



Universiteit  
Leiden  
The Netherlands

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

Aizenberg, E.

### **Citation**

Aizenberg, E. (2019, March 14). *Computer-aided techniques for assessment of MRI-detected inflammation for early identification of inflammatory arthritis*. Retrieved from <https://hdl.handle.net/1887/68704>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/68704>

**Note:** To cite this publication please use the final published version (if applicable).

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

# 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. 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.

## 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 Steenberg *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 Steenberg *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  $\geq 60$  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  $\geq 3$  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  $\geq 3$  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

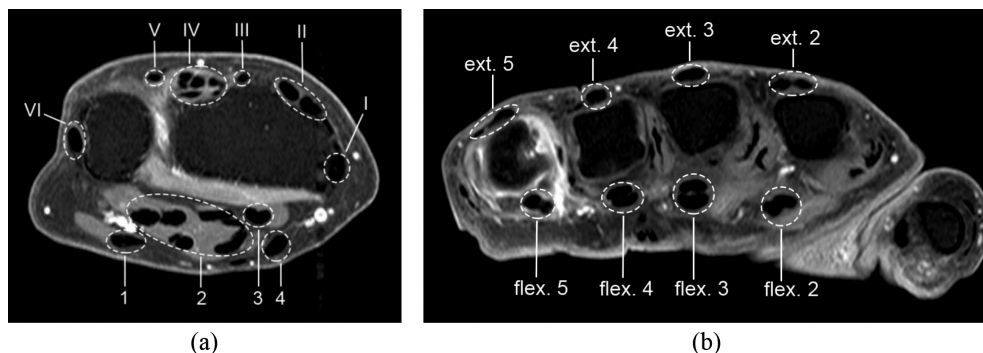


foot had 20 slices with a slice thickness of 3 mm and a slice gap of 0.3 mm. All axial sequences had a slice thickness of 3 mm and a slice gap of 0.3 mm with 20 slices for the wrist, 16 for the MCP joints, and 14 for the foot. Further information about the MRI protocol and some exceptions are described 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  $T_{\text{infl}}$ . Construct an ROC curve by varying the value of  $T_{\text{infl}}$  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  $T_{\text{infl}}$ . An ROC curve was constructed by varying the value

of  $T_{\text{inf}}$ . Diagnostic performance was quantified by AUC, sensitivity, specificity, PPV, and NPV. The latter four measures were computed for the  $T_{\text{inf}}$  value obtained from the case-control setting. For comparison of diagnostic performance, the method of Van Steenberghe *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<br>(n = 199) | Symptom-free persons<br>(n = 193) | CSA patients<br>(n = 225) |
|---------------------------------------|--------------------------|-----------------------------------|---------------------------|
| Age in years, mean (SD)               | 56.1 (14.4)              | 49.8 (15.8)                       | 44.2 (13.0)               |
| Female, n (%)                         | 127 (63.8)               | 136 (70.5)                        | 174 (77.3)                |
| BMI in kg/m <sup>2</sup> *, mean (SD) | 26.6 (4.3)               | 24.8 (3.9)                        | 27.0 (4.9)                |
| Elevated CRP, n (%)                   | 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, n (%)                | 121 (60.8)               | <i>Not assessed</i>               | 46 (20.4)                 |
| ACPA positive, n (%)                  | 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.



### **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  $T_{\text{Infl}}$  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.

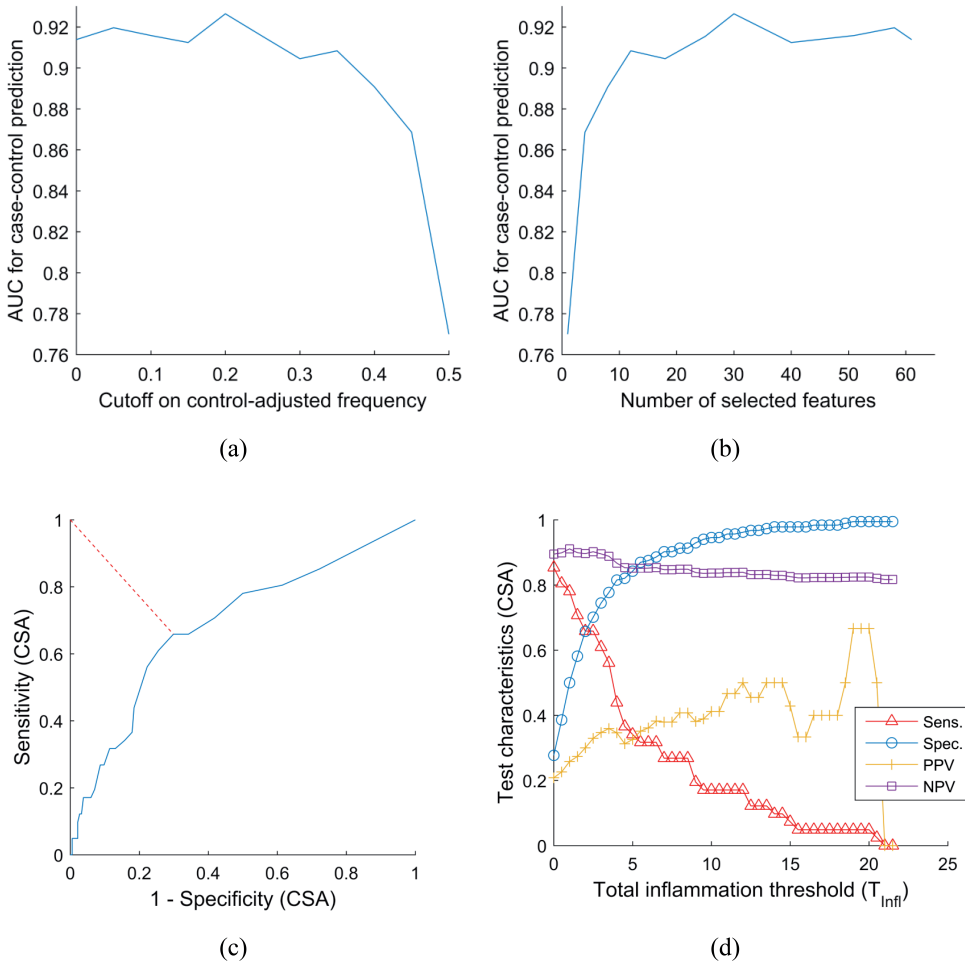
As a sub-analysis, to further explore the ROC curve produced by the subset of 30 features, the test characteristics were computed again for the point on the CSA ROC curve that was closest to (0,1), which corresponded to  $T_{\text{Infl}} = 2.5$ . This



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

| <b>Feature</b>              | <b>Location</b>         | <b>Control-adjusted frequency</b> |      |
|-----------------------------|-------------------------|-----------------------------------|------|
| 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

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

|   |          |          |          |          |
|---|----------|----------|----------|----------|
| <b>Inter-reader ICC for the CSA cohort</b>            | Reader 1 | Reader 2 |          |          |
| Reader 1  | x        | 0.97     |          |          |
| Reader 2  | 0.97     | X        |          |          |
|   |          |          |          |          |
| <b>Inter-reader ICC for the EAC cohort</b>            | Reader 3 | Reader 4 |          |          |
| Reader 3  | x        | 0.95     |          |          |
| Reader 4  | 0.95     | X        |          |          |
|   |          |          |          |          |
| <b>Inter-reader ICC for the symptom-free controls</b> | Reader 1 | Reader 2 |          |          |
| Reader 1  | x        | 0.96     |          |          |
| Reader 2  | 0.96     | X        |          |          |
|   |          |          |          |          |
| <b>Intra-reader ICC</b>                               | 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, 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.
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, Reijnierse 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, 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.
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

## Chapter 5

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, Breitschneider 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.

