertebrae segmentation. Furthermore, the labeling of VUs and consequently the interactive scoring features are entirely dependent on whether all evaluated vertebrae were detected. Within the context of automatic quantification, segmentation accuracy affects the variability and reliability of quantitative measurements. In addition to overall accuracy improvements, mislabeling errors should be investigated in detail. To ensure wide applicability in clinical practice, segmentation and quantification should also be robust to variations in MRI acquisition protocols and scanners. One possible way to account for this, as part of an atlas-based framework, could be to form sub-atlases of images acquired under a range of echo/repetition times and magnetic field strengths. Then, prior to segmenting a target image, the most appropriate sub-atlas would be identified based on acquisition parameters recorded in the image's DICOM data. It should be noted that the underlying approach of segmenting by drawing on knowledge accrued in a set of annotated images that constitute an atlas, or more generally a training set, is not unique to atlas-based segmentation. The atlases of the methods presented in this thesis can be used with other knowledge-based techniques, such as active appearance models or convolutional neural networks.

The presented quantification methods for RA focused on bone marrow edema and tenosynovitis in the wrist joint. However, as indicated by the results of **Chapter 5**, features that are predictive for progression to clinical arthritis also include synovitis and are spread not only across the wrist joint, but also the metacarpophalangeal joints in the hand and metatarsophalangeal joints in the foot. Therefore, it is important to expand the quantification framework to these joints and include the measurement of synovitis. The atlas-based nature of the framework provides a straightforward path for including additional joints by adding manually annotated atlases of these joints to the wrist atlas. To incorporate synovitis measurements, inflammation could be measured in the anatomical regions bounded by bones and tendon regions, which are already segmented.

In this thesis, we assessed the consistency between quantitative measurements and semi-quantitative visual scores by evaluating the correlation between these measures. This allowed us to establish that automatic quantification of MRI-detected inflammatory features frequently seen in RA patients is feasible and is largely consistent with visual scoring. To go beyond feasibility and towards rigorous evaluation of true positive versus false positive detections of inflammation, it is important that future studies assess absolute agreement between regions identified as inflammation by quantitative measurements and ground truth manual segmentations of inflammatory features by human experts. This would demand a large investment of resources, since human experts would need to manually segment all voxels considered to be part of each inflammatory feature. However, such studies may be essential to demonstrate a convincing level of agreement between automatic techniques and human experts, in order to facilitate the use of such computer-aided methods in clinical practice.

Ideally, application of automatic methods as part of a future clinical routine should be possible directly after acquisition of an MR scan, as soon as the DICOM image is stored in the patient database. However, at the present this would not be possible due to a number of artifacts that occur during acquisition and require correction prior to running quantitative analysis. For example, form entry errors can be made in the DICOM fields with regard to which location was scanned (e.g. wrist, foot) and on which side (left, right). A mistake in one of these fields will cause the atlas-based frameworks to use the wrong atlas for segmentation and result in failed quantification. Another issue is that MR scans acquired with a frequency-selective fat-saturated sequence sometimes suffer from fat suppression inhomogeneity, which may cause quantitative measurements to confuse regions of fat tissue for inflammation. There can be two general approaches to addressing these acquisition issues: 1) improve acquired image quality requirements and systematically minimize the possibility of DICOM field errors through stricter protocols and software interface, or 2) develop automatic methods for handling acquisition issues as part of the overall computer-aided framework. The choice

boils down to resource management. As computer-aided techniques become more integrated in clinical practice, a balanced approach could be to let operating technicians manage those acquisition issues that can be fixed with minimal additional inconvenience and cost for patients at acquisition time. On the other hand, correction of acquisition issues that would substantially increase patient inconvenience, cost, and procedure time can be delegated to the image processing engineers.

In conclusion, this thesis has explored the prospect of computer-aided assessment of MRI-detected inflammation for early identification of inflammatory arthritis. The presented studies showcase the potential of comparative visualization and automatic quantification to overcome the limitations of visual scoring and lay out a fertile ground for future improvements. Additionally, the understanding of the diagnostic role of individual inflammatory features in prediction of RA development is further advanced. Collectively, these findings can help facilitate the use of MRI for early diagnosis of inflammatory arthritis and potentially increase chances of better outcome and quality of life for patients.

7

Samenvatting en algemene discussie

In dit proefschrift hebben wij verschillende methoden ontwikkeld voor de beoordeling van MRI-gedetecteerde ontstekingen in patiënten met inflammatoire artritis. De beschreven onderzoeken zijn erop gericht om taken uit te voeren voor de vergelijkende visualisatie, automatische kwantificatie en kenmerkextractie, met het onderliggende doel om de vroegdiagnostiek van spondylartritis (SpA) en reumatoïde artritis (RA) te ondersteunen.

Hoofdstuk 2 presenteerde een interactieve beoordelingsmethode voor de evaluatie van inflammatoire veranderingen over de tijd in patiënten met axiale SpA. Lokaal-rigide beeldregistratie was toegepast om te compenseren voor verschillen in houding van de patiënt tussen scansessies en om de baseline en follow-up MR-scans te fuseren tot één kleur-gecodeerd beeld. De resulterende visualisatie gaf een duidelijk onderscheid weer tussen gebieden van verhoogde en verlaagde ontstekingsgraad over de tijd, gekoppeld aan het automatisch labelen van de werveleenheid (WEs) en een interactieve scoringsmodule, waarvan de invulvelden synchroon werden geactiveerd corresponderend met door de beeldbeoordelaar geselecteerde WE. Expert beoordelaars verklaarden dat het belangrijkste voordeel van zo'n computerondersteunde beoordeling ligt in de mogelijkheid om direct de inflammatoire veranderingen te visualiseren en te beoordelen in één enkel beeld, in plaats van twee aparte beelden. Daarnaast zorgde de synchronisatie tussen het beeld en de beoordelingsmodule ervoor, dat de kans op typefouten bij het invullen van het digitale invulformulier was verminderd. Tegelijkertijd toonde de matige overeenstemming tussen beoordelaars v.w.b. de mate van inflammatoire veranderingen aan, dat verdere standaardisatie van de

interpretatie van dergelijke kleur-gecodeerde visualisaties nodig is. Daarom zou een automatische kwantificatie van de mate van veranderingen het uiteindelijke gewenste doel zijn.

Hoofdstuk 3 stelde een raamwerk voor om beenmergoedeem in MRI van de pols automatisch te kwantificeren, voor de vroege detectie van RA. Voor het combineren van beelddata van coronale en axiale sequenties tot één enkel 3D beeld was super-resolutie-reconstructie toegepast. De carpale botten werden gelokaliseerd door gebruik te maken van atlas-gebaseerde segmentatie en de signalen, die geassocieerd zijn met beenmergoedeem, werden geïdentificeerd door fuzzy-clustering. De correlatie tussen kwantitatieve metingen en visuele scores was vastgesteld in een groot cohort van patiënten met vroege artritis. De resulterende metingen waren grotendeels consistent met de visuele scores, hetgeen aangeeft dat automatische kwantificatie van beenmergoedeem in MRI van de pols haalbaar is. Er werd echter opgemerkt dat onvolledige vetsuppressie tijdens de MRI acquisitie een nadelig effect kan hebben op de meetnauwkeurigheid. De oplossing hiervoor vergt een verdere verbetering van de acquisitiemethode.

Hoofdstuk 4 breidde de methode van **Hoofdstuk 3** uit en ontwikkelde hem verder om tenosynovitis van de extensor- en flexorpezen in de pols te meten. Atlas-gebaseerde segmentatie was gebruikt om de botten te lokaliseren en initiële oriëntatiepunten voor de peesgebieden te plaatsen. De initiële oriëntatiepunten werden vervolgens gebruikt als invoer voor marker-gebaseerde ‘watershed’-segmentatie. Een meetgebied was gedefinieerd rond de pezen. Zoals in **Hoofdstuk 3**, het signaal dat geassocieerd is met ontstekingen was geïdentificeerd door fuzzy-clustering met de aanpassing dat een afbeelding van enkelzijdige kansen was verkregen. De correlatie tussen de kwantitatieve metingen en visuele scores was vastgesteld in een groot cohort van patiënten met vroege artritis. Het feit dat een sterke correlatie was gevonden geeft aan dat automatische kwantificatie van tenosynovitis in MRI van de pols mogelijk is. De studie bracht ook verschillende uitdagingen aan het licht die betrekking hebben op het kwantificeren, zoals het beperkte succes van de segmentatie en oorzaken van onjuiste detectie. In het

bijzonder waren bloedvaten en synovitis, die aanwezig waren in het meetgebied, twee belangrijke bijdragen aan de consequente afwijkingen in de kwantitatieve metingen.

Hoofdstuk 5 onderzocht de mogelijkheid of de gebruikelijke verzameling van 61 MRI-gedetecteerde inflammatoire kenmerken die visueel beoordeeld zijn in de pols, metacarpofalangeale en metatarsofalangeale gewrichten, zou kunnen worden versimpeld tot een kleinere deelverzameling van RA-specifieke eigenschappen, gezien het feit dat sommige eigenschappen vaak ook aanwezig zijn in symptoom-vrije personen. Het verschil in frequentie van aanwezige ontstekingen tussen 199 RA-patiënten en 193 controles was bestudeerd. Een deelverzameling van 30 RA-specifieke kenmerken was verkregen (vooral locaties met tenosynovitis en synovitis) door exclusie van kenmerken op grond van de frequentieverschillen, terwijl de onderscheidende waarde werd gemaximaliseerd. Voor de validatie was deze deelverzameling gebruikt om artritis te voorspellen in 225 klinisch verdachte artralgie (CSA) patiënten. De kleinere deelverzameling liet een voorspellingsnauwkeurigheid zien die vergelijkbaar was met de originele verzameling. Deze resultaten suggereren dat het mogelijk is om de diagnostische waarde van MRI te behouden met betrekking tot de voorspelling van de progressie van CSA naar klinische artritis, terwijl slechts de helft van de gebruikelijke kenmerken hoeft te worden gescoord. Daarnaast leidt dit onderzoek tot nieuwe onderzoeksvragen over de processen die ontstekingen in de geïdentificeerde anatomische locaties aansturen, en of dit kan helpen om een beter inzicht te verkrijgen in de pathogenese van artritis.

Algemene discussie

Dit proefschrift draagt bij aan de computerondersteunde beoordeling van MRI-gedetecteerde ontstekingen in patiënten met inflammatoire artritis. De haalbaarheidsstudies van **Hoofdstuk 2-4** geven referentiepunten voor interactieve visualisatie in axiale SpA en automatische kwantificatie van ontstekingen in RA, terwijl **Hoofdstuk 5** de diagnostische rol duidelijk maakt van individuele

ontstekingseigenschappen in het voorspellen van RA ontwikkeling. De resultaten geven de belofte dat computerondersteunde technieken de beperkingen van het visueel scoren, zoals die in **Hoofdstuk 1** zijn beschreven, kan overwinnen. In het bijzonder op het gebied van automatische kwantificatie zijn de gevonden correlaties tussen kwantitatieve metingen en visuele scores bemoedigend, gezien het feit dat de methoden waren gevalideerd in een groot cohort van patiënten met vroege artritis. Desondanks zijn deze technieken nog niet robuust en nauwkeurig genoeg om te worden gebruikt in de klinische praktijk. We hebben een aantal belangrijke uitdagingen geïdentificeerd die aangepakt moeten worden om dit doel te bereiken.

Allereerst is de verbetering van de segmentatienauwkeurigheid een belangrijke richting voor toekomstig onderzoek, omdat het een essentieel onderdeel is van het gepresenteerde beeldverwerkings raamwerk. In de vergelijkende visualisatie van MRI van de wervelkolom is de nauwkeurigheid van de lokaal-rigide registratie direct afhankelijk van de precisie van de wervelsegmentatie. Daarnaast zijn het labelen van de WEs en derhalve de interactieve scoringseigenschappen volledig afhankelijk van het garandeert dat alle geëvalueerde wervels gedetecteerd zijn. Bij de automatische kwantificatie beïnvloedt de segmentatienauwkeurigheid de variabiliteit en betrouwbaarheid van de kwantitatieve metingen. Naast de algemene verbeteringen in nauwkeurigheid moeten de fouten bij het labelen gedetailleerd onderzocht worden. Om een uitgebreide toepasbaarheid in de klinische praktijk te garanderen, zouden de segmentatie en kwantificatie ook robuust moeten zijn tegen variaties in het MRI-acquisitieprotocol en scanners. Om hiermee rekening te houden, als onderdeel van de atlas-gebaseerde raamwerk, zou het mogelijk kunnen zijn om, sub-atlassen te vormen van beelden die verkregen zijn met een reeks van echo/repetitietijden en magnetische veldsterktes. Voordat het betreffende beeld gesegmenteerd wordt, zou er dan eerst de meest geschikte sub-atlas geïdentificeerd worden, gebaseerd op de acquisitieparameters die zijn opgeslagen in de DICOM-gegevens van het beeld. Er moet worden opgemerkt dat de onderliggende segmentatieaanpak die de vergaarde

kennis weergeeft in een verzameling van geannoteerde beelden die samen een atlas (of meer algemeen een training set) vormen, niet uniek is voor atlas-gebaseerde segmentatie. De atlassen van de methoden die in dit proefschrift zijn gepresenteerd, kunnen worden gebruikt met andere kennis-gebaseerde technieken, zoals ‘active appearance’ modellen of convolutionele neurale netwerken.

De gepresenteerde kwantificatiemethoden voor RA waren gericht op beenmergoedeem en tenosynovitis in de polsgewrichten. Echter zoals aangeven bij de resultaten van **Hoofdstuk 5**, omvatten de kenmerken die voorspellend zijn voor de progressie naar klinische artritis, ook synovitis en zijn die niet alleen over de polsgewrichten verspreid, maar ook over de metacarpofalangeale gewrichten in de hand en metatarsofalangeale gewrichten in de voet. Daarom is het belangrijk om het kwantificatie raamwerk uit te breiden met die gewrichten en een synovitismeting toe te voegen. Het atlas-gebaseerde karakter van het raamwerk verschaft een eenvoudige mogelijkheid om extra gewrichten toe te voegen door toevoeging van handmatig geannoteerde atlassen van deze gewrichten aan de polsatlas. Om synovitismetingen toe te voegen, zouden ontstekingen gemeten kunnen worden in het anatomische gebied begrensd door botten en pees-gebieden die reeds gesegmenteerd zijn.

In dit proefschrift hebben wij de consistentie tussen kwantitatieve metingen en semi-kwantitatieve visuele scores vastgesteld door de correlatie tussen beide metingen te evalueren. Dit stelde ons in staat om vast te stellen dat automatische kwantificatie van MRI-gedetecteerde ontstekingskenmerken die vaak aangetroffen worden in RA-patiënten, haalbaar is en grotendeels consistent is met het visueel scoren. Om verder te gaan dan alleen haalbaarheid, naar een strikte evaluatie van ‘true positive’ versus ‘false positive’ detectie van ontstekingen, is het van belang dat toekomstige studies de absolute overeenstemming vaststellen tussen de gebieden, die als ontsteking geïdentificeerd zijn door de kwantitatieve metingen, en de gouden standaard handmatige segmentaties van ontstekingen door menselijke experts. Dit zou een grote investering in middelen vragen, aangezien menselijke experts alle voxels die als onderdeel beschouwd worden van elk

ontstekingskenmerk, handmatig zouden moet segmenteren. Echter zulke studies zouden essentieel kunnen zijn om een overtuigende mate van overeenkomst aan te tonen, om het gebruik van zulke computerondersteunde methoden in de klinische praktijk te faciliteren.

In het ideale geval zou de toepassing van automatische methoden als onderdeel van toekomstig klinische routine mogelijk zijn, direct na het verkrijgen van een MR-scan zodra het DICOM-beeld is bewaard in de patiëntendatabase. Echter, momenteel, zou dit niet mogelijk zijn vanwege het aantal artefacten die zich voordoen tijdens de acquisitie en vereist correcties voorafgaand aan het uitvoeren van de kwantitatieve analyse. Bijvoorbeeld invoerfouten in een formulier zouden gemaakt kunnen worden in de DICOM-velden, bij de locatie dat gescand was (bijv. pols, voet) en bij welke zijde (links, rechts). Een vergissing in één enkel veld zal ervoor zorgen dat het atlas-gebaseerde raamwerk de verkeerde atlas gebruikt en resulteren in een mislukte kwantificatie. En ander kwestie is dat MR-scans verkregen met een frequentie-selectieve ‘fat-saturated’ sequentie soms vet-suppressie inhomogeniteiten ondervinden, die ervoor zorgen dat de kwantitatieve metingen vetweefselgebieden verwarren met ontstekingen. Er kunnen twee algemene benaderingen zijn om deze acquisitieproblemen aan te pakken: 1) verbeter de verkregen beeldkwaliteitseisen en minimaliseer systematisch de mogelijkheid van fouten in de DICOM-velden door striktere protocollen en software interface, of 2) ontwikkel een automatische methode om met acquisitieproblemen om te gaan als onderdeel van het algemene computerondersteunde raamwerk. De keuze komt neer op middelenbeheer. Aangezien computerondersteunde technieken steeds meer geïntegreerd worden in de klinische praktijk, zou het een evenwichtige aanpak kunnen zijn om de dienstdoende laborant deze acquisitieproblemen te laten hanteren, die opgelost kunnen worden tijdens de beeldacquisitie met minimale extra ongemakken en kosten voor de patiënt. Aan de andere kant zou de correctie van acquisitieproblemen overgelaten kunnen worden aan de beeldverwerkings-

ingenieur, als die problemen het patiëntongemak, de kosten en proceduredtijd substantieel zouden vergroten.

Tot slot, heeft dit proefschrift het vooruitzicht van computerondersteunde beoordeling van MRI-gedetecteerde ontstekingen onderzocht voor de vroege identificatie van inflammatoire artritis. De gepresenteerde studies demonstreren de mogelijkheid van vergelijkende visualisatie en automatische kwantificatie om de beperkingen van visuele beoordeling te overwinnen en legt een vruchtbare grond neer voor toekomstige verbeteringen. Daarnaast is het inzicht in de diagnostische rol van individuele ontstekingskenmerken in de voorspelling van RA ontwikkeling verder verbeterd. Gezamenlijk kunnen deze bevindingen het gebruik van MRI bij de vroegdiagnostiek van inflammatoire artritis helpen vergemakkelijken en mogelijk de kansen op een betere uitkomst en kwaliteit van leven van patiënten verhogen.

List of publications

E. Aizenberg, R. van den Berg, Z. Ez-Zaitouni, D. van der Heijde, M. Reijnierse, O. Dzyubachyk, B.P.F. Lelieveldt, “Computer-aided evaluation of inflammatory changes over time on MRI of the spine in patients with suspected axial spondyloarthritis: a feasibility study,” *BMC Medical Imaging*, 17:55, 2017.

E. Aizenberg, E.A.H. Roex, W.P. Nieuwenhuis, L. Mangnus, A.H.M. van der Helm–van Mil, M. Reijnierse, J.L. Bloem, B.P.F. Lelieveldt, B.C. Stoel, “Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: a feasibility study,” *Magnetic Resonance in Medicine*, vol. 79(2), pp. 1127–1134, 2018.

E. Aizenberg^{*}, D.P. Shamonin^{*}, M. Reijnierse, A.H.M. van der Helm–van Mil, B.C. Stoel, “Automatic quantification of tenosynovitis on MRI of the wrist in patients with early arthritis: a feasibility study,” *European Radiology*, doi: 10.1007/s00330-018-5807-2, 2018.

^{*}Authors contributed equally.

E. Aizenberg^{*}, R.M. ten Brinck^{*}, M. Reijnierse, A.H.M. van der Helm–van Mil, B.C. Stoel, “Identifying MRI-detected inflammatory features specific for rheumatoid arthritis: two-fold feature reduction maintains predictive accuracy in clinically suspect arthralgia patients,” *Seminars in Arthritis and Rheumatism*, doi: 10.1016/j.semarthrit.2018.04.005, 2018.

^{*}Authors contributed equally.

Acknowledgements

The PhD journey is full of ups and downs, and like any journey in life, the people you share it with can make all the difference in the world on how you come out at its end. I am very fortunate with the people I got to share my journey with.

First of all, I would like to thank my supervisors, Boudewijn and Berend, for providing me with the opportunity to pursue my PhD at LKEB and their continuous support throughout the past four years. Berend, thank you for always being available for open-minded and frank discussions, for asking questions that encouraged me to think critically about my work, and for giving me the freedom to pursue this research creatively. I am also very thankful to you for volunteering your time and effort to translate the summary and discussion chapter into Dutch!

This research would not be possible without the awesome, insightful, and fruitful collaboration with my co-supervisors and PhD colleagues from the departments of Rheumatology and Radiology. Annette and Monique, thank you so much for your guidance, support, and advice, for shaping and paving this research together. To my fellow PhDs, Rosaline, Freek, Zineb, Wouter, Lukas, Robin, it was a pleasure to work together. Thank you so much for your time, your insights, and your openness. Rosaline and Robin, I got to spend a lot of time with each of you, and I am so grateful for your patience and willingness to explore together. I would also like to thank Prof. Désirée van der Heijde for helping shape the first chapter of this thesis and our numerous discussions leading up to its publication.

I could not have been more fortunate with the amazing fellow colleagues at LKEB who created a warm, welcoming, and engaging environment to work in every day. To my fellow Terminalzaal'ers, Baldur, Qian, Paulien, Ronald, Trung, Floris, Niels, Thomas, Qing, thank you for being the awesome office mates you are since

my first day at LKEB and throughout this whole journey. Baldur, thanks for sharing your knowledge and expanding my horizons during our many conversations, whether it was about programming, history, politics, or life. To Oleh, I am deeply grateful for your supervision during my first year at LKEB, your support and help during the PhD, and your warm friendship. Edgar, your work laid out important foundations for this research, and I am very thankful for the many productive and enlightening discussions we held. Denis, your vast knowledge and skills with programming were a life savior throughout these years, thank you so much for always willing to help out! I am happy we got to work a lot together and learn from each other during this project. To Leo, thanks for all the cool conversations we shared about many a curious topic and your sincerely welcoming attitude and support. Michèle, thank you so much for making it possible for us to have all the equipment we as PhDs need to do the job and always helping out to make sure it works smoothly. To all the great people of LKEB with whom I have been fortunate to share this journey together, thank you for making our team a diverse, creative, and most importantly warm and welcoming workplace for everyone.

Moving to another country is hard, for a whole bunch of reasons. But I have been lucky to meet other wandering kindred spirits, here in the Netherlands, who became dear friends. Giorgos, Nefeli, Yannis, Alexia, Roberta, Eleni, Ángela, Aleix, Emilija, thank you for your honest friendship, support, and all the fun moments we shared here in our life travels in the Low Lands.

Last but most certainly not least, I want to thank my family, my dear loved ones, for always being there for me no matter the distance between us, for believing in me, and for inspiring me to strive for something bigger than myself by being a citizen of this world.

Curriculum Vitae

Evgeni Aizenberg was born in Moscow, USSR, on June 28, 1990. In 2012, he received his B.Sc. degree in Electrical Engineering from Boston University, USA, where he worked as an undergraduate researcher at the Biomedical Optics Lab and conducted his bachelor thesis on automated detection of colon pre-cancer based on in vivo endomicroscopy images. In 2014, he received his M.Sc. degree in Electrical Engineering from Delft University of Technology, The Netherlands. His master thesis on computer-aided evaluation of inflammatory changes on MRI of the spine in patients with suspected axial spondyloarthritis was carried out at the Division of Image Processing (LKEB) at Leiden University Medical Center, The Netherlands, and the same year he joined LKEB as a PhD researcher. The goal of his PhD work was development of computer-aided techniques for assessment of MRI-detected inflammation for early identification of inflammatory arthritis. The results of his research are published in various scientific journals and are included in this thesis.

