Rheumatology Advances in Practice, 2024, **8(3)**, rkae073
https://doi.org/10.1093/rap/rkae073
Advance access publication 20 June 2024
Original Article

British Society for Rheumatology
RHEUMATOLOGY ADVANCES IN PRACTICE
OXFORD

Clinical science

Automated quantification of wrist bone marrow oedema, pre- and post-treatment, in early rheumatoid arthritis

Chungwun Yiu[1], James Francis Griffith [1,*], Fan Xiao[1], Lin Shi[1], Bingjing Zhou [1], Su Wu [1], Lai-Shan Tam [2]

[1]Department of Imaging and Interventional Radiology, Chinese University of Hong Kong, Hong Kong, China
[2]Rheumatology Division, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, China

*Correspondence to: James Francis Griffith, Department of Imaging and Interventional Radiology, Chinese University of Hong Kong, Rm 27023, G/F, Prince of Wales Hospital, N.T., Hong Kong, China. E-mail: griffith@cuhk.edu.hk

Abstract

Objective: Bone inflammation (osteitis) in early RA (ERA) manifests as bone marrow oedema (BME) and precedes the development of bone erosion. In this prospective, single-centre study, we developed an automated post-processing pipeline for quantifying the severity of wrist BME on T2-weighted fat-suppressed MRI.

Methods: A total of 80 ERA patients [mean age 54 years (S.D. 12), 62 females] were enrolled at baseline and 49 (40 females) after 1 year of treatment. For automated bone segmentation, a framework based on a convolutional neural network (nnU-Net) was trained and validated (5-fold cross-validation) for 15 wrist bone areas at baseline in 60 ERA patients. For BME quantification, BME was identified by Gaussian mixture model clustering and thresholding. BME proportion (%) and relative BME intensity within each bone area were compared with visual semi-quantitative assessment of the RA MRI score (RAMRIS).

Results: For automated wrist bone area segmentation, overall bone Sørensen–Dice similarity coefficient was 0.91 (S.D. 0.02) compared with ground truth manual segmentation. High correlation (Pearson correlation coefficient $r = 0.928$, $P < 0.001$) between visual RAMRIS BME and automated BME proportion assessment was found. The automated BME proportion decreased after treatment, correlating highly ($r = 0.852$, $P < 0.001$) with reduction in the RAMRIS BME score.

Conclusion: The automated model developed had an excellent segmentation performance and reliable quantification of both the proportion and relative intensity of wrist BME in ERA patients, providing a more objective and efficient alternative to RAMRIS BME scoring.

Lay Summary

What does this mean for patients?

Increased fluid in the bone marrow, known as bone marrow oedema, is a measure of bone inflammation in people with rheumatoid arthritis (RA). This bone marrow oedema can be clearly seen on magnetic resonance imaging examinations of the joints. The wrist joint is one of the more commonly affected joints in RA. Bone marrow oedema is a measure of disease activity and is reversible. It is a precursor of non-reversible bone erosion or deformity. One of the treatment aims of RA is to reduce bone oedema. In this study, we present an automated method of measuring bone marrow oedema that is much faster, more reliable and more responsive than traditional visual methods. This will be helpful to accurately gauge the amount of bone inflammation in RA patients at initial presentation and following treatment.

Keywords: rheumatoid arthritis, bone marrow oedema, wrist, quantification, segmentation.

Key messages

- A method of automatically measuring BME proportion and relative intensity is presented.
- Strong correlation between the automated BME proportion and RAMRIS BME score was found.
- Decreased BME proportion and intensity with treatment correlated strongly with the RAMRIS BME score.

Introduction

RA is a chronic systemic inflammatory arthropathy with a worldwide prevalence of ≈1%. The wrist is the most affected joint. Inflammation in the wrist is a marker of systemic disease activity [1, 2]. The main manifestations of RA are inflammation in either the joint synovium ('synovitis'), tenosynovium ('tenosynovitis') or bone ('osteitis') [3]. Bone inflammation, which manifests as bone marrow oedema (BME) on MRI, is a recognized precursor of bone erosion and structural deterioration [4].

While BME is not a feature of healthy wrists [5], it is seen in at least 40% of patients with early RA and is a more specific marker of inflammatory arthritis than bone erosion [6]. In RA, BME reflects bone marrow inflammation (osteitis)

Received: 27 February 2024. **Accepted:** 22 May 2024

with infiltration by macrophages, plasma cells, T cells and B cells [7]. It is a marker of an aggressive RA phenotype associated with increased bone erosion and worse functional outcome [8]. Unlike erosions, osteitis is reversible, with a reduction in BME indicating good therapeutic response.

BME is generally scored in everyday practice as mild, moderate or severe, while in research studies it is scored semi-quantitatively using the RA MRI score (RAMRIS). For RAMRIS, each wrist bone is scored as 1–3 according to the percentage (0–100% in 33% increments) of bone showing high signal intensity on T2-weighted fat-suppressed MRI. Although the reliability of RAMRIS scoring is high [intraclass correlation coefficient 0.84 (90% CI 0.73, 0.94)], this scoring method is semi-quantitative, requires reader training and is time-consuming [9].

There are two features of BME severity that can be measured, namely the extent of bone affected by BME and the intensity of this BME. A quantitative measure of BME severity would help to objectively gauge the degree of bone inflammation and assess treatment response, as well as being helpful when investigating the aetiology and effect of BME.

The aim of this study was to develop an automated method of quantifying the extent and intensity of BME in wrist bone areas using T2-weighted fat-suppressed MRI. Such a system would allow BME on MRI to be evaluated with a quick, reliable method that would help to move BME quantification from the research to the everyday clinical arena.

Materials and methods

Data acquisition

The study flow chart is shown in Fig. 1. MRI data from a prospective longitudinal study of treatment-naïve ERA patients recruited between October 2012 and January 2016 was utilized. This study complied with all relevant ethical standards. Written informed consent was obtained from all patients and this study received ethical approval from the Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee (Joint CUHK-NTEC CREC; CREC ref. no.: 2021.398). Eighty ERA patients, defined as having a symptom duration of <24 months, were studied pre- and post-treatment. From this cohort of 80 patients, 60 were suitable for automated segmentation, with 20 MRI datasets deemed unsuitable due to movement artefact or incomplete fat suppression. A total of 49 (82%) of the 60 patients underwent an identical follow-up MRI and assessment after 1 year of treatment. Patient details

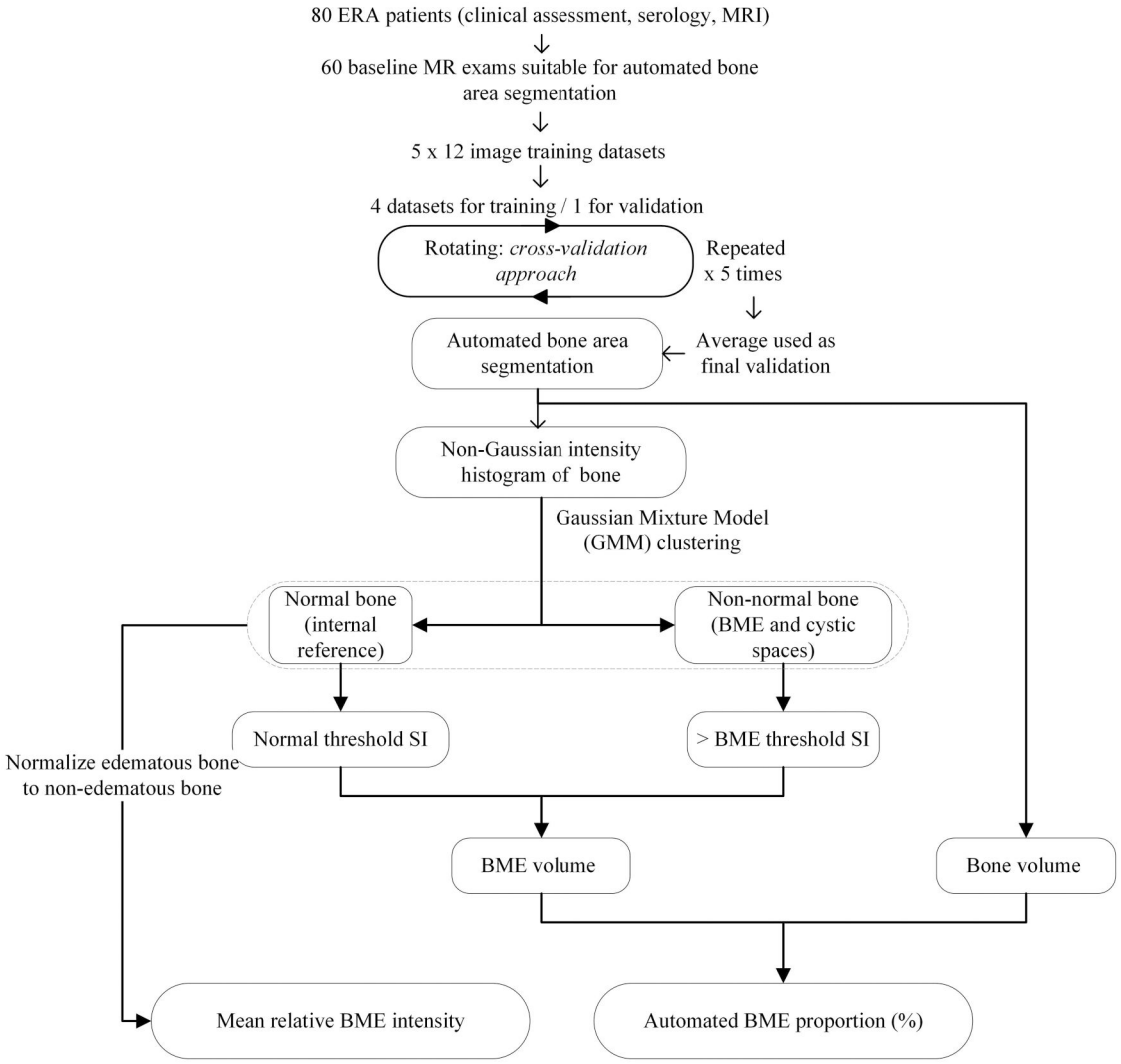

Figure 1. Study flow chart. SI: signal intensity

are summarized in Supplementary Table S1, available at *Rheumatology Advances in Practice* online.

MRI

MRI of the most symptomatic wrist was performed in a prone position on a 3.0 T system (Achieva TX, Philips Healthcare, Andover, MA, USA) with a dedicated wrist coil. Segmentation (manual and automated) was performed on T2-weighted fat-suppressed coronal images with an echo time (TE) of 70 ms, repetition time (TR) of 3121 ms, flip angle 90°, field of view 80 mm × 80 mm and a 448 × 448 reconstruction matrix yielding a pixel size of 0.178 mm × 0.178 mm with 1.65 mm slice thickness. Each MR examination comprised 20 consecutive coronal images.

Segmentation

Manual segmentation of 15 wrist bone areas in 60 wrists was undertaken by trained research assistants using the open-source software ITK-SNAP [10] supervised by a musculoskeletal radiologist with 29 years of experience in MR wrist imaging (Fig. 2).

Automated segmentation employed a convolutional neural network (nnU-Net) framework [11], implemented and trained using the Pytorch API [12] on a Tesla A100 40 GB graphics processing unit (NVIDIA, Santa Clara, CA, USA). T2-weighted coronal fat-saturated images were uploaded to the framework for training. The segmentation method used was a modification of recognized approaches to image segmentation [11]. To address limitations of a relatively small training dataset, a 5-fold cross-validation cycle approach was adopted. The 60-image dataset was divided into five sets of 12 image datasets, each dataset comprising all the T2-weighted fat-supressed images for that patient. Four sets were utilized for training and the fifth for validation. This process was repeated five times, rotating through different sets for training and validation. The average of the five results yielded the final validation result. A total of 1000 epochs

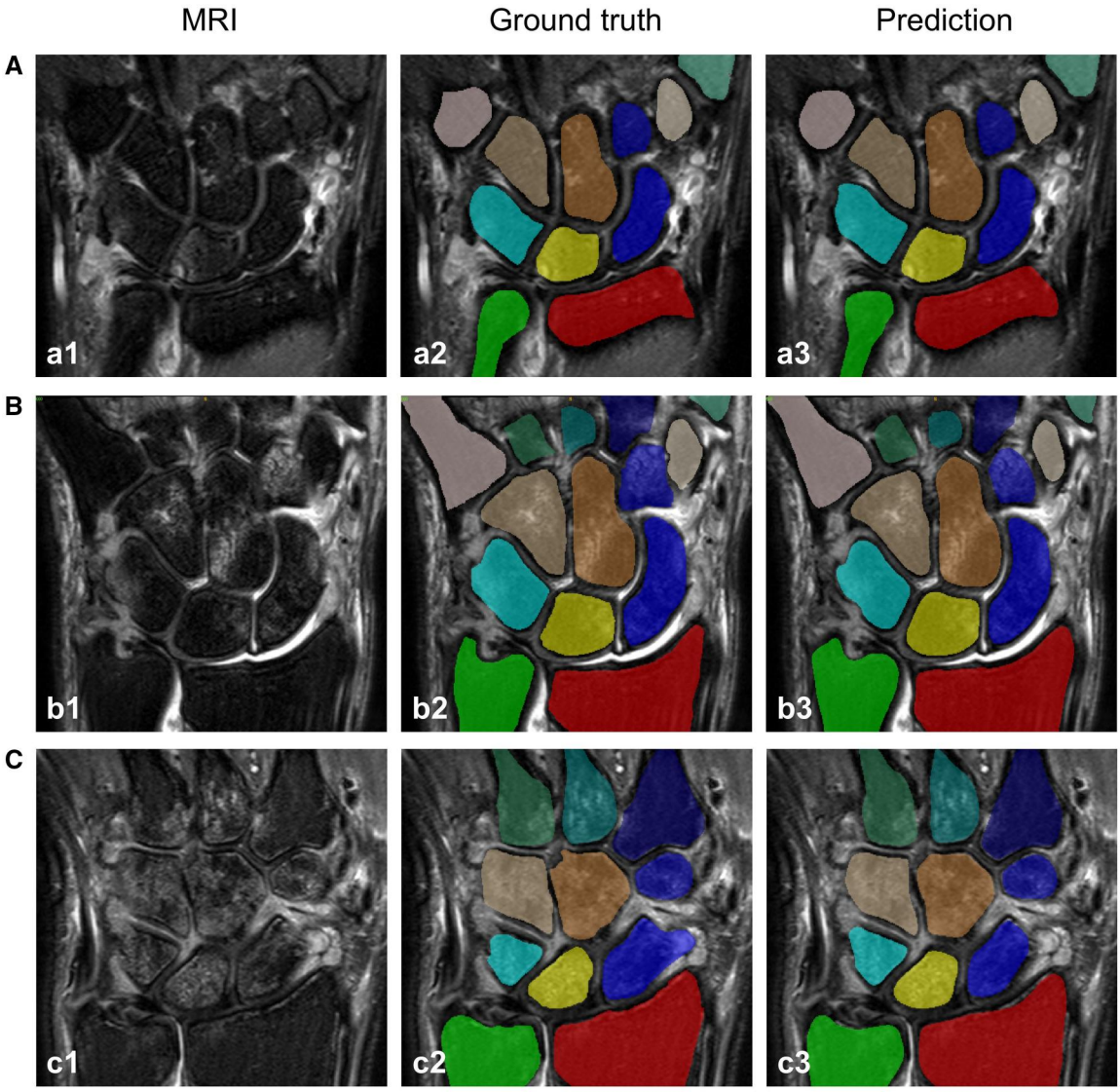

Figure 2. Segmentation results of a (**A**) mild (RAMRIS BME = 3), (**B**) moderate (RAMRIS BME = 11) and (**C**) severe (RAMRIS BME = 25) BME case. The left column (a1–c1) shows the original T2-weighted fat-saturated coronal image. The middle column (a2–c2) shows the manually segmented model (ground truth). The right column (a3–c3) shows the predicted model following automated wrist bone segmentation. The same bone areas in the ground truth and predictive columns are colour coded

were used to train the model. After the model was fully trained, T2-weighted fat-suppressed coronal image datasets were uploaded for automated quantification of BME proportion (%) and intensity.

Quantification

Visual scoring

Visual scoring of BME extent was undertaken by a musculoskeletal radiologist using the RAMRIS BME subscore. Fifteen wrist bone areas were scored on a 0–3 scale according to the estimated proportion of bone area affected by BME, with 0 being no BME, 1 being <33% bone area affected, 2 being 34–66% affected and 3 being >66% affected. Summing scores for all 15 bone areas yielded an overall BME score (maximum score of 45).

Automated scoring of BME

The bone area proportion affected by BME and the BME intensity were evaluated. After wrist bone area segmentation, a non-Gaussian intensity histogram of the overall bone was automatically obtained (Fig. 3). The mean and S.D. of BME intensity were calculated and the BME proportion of oedematous wrist bone was automatically calculated. Initially a Gaussian mixture model (GMM) clustered the histogram into two groups: normal bone and non-normal bone (BME and cystic spaces). Signal intensity of the normal bone cluster served as the internal reference. BME intensity was then determined using a two-way thresholding approach using the automatically defined threshold for normal bone to distinguish it from BME:

$$normal\ threshold = \mu + 1.96\sigma, \tag{1}$$

where μ is the mean and σ the S.D. of the normal bone cluster intensity; 1.96 yielded a 95% confidence level. BME was defined as any pixel with a signal intensity above the normal threshold.

A higher threshold was manually set to distinguish BME from bone marrow cystic areas, i.e. to exclude cystic areas within bone. This threshold was manually adjusted until the cystic areas were virtually removed from the segmented image. If no cystic areas were present, there was no need to adjust the upper threshold. After obtaining the histogram along with the lower and upper thresholds, BME proportion was calculated for both the overall wrist bone areas and individual bone areas using the following equation:

$$BME\ proportion\ (\%) = \frac{BME\ volume}{Bone\ volume} \times 100\%. \tag{2}$$

To compare BME intensity, normalization was performed by referencing oedematous bone to non-oedematous bone, as MR signal intensities are relative, preventing direct absolute comparison:

$$I\prime = \frac{I - \mu}{\sigma}, \tag{3}$$

where $I\prime$ is the normalized intensity, I is the non-normalized intensity and μ and σ are the mean and S.D., respectively, of the normal bone cluster. After normalization, mean BME intensity could be regarded as the average 'distance' of the BME intensity to the normal bone intensity.

Quantification of BME at baseline and following 1 year

Visual and automated quantification of BME proportion and intensity were compared for 49 patients at baseline and after 1 year of standard RA treatment.

Statistical analysis

SPSS version 10 [13] (IBM, Armonk, NY, USA) was utilized for statistical analysis. The Pearson correlation coefficient (r) was used to evaluate the correlation between visual RAMRIS of BME and BME proportion. Similarly, manual (ground truth) and automated segmentation ('output') was evaluated using the Sørensen–Dice similarity coefficient (DICE) [14] and Jaccard similarity coefficient (JAC) [15]. Error was evaluated by volumetric distance (VD) [16] and global consistency error (GCE) [17]. All four metrics were calculated for the ground truth and the output by case, which are defined as follows for a particular bone label:

$$DICE = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$JAC = \frac{DICE}{2 - DICE} \tag{5}$$

$$VD = \frac{|FN - FP|}{2TP + FP + FN} \tag{6}$$

$$GCE = \frac{1}{n}\min\left[\frac{\frac{FP(FP+2TN)}{TN+FP}\quad\frac{FN(FN+2TP)}{TP+FN}}{\frac{FP(FP+2TP)}{TP+FP}\quad\frac{FN(FN+2TN)}{TN+FN}}\right]. \tag{7}$$

TP, FP, TN and FN represent true-positive, false-positive, true-negative and false-negative statistics, respectively, and n is the number of voxels ($448 \times 448 \times 20$). Quantitative segmentation results were evaluated for overall wrist bone rather than individual bone areas, as the clinical relevance of BME relates to the whole wrist. Wilcoxon signed-rank tests for overall bone RAMRIS BME score, BME proportion and mean BME intensity pre-treatment and post-treatment were evaluated. The Pearson correlation coefficient between ΔRAMRIS BME (change in the RAMRIS BME pre- and post-treatment), ΔBME proportion [change in the BME proportion pre- and post-treatment] and Δmean BME relative intensity (change in BME relative intensity pre- and post-treatment) was calculated. P-values <0.05 were considered statistically significant.

Reliability and minimal detectable difference

The intraclass correlation coefficient (ICC) [18] for ΔBME proportion and Δmean BME relative intensity for 15 randomly selected patients between two observers were calculated. Baseline and 1-year examinations were assessed for each patient. Each T2-weighted fat-saturated coronal dataset was uploaded to the model and the upper threshold manually adjusted. We used the ICC [1, 2] (two-way random, single measures, absolute agreement) to assess reliability. The smallest detectable difference (SDD) [19, 20] was computed using the following formula:

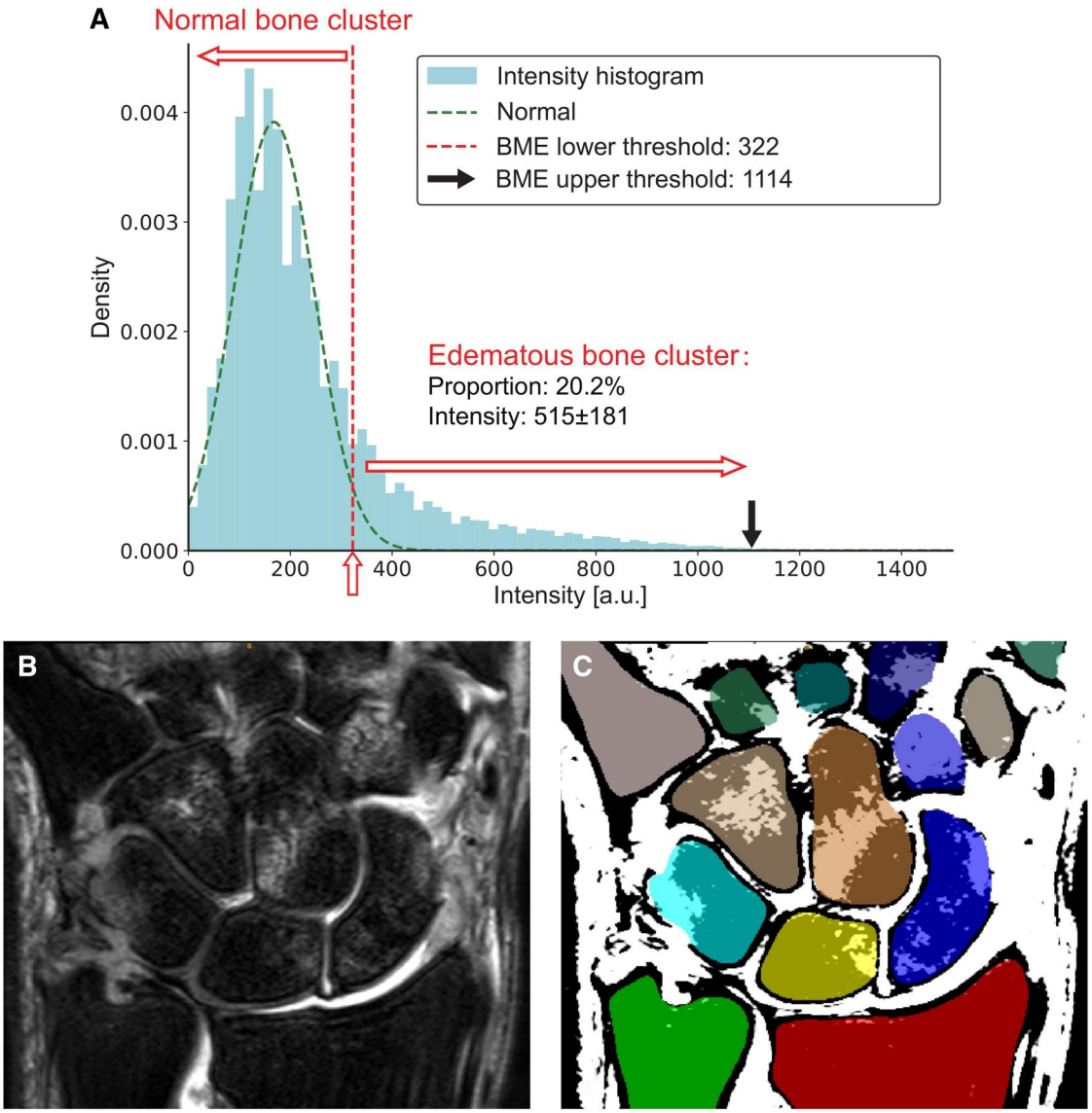

Figure 3. Histogram of BME with thresholds. (**A**) Intensity histogram of overall bone and the GMM clustering result of a single representative case. α, μ and σ are the proportion, mean and S.D. of the cluster, respectively. The lower BME intensity threshold is obtained from the normal bone cluster using the formula $\mu + \sigma * 1.96$. The upper BME intensity threshold was 1114. (**B**) T2-weighted fat-suppressed coronal MRI. (**C**) Same binary MRI with predicted labels after thresholding with upper and lower BME intensity thresholds

$$SDD = 2 \times \left(SD\left[obs^{b}_{post} - obs^{b}_{pre}\right] - \left[obs^{a}_{post} - obs^{a}_{pre}\right] \right), \quad (8)$$

where SD is the standard deviation, obs^{a} is the score at the first observer reading and obs^{b} is the score at second observer reading.

Results

Segmentation (manual and automated)

Examples of manual and automated segmentation are presented in Fig. 2, depicting three cases with mild, moderate and severe BME. The automated model consistently identified individual wrist bone areas, even when these bone areas were oedematous, and was not impeded by low T2 signal intensity tendons, high T2 signal joint fluid or synovitis.

Segmentation accuracy (DICE) for the overall wrist bone was 0.91 (S.D. 0.02), while that for the individual wrist bone areas is shown in Supplementary Fig. S1 and Supplementary Table S2, available at *Rheumatology Advances in Practice* online. The pisiform, trapezoid and fifth metacarpal base had lower segmentation reliabilities than other bone areas. The correlation between RAMRIS BME and overall wrist bone DICE ($r = -0.644$, $P < 0.001$, $n = 60$) (Supplementary Fig. S1, available at *Rheumatology Advances in Practice* online) illustrates how more severe degrees of BME tend to slightly compromise segmentation accuracy.

Quantification (visual and automated)

Fig. 3A shows an intensity histogram of overall wrist bone areas and the GMM clustering results for a single case with moderate BME and a RAMRIS BME score of 11 (same case as Fig. 2B). BME proportion for this case was 20.2%

Table 1. Example of RAMRIS BME, automated BME proportion and intensity in a single patient

Bone area	RAMRIS BME score	Quantitative score	
		BME proportion (%)	BME intensity[a], mean (S.D.)
Radius	0	7.2	5.02 (0.90)
Ulna	1	1.3	4.95 (0.90)
Scaphoid	1	23.9	6.05 (1.77)
Lunate	1	19.9	5.76 (1.60)
Triquetrum	1	16.2	5.77 (1.28)
Pisiform	0	14.3	5.05 (0.85)
Trapezium	1	13.6	5.72 (1.55)
Trapezoid	1	85.0	7.99 (2.54)
Capitate	1	25.2	6.16 (2.07)
Hamate	1	45.9	6.82 (2.58)
BOM1	0	0.5	4.89 (0.70)
BOM2	1	77.7	8.60 (3.05)
BOM3	1	37.7	6.52 (2.81)
BOM4	1	25.4	6.69 (2.43)
BOM5	0	0.3	4.48 (0.28)
Overall	11	20.2	4.40 (2.29)

For a single case of moderate-severity BME (RAMRIS BME = 11) (the same case as in Fig 2b). Comparison of visual RAMRIS BME with automatic quantification of BME proportion and relative intensity for individual wrist bone areas and overall wrist bone areas.
BOM: base of metacarpal.
[a] Normalized by reference to normal bone cluster.

representing the percentage of voxels in the segmented bone areas occupied by BME. The normalized intensity of this BME was calculated as 4.40 (S.D. 2.29) (Table 1). Fig. 3B shows the corresponding MRI and Fig. 3C shows the binary MRI with predicted labels after thresholding. The RAMRIS BME scores for each bone area and overall bone area are compared with automated quantitative BME scores for bone marrow proportion and BME relative intensity in Table 1.

In 60 validation cases, strong correlation ($r = 0.928$, $P < 0.001$) between the RAMRIS BME score and automated BME proportion was found, with moderate correlation ($r = 0.677$, $P < 0.001$) between RAMRIS BME score and automated BME relative intensity (Fig. 4A, B). Strong correlation ($r = 0.752$, $P < 0.001$) between BME proportion and BME relative intensity was found (Fig. 4C).

BME quantification at baseline and after 1 year of treatment

With treatment in 49 patients, the overall RAMRIS BME score decreased from 6.60 (S.D. 7.36) to 3.41 (S.D. 5.88) ($P = 0.011$), BME proportion decreased from 9.6 (S.D. 6.6) to 6.2 (S.D. 5.88) ($P < 0.001$) and BME relative intensity decreased from 3.43 (S.D. 0.57) to 3.29 (S.D. 0.51) ($P = 0.049$) (Table 2). A total of 19 (39%) of the 49 patients showed no treatment change with RAMRIS BME score, while all 49 patients showed some change in BME proportion and BME relative intensity (Table 2, Supplementary Fig. S2, available at *Rheumatology Advances in Practice* online). Strong ($r = 0.852$, $P < 0.001$) correlation was found between pre- and post-treatment changes in visual RAMRIS BME score and automated BME proportion, with moderate correlation ($r = 0.428$, $P = 0.002$) between changes in visual RAMRIS BME scores and automated BME relative intensity. There was modest ($r = 0.618$, $P < 0.001$) correlation between pre- and post-treatment changes in BME proportion and BME relative intensity.

Reliability and minimal detectable difference

There was high reliability between both observers, for pre- and post-treatment changes in BME proportion (ICC 0.980, $P < 0.001$) and for pre- and post-treatment relative changes in BME mean intensity (ICC 0.735, $P < 0.001$) (Supplementary Table S3, available at *Rheumatology Advances in Practice* online). The smallest detectable difference (SDD) of pre- and post-treatment changes in BME proportion and BME mean intensity was 2.04% and 0.82, respectively (Supplementary Table S3, available at *Rheumatology Advances in Practice* online).

Discussion

BME is increased fluid content within the bone marrow [21]. It is seen as high signal intensity on T2-weighted fat-suppressed sequences and, less conspicuously, as low signal intensity on T1-weighted images [22]. In RA, BME is usually the result of bone inflammation (osteitis), being histologically associated with a patchy lymphoplasmacytic marrow infiltrate similar to that seen in inflamed synovial tissue [23]. BME occurs in two-thirds of ERA patients and is nearly always accompanied by synovitis or tenosynovitis [17, 18]. BME is the strongest predictor of bone erosion [24], associated with a 6.5-fold site-specific higher erosion risk at 6 years [25]. The likelihood of developing bone erosion in non-oedematous bone is very low [26]. As BME is reversible and bone erosions are generally irreversible, one should aim to treat BME before irreversible structural damage ensues [27]. Although a reduction in BME is usually accompanied by reduced synovial proliferation [28], BME reduction may be an earlier indicator of treatment response than synovitis reduction [27].

Semi-quantitative RAMRIS BME scoring [22, 29] is a most commonly used method to quantify BME severity. However, RAMRIS assessment is limited in detecting subtle change, requires assessment by experienced observers, has modest interreader reliability, does not quantify BME intensity and is relatively time-consuming, with each assessment taking at least 5 min. Several studies have attempted to address these limitations. Roemer *et al.* estimated BME volume by drawing three orthogonal lines spanning the lesion's margins [30]. Li *et al.* used a threshold-based method, relying on a region of interest in normal bone to calculate a threshold to generate BME voxels [31, 32]. Although this method had high reproducibility and correlated well with the RAMRIS BME score, it lacked automation and was relatively time-consuming.

The highly automated pipeline present in this study is time efficient, with the whole process completed in 1 s and the full quantification result displayed in real time. The only step occasionally requiring manual interaction is setting the upper threshold of BME intensity by placing a marker in one cystic area when cystic bone change is present. No other intervention is needed. As each bone area following segmentation is analysed on a pixel-by-pixel basis, areas of mild, moderate and severe BME can be handled satisfactorily by the proposed pipeline. Segmentation is the most critical and technically difficult part of the pipeline. Accurate segmentation is necessary before BME quantification to avoid erroneously including joint fluid, inflamed synovium and soft tissue oedema as BME. Wrist bones are not easily segmented, as they are small, variably sized, with narrow intervening spaces and aligned curvilinearly in both transverse and sagittal planes.

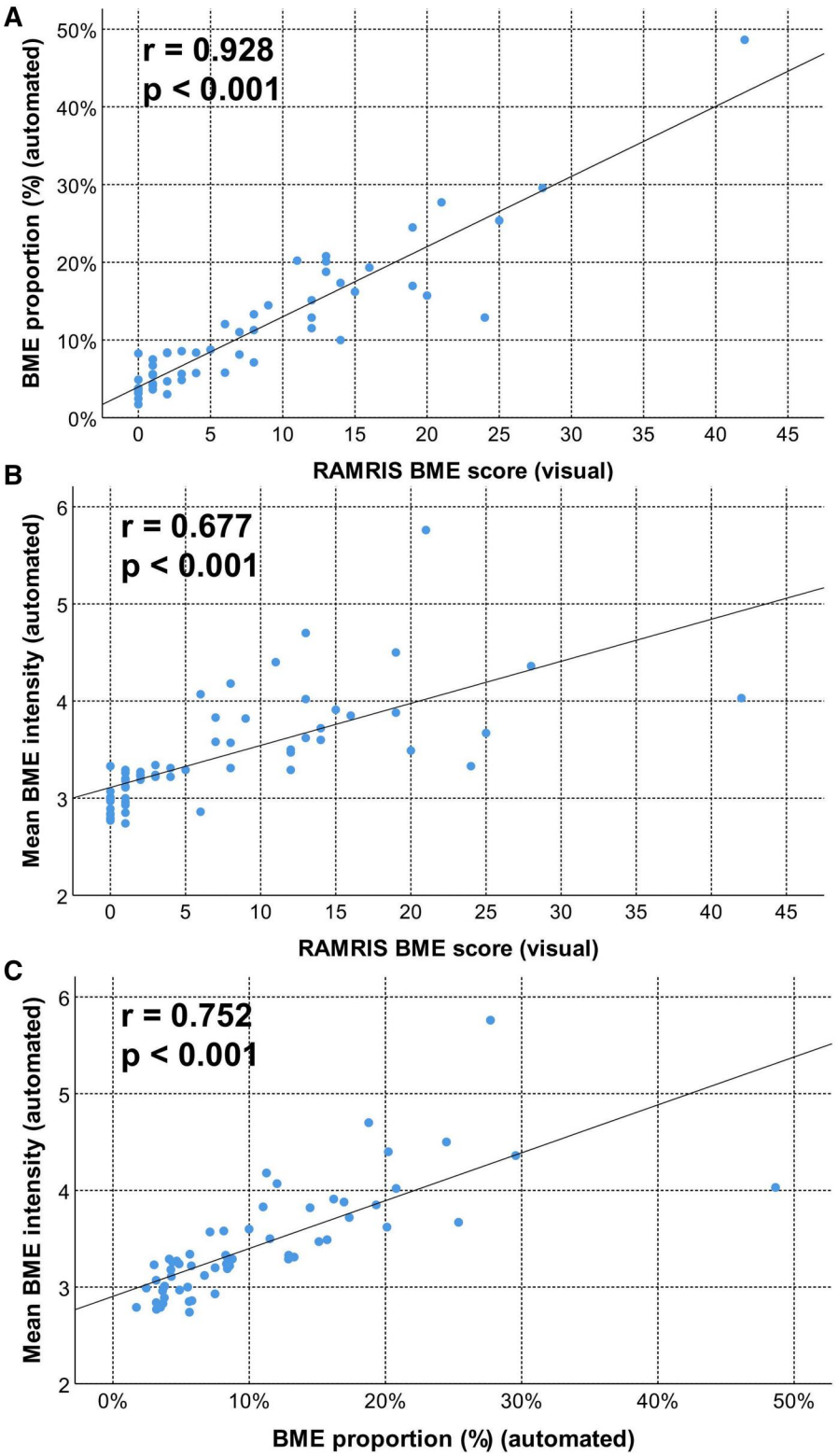

Figure 4. BME quantification for overall wrist bone (i.e. all wrist bone areas considered together) in 60 validation cases. (**A**) Scatter plot showing strong correlation ($r = 0.928$, $P < 0.001$) for overall BME proportion and RAMRIS BME. In 12 (20%) of the 60 cases, RAMRIS BME was 0 but BME proportion provided a low value for BME. (**B**) Scatterplot showing moderate ($r = 0.677$, $P = 0.033$) correlation between mean BME intensity and RAMRIS BME. (**C**) Scatterplot showing high ($r = 0.752$, $P < 0.001$) correlation between mean BME intensity and BME proportion

Although the wrist bones are more easily demarcated on T1 spin-echo rather than T2-weighted fat-suppressed images, it is preferrable to segment T2-weighted fat-suppressed images, as these better depict BME. Several previous studies have undertaken wrist bone segmentation on T1-weighted spin echo sequences using atlas-based [33, 34], seed-based [35, 36] or machine learning–based [37] algorithms. To address issues related to less distinct boundary marking on

Table 2. Ranking of treatment response for RAMRIS BME, BME proportion and relative intensity

Quantification metric	Descriptive statistics, mean (S.D.)		Wilcoxon signed-rank test				
	Pre-treatment	Post-treatment	Negative ranks[b]	Positive ranks[c]	Ties[d]	Total	*P*-value
RAMRIS BME	6.06 (7.36)	3.41 (5.88)	21	9	19	49	0.011
BME proportion	9.6 (6.6)	6.2 (5.4)	34	15	0	49	0.000
BME intensity[a]	3.43 (0.57)	3.29 (0.51)	32	17	0	49	0.049

Wilcoxon signed-rank test for RAMRIS BME, BME proportion and relative BME intensity of all wrist bone areas in 49 patients pre- and post-treatment.
[a] Normalized by reference normal bone cluster.
[b] Quantification metric post-treatment < pre-treatment.
[c] Quantification metric post-treatment > pre-treatment.
[d] Quantification metric post-treatment = pre-treatment.

T2-weighted fat-suppressed images, especially when the bones were oedematous, we applied a deep learning algorithm, including convolutional neural network–based segmentation algorithms. Building on our previous work [38], an nnU-Net framework [11] was implemented in this study to good effect.

Compared with manual segmentation, the proposed pipeline was very accurate (DICE 0.91) at segmenting individual wrist bone areas on T2-weighted fat-suppressed MRI. The proposed algorithm enables the proportion of the overall wrist bone area affected by BME to be provided as a percentage, as well as the average relative intensity of this BME. This is the first algorithm able to measure these features. Automated BME proportion had a strong ($r = 0.928$, $P < 0.001$) correlation with visual RAMRIS BME scoring. BME intensity had a strong ($r = 0.752$, $P < 0.001$) correlation with BME proportion. As well as considering all the wrist bone areas as a single unit, the pipeline also enables quantification of the BME proportion and relative intensity of individual wrist bone areas, enabling areas of incomplete fat suppression to be excluded and, if necessary, the relationship between BME in specific bone areas and subsequent erosion to be readily evaluated. While RAMRIS takes at least 5 min to complete and is semi-quantitative and manual bone segmentation of wrist bone areas with quantification takes at least 1 h, automated BME quantification following upload of T2-weighted fat-suppressed images can be completed in 1 s.

Automated quantification of BME also seemed more sensitive at detecting treatment change than the RAMRIS BME. Treatment effect was analysed in 49 ERA patients who underwent follow-up MRI after 1 year of standard RA treatment. Following treatment, the proportion of bone affected by BME decreased by about one-third, from 9.6% (S.D. 6.6) to 6.2% (S.D. 5.4), while BME intensity decreased by a lesser degree. Using RAMRIS BME scoring, two-fifths of patients showed no treatment change, while all patients showed treatment change of both BME proportion and BME relative intensity, with about three-quarters of patients showing a reduction in these parameters.

There are some limitations to this study. First, only a limited image dataset of 60 patients was used for training. Due to motion artefact and inadequate fat suppression, one-quarter of image datasets from the original cohort were unsuitable for accurate automated segmentation. In other words, high-quality MRI data are currently necessary to apply this method, which needs to be made more robust to deal with motion artefacts and inadequate fat suppression. Second, this was a feasibility study with no independent external testing of the proposed method being undertaken. This automated assessment method needs to be tested on datasets from different MRI systems to ensure its widespread applicability, including its applicability on lower-resolution MR acquisitions, alternative acquisition protocols and low-field MRI systems. While BME proportion was measured both manually and automatically, BME intensity was only measured automatically, as no accepted method of manually measuring relative BME intensity exists. While the automated system provided high reliability with a small minimal detectable change for BME proportion, reliability was only moderate for BME intensity, possibly related to non-uniformity of the internal reference standard.

In conclusion, we present a reliable method of automatically quantifying BME proportion and intensity in ERA patients based on T2-weighted fat-suppressed coronal MRI sequences. Automated BME quantification can be completed in 1 s, had a high correlation with RAMRIS BME scoring and was sensitive to treatment change.

Supplementary material

Supplementary material is available at *Rheumatology Advances in Practice* online.

Data availability

All data relevant to the study are included in the article or uploaded as supplementary information.

Authors' contributions

All authors participated in drafting the article or revised it critically for important intellectual content. All authors approved the final version to be published. C.Y. had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. J.F.G. and L.T. were responsible for the conception and design of the study. F.X. and L.S were responsible for data acquisition. B.Z, and W.S were responsible for the analysis and interpretation of the data.

Funding

This work was supported by the Health Services Research Fund (grant project 6906217).

Disclosure statement: The authors have declared no conflicts of interest.

Acknowledgements

This research was conducted in part at Chinese University of Hong Kong Department of Imaging and Interventional Radiology MRI facility, which is jointly funded by Kai Chong Tong, Hong Kong Special Administrative Region Research Matching Grant Scheme and the Department of Imaging and Interventional Radiology, Chinese University of Hong Kong. It was conducted with the aid of the Health Services Research Fund (grant project 6906217).

References

1. McQueen FM, Benton N, Perry D *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. Ostergaard M, Møller Døhn U, Duer-Jensen A *et al.* Patterns of magnetic resonance imaging bone erosion in rheumatoid arthritis—which bones are most frequently involved and show the most change? J Rheumatol 2011;38:2014–7.
3. Rubin DA. MRI and ultrasound of the hands and wrists in rheumatoid arthritis. I. Imaging findings. Skeletal Radiol 2019; 48:677–95.
4. McQueen FM, Stewart N, Crabbe J *et al.* Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset. Ann Rheum Dis 1998;57:350–6.
5. Ejbjerg B, Narvestad E, Rostrup E *et al.* Magnetic resonance imaging of wrist and finger joints in healthy subjects occasionally shows changes resembling erosions and synovitis as seen in rheumatoid arthritis. Arthritis Rheum 2004;50:1097–106.
6. Dakkak YJ, Boer AC, Boeters DM. MRI-detected osteitis is not associated with the presence or level of ACPA alone, but with the combined presence of ACPA and RF. Arthritis Res Ther 2016;22:67.
7. Narváez JA, Narváez J, De Lama E, De Albert M. MR imaging of early rheumatoid arthritis. Radiographics 2010;30:143–63; discussion 163–5.
8. Tamai M, Kawakami A, Uetani M *et al.* The presence of anticyclic citrullinated peptide antibody is associated with magnetic resonance imaging detection of bone marrow oedema in early stage rheumatoid arthritis. Ann Rheum Dis 2006;65:133–4.
9. Crowley AR, Dong J, McHaffie A *et al.* Measuring bone erosion and edema in rheumatoid arthritis: a comparison of manual segmentation and RAMRIS methods. J Magn Reson Imaging 2011; 33:364–71.
10. Yushkevich PA, Piven J, Hazlett HC *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116–28.
11. Isensee F, Jäger PF, Kohl SAA, Petersen J, Maier-Hein KH. Automated design of deep learning methods for biomedical image segmentation. arXiv:1904.08128v2.
12. Paszke A, Gross S, Chintala S *et al.* Automatic Differentiation in PyTorch [Internet]. In: NIPS 2017 Workshop on Autodiff. https://openreview.net/forum?id=BJJsrmfCZ.
13. IBM. IBM SPSS Statistics for Windows, version 21.0. Armonk, NY: IBM, 2012.
14. Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.
15. Crum WR, Camara O, Hill DLG. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans Med Imaging 2006;25:1451–61.
16. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015;15:29.
17. Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proc IEEE Int Conf Comput Vis 2001;2:416–23.
18. Landers R. Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. Authorea. June 24, 2015.
19. Lassere M, Boers M, van der Heijde D *et al.* Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731–9.
20. Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Østergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: a comparison of the OMERACT rheumatoid arthritis magnetic resonance imaging score applied to different joint combinations and the Sharp/van der Heijde radiographic score. Arthritis Rheum 2005;52:2300–6.
21. Gorbachova T, Amber I, Beckmann NM *et al.* Nomenclature of subchondral nonneoplastic bone lesions. Am J Roentgenol 2019; 213:963–82.
22. Østergaard M, Peterfy C, Conaghan P *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.
23. McQueen FM. Bone marrow edema and osteitis in rheumatoid arthritis: the imaging perspective. Arthritis Res Ther 2012;14:224.
24. Hetland ML, Ejbjerg B, Hørslev-Petersen K *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.
25. McQueen FM, Benton N, Perry D *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.
26. Mundwiler ML, Maranian P, Brown DH *et al.* The utility of MRI in predicting radiographic erosions in the metatarsophalangeal joints of the rheumatoid foot: a prospective longitudinal cohort study. Arthritis Res Ther 2009;11:R94–10.
27. Sudoł-Szopińska I, Kontny E, Maśliński W *et al.* Significance of bone marrow edema in pathogenesis of rheumatoid arthritis. Pol J Radiol 2013;78:57–63.
28. Conaghan PG, Emery P, Østergaard M *et al.* Assessment by MRI of inflammation and damage in rheumatoid arthritis patients with methotrexate inadequate response receiving golimumab: Results of the GO-FORWARD trial. Ann Rheum Dis 2011;70:1968–74.
29. Østergaard M, Peterfy CG, Bird P *et al.* The OMERACT rheumatoid arthritis magnetic resonance imaging (MRI) scoring system: updated recommendations by the OMERACT MRI in arthritis working group. J Rheumatol 2017;44:1706–12.
30. Roemer FW, Bohndorf K. Long-term osseous sequelae after acute trauma of the knee joint evaluated by MRI. Skeletal Radiol 2002; 31:615–23.
31. Li X, Ma BC, Bolbos RI *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.
32. Li X, Yu A, Virayavanich W *et al.* Quantitative characterization of bone marrow edema pattern in rheumatoid arthritis using 3 Tesla MRI. J Magn Reson Imaging 2012;35:211–7.
33. Włodarczyk J, Czaplicka K, Tabor Z, Wojciechowski W, Urbanik A. Segmentation of bones in magnetic resonance images of the wrist. Int J Comput Assist Radiol Surg 2015;10:419–31.
34. Gemme L, Nardotto S, Dellepiane SG. Automatic MPST-cut for segmentation of carpal bones from MR volumes. Comput Biol Med 2017;87:335–46.
35. Conte D, Foggia P, Tufano F, Vento M. An enhanced level set algorithm for wrist bone segmentation [Internet]. Image Segmentation. InTech 2011. http://dx.doi.org/10.5772/15437.

36. Włodarczyk J, Wojciechowski W, Czaplicka K, Urbanik A, Tabor Z. Fast automated segmentation of wrist bones in magnetic resonance images. Comput Biol Med 2015;65:44–53.
37. Koch M, Schwing AG, Comaniciu D, Pollefeys M. Fully automatic segmentation of wrist bones for arthritis patients. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA 2011:636–40.
38. Wong LM, Shi L, Xiao F, Griffith JF. Fully automated segmentation of wrist bones on T2-weighted fat-suppressed MR images in early rheumatoid arthritis. Quant Imaging Med Surg 2019;9:579–89.

Rheumatology Advances in Practice, 2024, 8, 1–10
https://doi.org/10.1093/rap/rkae073
Original Article