DELFT UNIVERSITY OF TECHNOLOGY

Early Detection of Rheumatoid Arthritis using extremity MRI: Quantification of Bone Marrow Edema in the Carpal bones
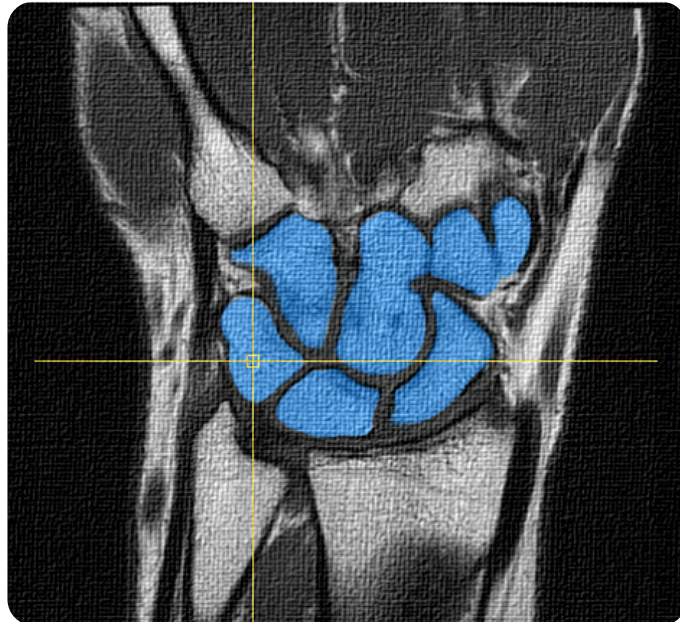
THESIS

submitted in partial fulfilment of the
requirements for the degree of

MASTER OF SCIENCE

in

BIOMEDICAL ENGINEERING

March 20, 2015

Edgar Alphonse Hendrik Roex
4253450

Faculteit Werktuigbouwkunde, Maritieme Techniek
& Technische Materiaalwetenschappen
Delft University of Technology
E.A.H.Roex@student.tudelft.nl

ii

EXAMINATION COMMITTEE MEMBERS:

Prof. Dr. Ir. B.P.F. Lelieveldt [a,b]	Chair, Supervisor
Dr. B. C. Stoel [b]	Daily Supervisor
Dr. Ir. Bart L. Kaptein [c]	External Member
Prof. Dr. M. Loog [a]	External Member

a   Delft University of Technology, Faculty of Electrical Engineering, Mathematics & Computer Science, Delft, The Netherlands
b   Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands
c   Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands

Abstract — Visual scoring of magnetic resonance images for the early detection of rheumatoid arthritis is prone to human subjectivity and lacks sensitivity. In a bid to develop an objective and quantitative alternative using digital image processing, this thesis proposes automatic segmentation of the carpal bones, followed by the quantification of bone marrow edema, which is an important inflammatory imaging biomarker. Segmentation of the carpal bones is achieved using multi atlas-based segmentation. Compared to manual segmentations of the training data, an average Dice overlap of 0.85 was achieved. By examining contrast-enhanced MR images of the wrist, edematous bone is classified from normal bone marrow using knowledge based fuzzy clustering. Validation of the quantitative score against the existing RA MRI Scoring (RAMRIS) system showed a significant positive correlation. Segmentation error was seen to be a confounding factor, limiting the specificity of the BME measure. To increase agreement and maximise the available information, it is recommended that data from a complementary imaging plane is included.

Acknowledgements

First and foremost, I would like to thank Boudewijn for the opportunity of completing my MSc thesis at LKEB and overseeing the project. Berend, I have been very fortunate to have you as my supervisor. Your availability, guidance and well-placed critique have undoubtedly lifted the quality of my work. I am grateful for the many lessons in good research practice you have given me along the way.

Evgeni, cheers for all your input and our many in-depth discussions. Working together on the early RA project made it all the more productive and importantly, enjoyable. I would like to thank Denis for all the technical assistance, and the registration gurus, Marius and Floris, for their expertise and of course everyone else at LKEB that offered me their time.

My thanks go to Annette, Lukas, Wouter, and Hanna from rheumatology and Monique from radiology for their collaboration and clinical expertise. I would also like to acknowledge all the patients and volunteers than underwent an MRI scan, and those who scored the data.

I thank all my friends and fellow students, especially the more than 20 different housemates I got to live with; you guys made my time as a (nomadic) student in Delft unforgettable.

Lastly, to my family and loved ones, near and far, thank you for your encouragement and support.

Abbreviations and acronyms

ABS	Atlas-Based Segmentation
BME	Bone Marrow Edema
CR	Conventional Radiography
CT	Computed Tomography
DCE-MRI	Dynamic Contrast Enhanced MRI
EULAR	European League Against Rheumatism
FOV	Field of View
(N)MI	(Normalised) Mutual Information
MCP	Metacarpophalangeal (joints)
MRI	Magnetic Resonance Imaging
OMERACT	Outcome Measures in Rheumatoid Arthritis Clinical Trials
RA	Rheumatoid Arthritis
RAMRIS	Rheumatoid Arthritis MRI Scoring system
ROI	Region of Interest
US	Ultrasonography

Glossary

Articular	Of or relating to a joint or joints
Arthralgia	Joint pain
Edema	Swelling caused by accumulation of fluid
Osteitis	Inflammation of bony tissue
Prognosis	A prediction of the probable course and outcome of a disease
Registration	The process of matching or aligning of one image to another
Segmentation	The process of subdividing an image into its constituent objects or regions of interest
Synovitis	Inflammation of the synovium
Synovium	A soft-tissue membrane that lines the joint capsule of synovial joints
Voxel	A volume element, the 3D analogy of a 'pixel'

x

Contents

1. Introduction

1.1 Clinical background

Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a chronic autoimmune disease, typified by inflammation of synovial joints. These are joints lined with synovium, a specialised tissue responsible for maintaining the nutrition and lubrication of the joint. The exact pathogenesis is unknown, making RA an active area of medical research. Approximately one percent of the world population is affected, with a relatively higher incidence in the elderly and women. Tell-tale signs of RA are (morning) stiffness, pain, swelling and fatigue. Although symptoms are systemic, the small joints of the hands, wrists and feet are those earliest and most frequently involved [1]. Pathophysiological features include (teno)synovitis, bone marrow edema (BME) and bone erosions. Synovitis is the inflammation of the joint lining or tendon sheaths due to the accumulation of fluid. Similarly BME describes the build-up of liquid inside the bone marrow cavity. Without intervention, erosion and destruction of the joints eventually ensues, resulting in deformities and loss of function. On top of this, the pain, discomfort and physical limitations place an emotional burden on patients, further diminishing their quality of life. From a socioeconomic standpoint RA places large direct and indirect costs on healthcare and society as a whole. This includes the high cost of treatment, as well as costs incurred due to sick leave, disability pensions and occupational therapy. In the Netherlands, as many as one in eight workers with RA is declared unfit for work [2]. A Dutch cohort study records 42% of RA patients as being work disabled after three years [3]. Fortunately, since the arrival of specialised new drugs known as biologicals, destructive RA may become a thing of the past. This treatment comes at a price however; estimates place the costs for biologicals in 2010 at 450 million euro in The Netherlands and 90 billion dollars worldwide [4]. In the same year in The Netherlands, biologicals ranked number one and two in terms of total expenditure and growth thereof [5]. The overall impact of RA emphasises the need to shorten disease life.

Treatment of RA

Since the root cause of RA is yet to be discovered, there is no outright cure. In recent years however, rheumatology has gained new insights into the complex biological pathways underlying the disease. There is growing evidence to support the concept of a "window of opportunity" early in the disease phase, when aggressive treatment can be seen to improve patient outcomes in the long term [6]–[10]. Said treatment involves the prescription of biologicals, a type of disease-modifying antirheumatic drug (DMARD). These agents inhibit the cell signalling protein called tumour necrosis factor-alpha (TNFα) responsible for provoking an inflammatory immune response. As a result, the progression of RA may be suppressed, preventing irreversible joint damage and subsequent disability. Achieving this state, known as remission, is the therapeutic goal in RA and has become a realistic outcome for a substantial portion of patients (Figure 1.1). Given the aforementioned "window of opportunity" rheumatologists focus on what is known as *early RA* or even *very early RA* and try to identify individuals who are likely to have a persistent and destructive disease course. A

treat-all approach is not viable due to the aggressive nature and high costs associated with the medication. Though the literature is not consistent in its definition of *early RA*, most authors use the term to mean disease duration of less than one year. Rheumatologists in the Leiden University Medical Center (LUMC) define *very early RA* as the first three months after symptoms first arise. In order to select those who are to receive treatment with biological DMARDs, screening is performed in the preclinical phase of arthritis. However current screening for (very) early RA lacks sensitivity, calling for novel clinical and radiological indicators.
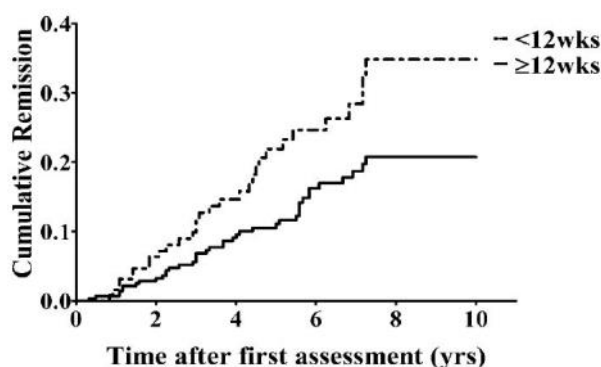
Figure 1.1: Greater cumulative remission is shown for patients who had their first assessment within 12 weeks of symptom onset compared to those who delayed. Source [11].

1.2 RA and diagnostic imaging

Current RA diagnosis depends on the aggregation of a range of clinical criteria, including physical examination results, laboratory data and radiologic findings [7]. For decades, the mainstay modality for assessing joint damage has been conventional radiography (CR). In RA, CR can help monitor disease progression by assessing bone erosions and joint space narrowing [12]. Although CR is an accessible and reproducible form of imaging, it does not capture soft-tissue lesions, such as synovitis and bone marrow edema, which precede joint damage. Moreover, erosions may be seen with other imaging modalities many months before abnormalities appear on radiographs [13]. Consequently, CR is not a reliable prognostic tool for patients with very early RA.

Like CR, ultrasonography (US) is readily available to the clinician and has the advantage of being low-cost and free from ionizing radiation. Used for diagnosis and even symptomatic treatment, it can be used to assess tendon and cartilage damage, synovitis and erosions [12][14]. US is easily repeated, allowing for frequent follow-ups. Unfortunately, not all areas of the joints and surrounding tissue are acoustically accessible by US, meaning imaging is limited to superficial structures. Another major drawback is that the scanning technique is highly operator dependent, leading to reproducibility issues [1].

Known for offering superior soft-tissue contrast, MRI is increasingly being used in the assessment of RA, particularly in its crucial early stages. MRI is more sensitive to bone erosions than CR, and therefore more accurate in monitoring erosive progression [1]. Moreover, MRI is able to detect pre-erosive inflammatory features such as synovitis and bone marrow edema (BME), which act as prognostic markers for RA [15], [16]. In a Norwegian study, MRI measures of inflammation were the most responsive to RA treatment when compared to CR and US [17].

Figure 1.2: The ONI MSK Extreme 1.5T by General Electric.

Extremity MRI

Dedicated extremity MRI (eMRI) scanners have considerably increased the usability of MRI as an imaging tool in RA. Featuring a smaller bore size compared to full-body MR system, the patient can sit or lie beside the scanner inserting only a hand or a foot. In general this provides a more comfortable patient experience. From an economic standpoint, eMRI machines are appealing due to their small size, lower cost and greater patient throughput. The ONI MSK Extreme 1.5T by General Electric (GE) is the eMRI scanner in use at the LUMC (shown in Figure 1.2). Although RA appears bilaterally, in the interest of reducing study time, only the hand and wrist (and corresponding foot) of the dominant or most symptomatic extremity are studied. To highlight inflammatory lesions, a gadolinium-based contrast agent is administered intravenously. Due to a shallow bore, acquisitions of the hand using the ONI eMRI are performed in two parts: one of the wrist and one of the metacarpophalangeal (MCP) joints (see Figure 1.3). The disjoint datasets may be fused using post-processing (image stitching), providing the two field-of-views (FOV) overlap.
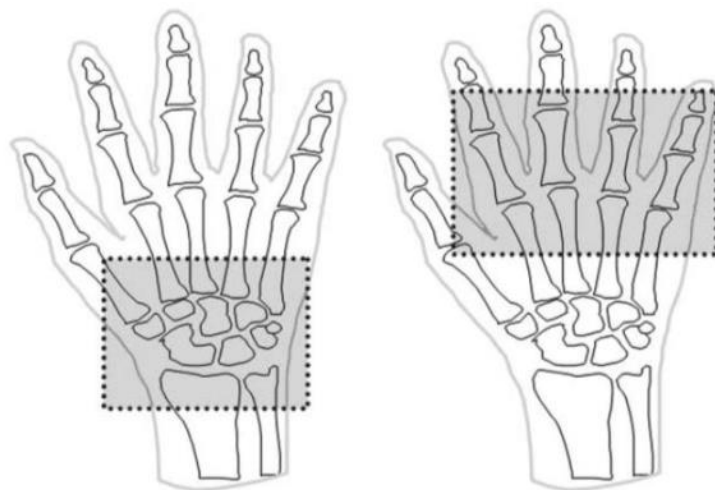
Figure 1.3: Approximate field-of-view (shaded area) for the wrist (left) and MCP joints (right) using eMRI. For image stitching a degree of overlap is required. Source: [7].

Scoring systems

MRI shows great promise in detecting early RA as it captures all relevant structures in arthritic disease and is sensitive to change. However, before it can become an integral part of clinical RA diagnosis, there is a need to standardise and quantify image assessment. A reproducible scoring scheme that captures disease-related activity facilitates comparisons in clinical trials involving early RA patients and healthy controls. In turn, the prognostic utility of imaging biomarkers can be evaluated. The ability to closely monitor disease progression will result in improved treatment decisions.

RAMRIS

Devised and validated by the OMERACT[1] group, the RA MRI Scoring (RAMRIS) system was introduced in 2002. It is a method to semi-quantitatively evaluate abnormalities in the wrist and MCP joints. The intention is to assess both inflammatory and destructive changes. RAMRIS functions as a standard comparator for RA MRI assessment in clinical studies [18]. This has increased the clinical value of MRI, allowing it to become a relevant outcome measure in clinical trials in patients with RA.

RAMRIS stipulates a core set of basic MRI sequences for evaluating RA joints to increase comparability [19]:

- Imaging in two planes with T1 weighted images before and after intravenous gadolinium contrast

---

[1] Outcome Measures in Rheumatoid Arthritis Clinical Trials

- A T2 weighted fat saturated sequence or, if the latter is not available, a STIR (short tau inversion recovery) sequence[2]

where two planes can be reconstructed from a single 3D acquisition with isometrical voxels, or acquired separately by obtaining a two dimensional sequence in two planes. An intravenous gadolinium injection is not critical if assessing bone erosions only.

The important RA joint pathologies are defined as follows (see Figure 1.4 for example images):

- *Synovitis*: an area in the synovial compartment that shows above normal post-gadolinium enhancement of a thickness greater than the width of the normal synovium
- *MRI bone erosion*: a sharply marginated bone lesion, with correct juxta-articular localisation and typical signal characteristics, which is visible in two planes with a cortical break seen in at least one plane
- *MRI bone edema*: A lesion within the trabecular bone, with ill-defined margins and signal characteristics consistent with increased water content
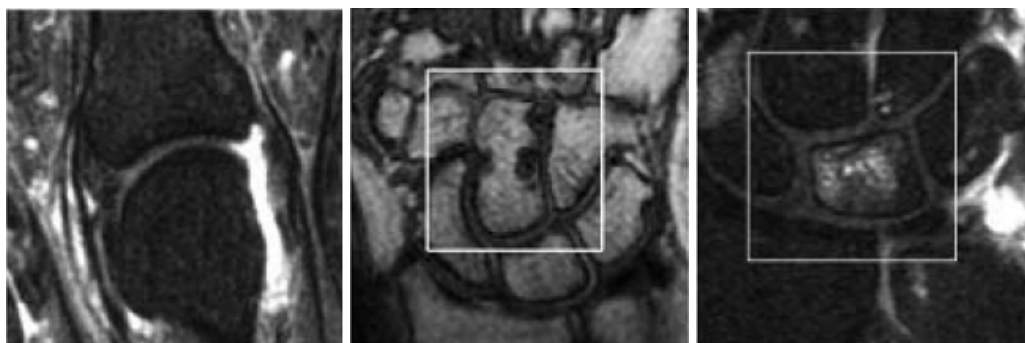
Figure 1.4: Example images of RA MRI features [20]. Left: bright edge indicates synovitis of an MCP joint. Middle: bone erosion (dark spot) of the capitate in the wrist. Right: high signal inside the bone indicates BME of the lunate. Note inflammatory lesions are best seen on contrast-enhanced MRI (left and right), whereas erosive features are best captured by pre-contrast T1 MRI (middle). See Figure 1.5 for bone anatomy of the hand.

These pathologies are scored semi-quantitatively, for specific bones and regions of the hand and wrist, as follows:

- *Synovitis* is scored 0–3. Score 0 is normal, and 1–3 (mild, moderate, severe) are by thirds of the presumed maximum volume of enhancing tissue in the synovial compartment.

---

[2] NB: a T1-weighted fat-suppressed sequence is used instead in this study; see Table 1.2, page 12.

- *Bone erosions* are scored 0–10, according to the proportion (in increments of 10%) of bone involved and compared to the ''assessed bone volume'':
  0: no erosion; 1: 1–10% of bone eroded; 2; 11–20%, ..., 10: 91-100%
- *Bone marrow edema* is scored 0–3 based on the proportion (in increments of 33%) of bone involved: 0: no edema; 1: 1–33%; 2: 34–66%; 3: 67–100%

Limitations of visual scoring

Whilst the introduction of RAMRIS has been a necessary and a welcome one, the system is not free from measurement error and is prone to reliability and reproducibility issues. In a paper on the most important pitfalls encountered in RA MRI scoring, McQueen et al. attribute the principal sources of error or confusion to imaging artefacts, the misinterpretation of normal features, the influence of slice thickness/field of view and the distortion of anatomy due to disease [21]. Furthermore, lesions frequently coexist, making them hard to classify.

The sensitivity of visual scoring, as done in RAMRIS, is inherently limited by the human visual system. Subtle changes in grey values are difficult to detect and depend on lighting conditions, display settings and other variables. Although studies have shown inter- and intra-observer agreement for RAMRIS to be "satisfactory" or "reasonable", they remain an issue and may depend on the relative level of MRI reading experience [18], [22]. The aforementioned issues undermine the reliability and reproducibility of RAMRIS.

Various sources of experimenter bias also undermine score reliability. An example of procedural bias is whereby baseline and follow-up data are viewed and scored concurrently, allowing the reader to (unknowingly) change his or her expectation of disease progression.

Practical issues include the time and cost to train and calibrate readers. Visual image scoring itself is also a tedious and time-consuming task, forcing readers to stop scoring when fatigue sets in. The issue of time also prevents RAMRIS from being adopted in clinical practice.

RAMRIS is a semi-quantitative score measured on an ordinal scale. An objective and quantitative assessment of MR images for RA is preferred but requires further development in order to ultimately enhance the predictive power of the aforementioned imaging biomarkers.
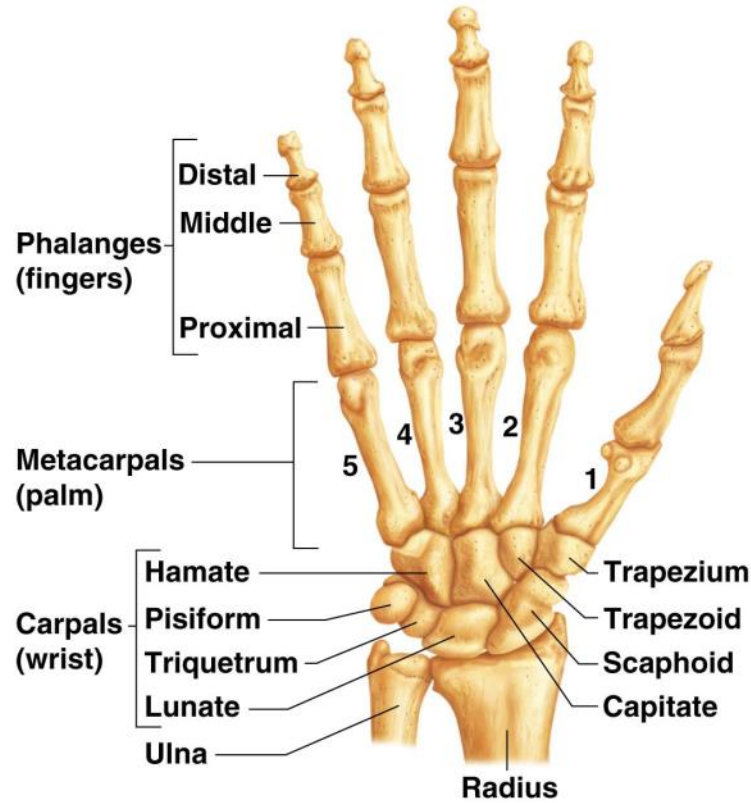
Figure 1.5: Bone anatomy of the left hand. The carpal bones form two transverse rows; the proximal row articulates with the radius and ulnar bones of the forearm. The distal row articulates with the metacarpals of the palm. Source — *Pearson Education, Inc.* (2009).

1.3 Research problem & goal

The visual scoring of eMRI for RA is inefficient, subjective and semi-quantitative. In order to make MRI assessment more objective, sensitive and reproducible, automated quantitative analysis, by means of computer software, is sought. This can reduce study time and is more readily distributable than having to train expert readers. Before any computational approach can be standardised, the automatic detection and quantification of key imaging biomarkers (and potentially new ones) needs to be developed. Hence, the goal of this project was to develop and validate image-processing tools that quantitatively assess pre-erosive inflammatory biomarkers for the early detection of RA. Specifically, this work focussed on characterising BME of the carpal bones using pre-contrast as well as and post-contrast fat-suppressed eMRI of the wrist. The reasons for which are explained below.

Bone marrow edema

Devised from MRI studies, the term bone marrow edema (BME) was first introduced in 1988 [23]. BME is defined by OMERACT as: "a lesion within the trabecular bone, with ill-defined margins and signal characteristics consistent with increased water content". In other words, BME describes the replacement of bone marrow fat by fluid (containing H^+ ions), a change readily detected by MRI. Edematous bone marrow is more vascularised and therefore experiences greater perfusion of contrast agent. After a dose of gadolinium, BME is seen as increased signal (see Figure 1.6). BME is best assessed on post-contrast fat-suppressed images, typically in the coronal plane [24]. A number of studies have followed patients with early RA and concluded that BME was the strongest predictor of radiographic progression [25], [26]. Visually, BME has an indistinct, "feathery" appearance, thus making it challenging to delineate. MRI BME is reported to be observed most frequently in the carpal (wrist) bones, namely the scaphoid, lunate, triquetrum and capitate [13], [27] (see Figure 1.5 for bone anatomy of the hand). This is supported by Li et al., who reported that the carpal bones contain significantly higher total volume of BME compared to other joint sites of the hand. For these reasons the carpalia are considered the most important joint site to characterise BME. Thus, this combination of RA lesion and joint site is the main subject of this study.

Figure 1.6: A contrast-enhanced eMRI of the right wrist (T1-weighted fat-suppressed) with BME of the triquetrum (white arrow). Using RAMRIS, this bone was given a score of 1 on a scale of 0 to 3.

eMRI acquisition protocol

In eMRI studies in healthy subjects and arthritis patients at the LUMC, the wrist is typically scanned in two orientations using a pre- and post-contrast sequence. Before gadolinium-contrast is administered, the wrist is scanned in the coronal direction and sometimes in the transverse direction (see Figure 1.7 for body plane definitions). Post-gadolinium contrast, a coronal scan is acquired as well as a transverse scan. Depending on the orientation, there is a different through-plane resolution; the transverse scan of the wrist has a thicker total slice thickness (slice thickness + inter-slice gap) (3.3mm) compared to the coronal scan (2.2mm). A summary of the data dimensions is provided in Table 1.1.

Table 1.1: Spatial resolution and image dimensions of coronal and transverse eMRI of the wrist.

Orientation	In-plane voxel size (mm^2)	Slice thickness (mm)	Inter-slice spacing (mm)	Image dimensions (voxels)
Coronal	0.195	2.0	0.2	512×512×18
Transverse	0.195	3.0	0.3	512×512×20

Given their higher spatial resolution, the coronal scans were used throughout this work, from segmentation to quantification of BME. For manual annotation, the coronal scan is more intuitive to navigate, especially to the untrained eye. The pre-contrast T1-weighted image contains the most information for detecting bone anatomy, whereas the T1-weighted post-contrast image (fat-suppressed) is needed to identify inflammation. The latter sequence was introduced to replace a T2-weighted scan (as stipulated by RAMRIS) allowing for similar images at shorter acquisition times [28]. Table 1.2 states the eMRI sequences used for pre- and post-contrast acquisitions along with their pulse timing parameters.

Table 1.2: Echo and Repetition times for coronal eMRI sequences of the wrist.

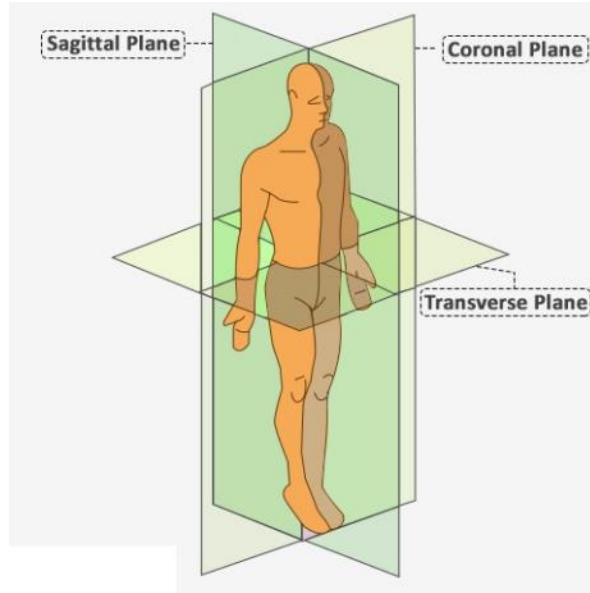|	eMRI sequence	Echo time TE (ms)	Repetition time TR (ms)
(1)	T1 pre-contrast	11	575
(2)	T1 post-contrast, fat-suppressed	9.2-9.6	700

Figure 1.7: The coronal, sagittal and transverse anatomical imaging planes.

1.4 Outline

To recap, the goal of this project is to develop an algorithm to automatically quantify the degree of BME in the carpal bones, using (contrast-enhanced) eMRI. In order to do so, the carpal bones need to be segmented in both the pre- and post-contrast images. This is because the pre-contrast T1-weighted sequence best captures anatomical information to segment bone, while the post-contrast T1-weighted image with fat-suppression gives a better picture of BME. Hence the combination of the two sequences provides the best possible segmentation and quantification. Furthermore, segmentation of individual carpal bones allows for comparison of the quantitative score to RAMRIS, which scores BME on a per-bone-basis. In Chapter 0 a proposed atlas-based segmentation (ABS) technique for carpal bone label assignment is given. This technique incorporates prior knowledge which helps assign distinct labels to each carpal bone. Several experiments were performed to optimise the ABS accuracy using manually labelled reference images as a benchmark.

Chapter 3 describes an automated process to comparatively assess the pre- and post-contrast images of the wrist for the presence of BME. To facilitate this comparison, the two images will need to be spatially aligned (registered). The BME quantification method was also optimised using the training set. The two aforementioned techniques are shown in the overall framework in Figure 1.8.

Chapter 4 details a validation study that compares the quantitative BME score to RAMRIS using an independent test set.

The final chapter, Chapter 5, provides an in-depth discussion of the overall findings of this thesis as well as suggestions for future work.

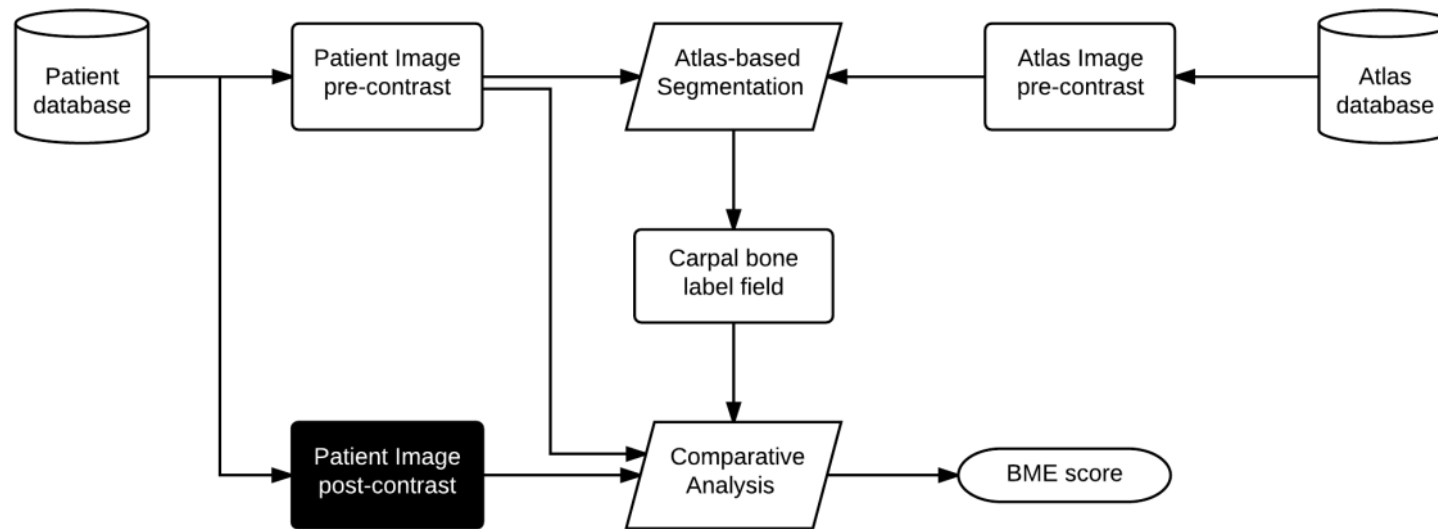**Figure 1.8:** High-level framework for the atlas-based segmentation of the carpal bones and the quantification of bone-marrow edema. The pre-contrast image is matched to a labelled atlas image to obtain a segmentation of the carpal bones (ABS). The resulting label field is used, together with the pre- and post-contrast images in a comparative analysis whereby the degree of BME is calculated per carpal bone.

2. Atlas-based Segmentation

2.1 Background

In this work, the carpal bones are automatically located using atlas-based segmentation (ABS). This technique uses a pre-segmented image, known as the atlas, as a reference to segment new images. Accurate segmentation is the goal and will affect the success or failure of subsequent BME quantification. As such, a significant effort was made to implement, automate and optimise ABS of the carpal bones, the details of which are described and discussed in this chapter. A more general and detailed overview of ABS is provided by T. Rohlfing et al. [29].

Rationale

A major challenge in processing MR images is the absence of a standardised signal intensity scale, such as the Hounsfield unit in CT. This is because the MR signal is affected by magnetic field inhomogeneities and positional changes within the RF-coil, as well as other scanner read-out artefacts. As a result, intensities for a specific tissue type will vary across patients. This effect is present even when examining the same patient, on the same scanner with an identical MR sequence [30], [31]. Hence, in MR images of the wrist, there is no clear and dependable relationship between a voxel's intensity value and its anatomical label. Furthermore, a carpal bone comprises more than one tissue type and thus displays a range of intensity values. Bone marrow, being a fatty tissue has a high intensity in a T1-weighted image, whereas the cortical bone surrounding it appears dark, like air or tendons.

Given the carpal bones have a highly inhomogeneous appearance, simple intensity-based segmentation such as a threshold or level sets is unlikely to render satisfactory results. ABS on the other hand, is well suited to the task, as it captures *a priori* information about anatomical structures. This includes the bone's shape, location and orientation in relation to neighbouring structures. Furthermore, as mentioned in 1.4, the segmentation ought to not only detect the carpal bones but also distinguish them from each other. ABS can easily accommodate for this by means of distinct label classes.

In short, ABS allows for automated segmentation in 3D and helps incorporate prior knowledge. The decision to use an ABS approach was made following a literature review on bone segmentation techniques in MR conducted at the start of the project. It is included as a reference in 0.
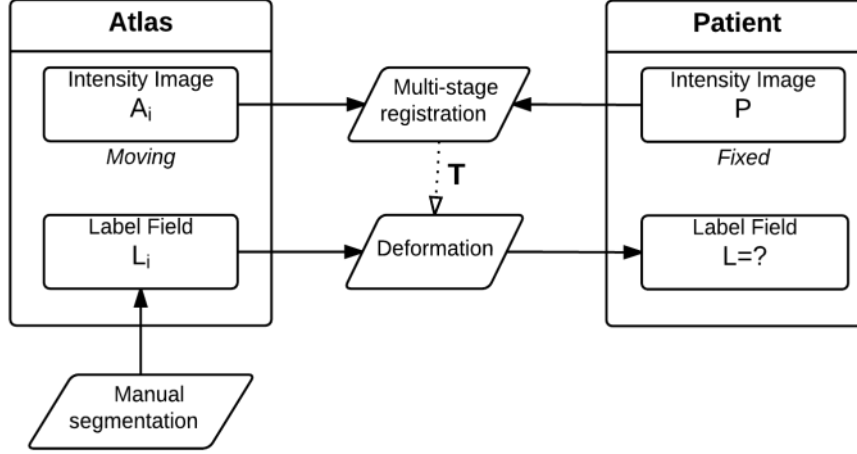
Figure 2.1: High level atlas-to-patient registration framework for ABS. The atlas image A_i is defined as the moving image and is deformed onto the patient image P. Registration is achieved in two stages, using an affine transformation for global alignment, followed by a non-rigid transformation parameterised by cubic B-splines. The final coordinate transformation T, which maximises the similarity between A_i and P is applied to the manually delineated label field L_i to obtain the segmentation estimate L.

Registration

ABS uses one or more pre-segmented reference images to segment new images. The set of reference images is known as the atlas. Through image registration, the atlas image A_i is matched to an unseen patient image P. Similarly the labels of the atlas (image L_i) can be propagated to the patient image to obtain the carpal bone segmentation L. The framework for this process is given in Figure 2.1. Image registration is the process of bringing two images (or volumes) into spatial alignment. It calculates a coordinate mapping T between the two images such that their similarity is maximised. In this work, all registrations between atlas and patient images were carried out using the `elastix`[3] image registration toolbox (open-source) [32]. For consistency, this thesis uses the mathematical notation as defined in `elastix` documentation and associated literature (see [32], [33]). Registration is achieved by iteratively warping or deforming the *moving image I_M* onto the *fixed image I_F*. This process involves a transformation and an interpolation step. The choice of transformation controls the allowable degrees of freedom in the deformation. In this work, atlas to patient registration is achieved by using two sequential registrations. The first uses an affine transformation model and aims to achieve global alignment. The second uses a non-linear transformation model, parameterised by cubic B-splines, further deforming the atlas image onto the patient image. The atlas image A is defined as the moving image so as to not distort the patient image P, where measurements are made. The total atlas-to-patient coordinate transfor-

---

[3] www.isu.uu.nl/elastix

18

mation is given by the functional composition of the affine and the non-rigid B-spline transform:

$$T = (\,T_{Bspline} \circ T_{affine}\,)(x).$$

An intensity-based similarity metric, defined as $\mathcal{S}$, measures the quality of alignment of each registration. Such a metric is selected based on the expected relation between intensity values of the two images. Common similarity measures for mono-modal registration are Sum of Squared Differences (SSD) and Normalized Cross-Correlation (NCC). The former assumes the fixed and moving image have an equal intensity distribution. NCC on the other hand, can handle some intensity inhomogeneities, assuming a linear relation between intensities values of the two images. For mono- but particularly for multi-modal registrations, Mutual Information (MI) is commonly used. This is an entropy-based measure stemming from information theory. It assumes no prior functional relationship between the image intensities. Section 2.3 offers a closer look at the different similarity measures and their effect on registration success.

The registration process is formulated as the minimisation of a similarity-based cost function $\mathcal{C}$ with respect to the vector of transformation parameters $\boldsymbol{\mu}$:

$$\hat{\boldsymbol{\mu}} = arg \underbrace{min}_{\boldsymbol{\mu}} \; \mathcal{C}\big(\boldsymbol{T}_{\boldsymbol{\mu}}; I_{fixed}; I_{moving}\big),$$

where $\mathcal{C} = -\mathcal{S}$ (i.e. high similarity, low cost). Commonly, an iterative optimisation strategy is employed based on gradient descent. Every iteration, voxels are sampled from the fixed image to compute the cost function derivatives. A step is taken in the direction of the negative derivative of the cost function until a minimum is reached.

In looping over the fixed image, all voxels may be used but it has been shown that a subset also suffices, reducing computational cost [34][35]. Samples can be chosen randomly and either on- or off-grid. When sampling off-grid (i.e. at between voxel locations) the grey-values are obtained by intensity interpolation.

Similarly, during the optimisation, samples taken from the moving image at non-voxel positions requiring interpolation for which a range of methods can be used, trading-off between quality and speed. For the best results, the final deformed image (if needed) is resampled using third-order B-spline interpolation.

Finally, a multi-resolution approach is used reduce the data size and complexity and increase the chance of successful registration. Smoothing and sometimes down-sampling are specified for a number of resolution levels forming what is known as an image pyramid or scale space. The exact specification of the pyramid schedule is dependent on the data and the application at hand. The same goes for all registration parameters; there are no clear-cut rules and the algorithm should be tuned to the specific needs. The choice of registration parameters is explained in detail in Chapter 0.

Affine transformation model
The initial registration uses an affine transformation model, allowing for translation, rotation, scaling and shearing. Shearing or skewing, has the effect of turning a square into a

slanted non-square parallelogram, or a circle into an ellipse. An affine transformation preserves the area of geometric figures, as well as the relative distances of collinear points. The combination of these four transformations, applied in all three directions (x, y, z) yields a total of twelve degrees of freedom.

B-spline transformation model

The second registration uses a non-linear transformation model of three-dimensional cubic B-splines. It takes the result of the first registration as the starting point. It is parameterised by a vector of control points which is iteratively optimised to achieve high similarity between the two volumes. These control points are defined on a regular grid, overlayed on the fixed image (which is unrelated to the sampling grid). The spacing of the control points (σ) is defined for each computation level and is a key parameter. It effectively determines the 'elasticity' of allowed deformations. A densely spaced control point grid allows for greater deformations than a sparsely spaced one. Chapter 3 explains how the optimal grid-size spacing was determined.

Optimisation

The transformation models (the B-spline model in particular) involve a large number of parameters that precisely define and guide each registration, allowing for complex deformations. Whilst giving a high amount of flexibility and control over the process, it also creates many possible combinations of settings. In order to select the optimal parameters, both registration steps were optimised. Details of these experiments are discussed in 2.2 and 2.3.

Multi-atlas

In principle, a single atlas is sufficient to segment unseen images. However, it has been shown that the inclusion of multiple atlas images yields better segmentation success [36]. This is analogous to gains seen in the combination of classifiers in the field of pattern recognition. Each transformed atlas image effectively represents a unique classifier, assigning a label value to the voxels in the target image [37]. The manual segmentation and registration can then be seen as the classifier-training phase. A multi-atlas approach helps reduce the impact of bias from a single labelled subject and to model anatomical variability [38]. Instead of a single estimate of the label-field, a multi-atlas framework renders one for each atlas subject. These labels are then combined to generate the final segmentation in a process known as label fusion.

Label fusion

Each pairwise registration between the patient image and three atlas images yields a coordinate transformation T_i, which can then be applied to the atlas label fields L_i, resulting in a set of deformed label images $\{L_1 \circ T_1, ..., L_n \circ T_n\}$ for $i = 1, ..., n$. Each deformed label field is an estimate of the final segmentation. A straight-forward but effective way to combine these estimates is by Majority Voting (MV); for every voxel in the patient image, each deformed label field casts a vote on whether said voxel belong to a certain carpal bone or not. Following label fusion, the segmentation mask may require some post-processing in order to

fill gaps or remove other artefacts. Figure 2.2 illustrates the overall ABS framework, including label fusion.
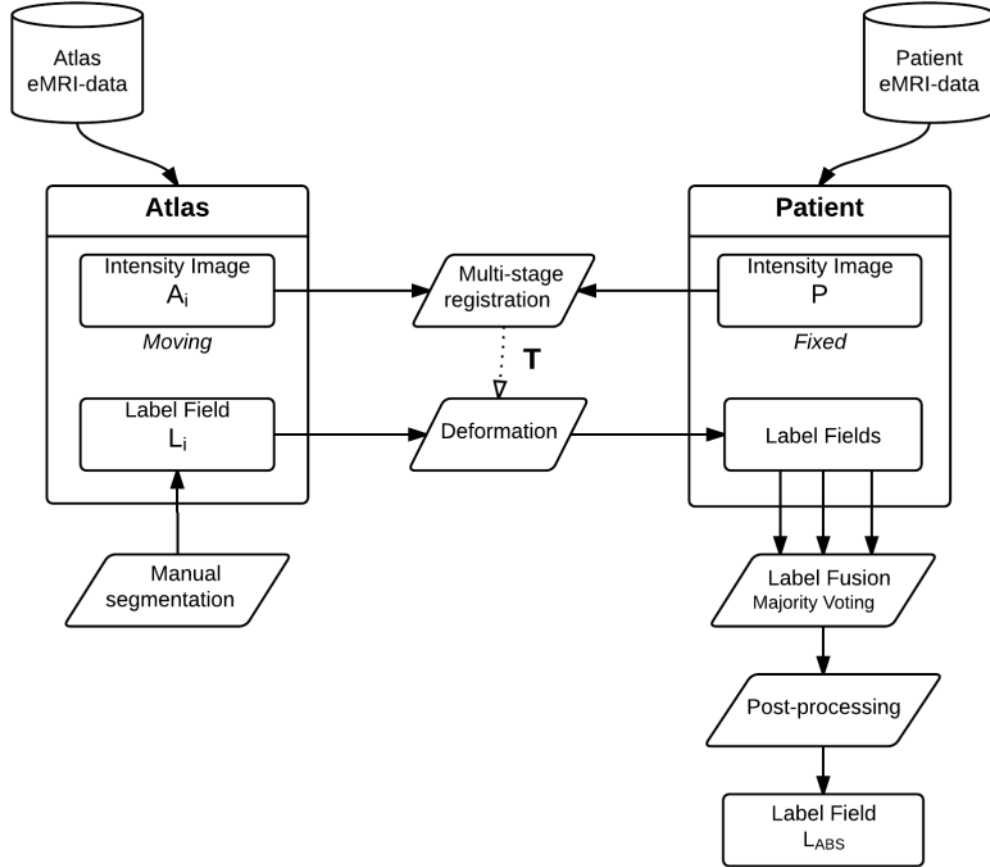

Figure 2.2: Framework for multi-ABS of the carpal bones, including MV label fusion and a post-processing step resulting in the final ABS label image L_{ABS}.
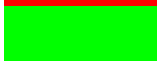
2.2 Method

The entire ABS framework was implemented using MeVisLab[4], a development environment for medical image processing and visualisation. Calls to `elastix` were made using a dedicated wrapper within MeVisLab. All automation was programmed using Python scripts within MeVisLab's integrated text editor MATE. Parameter files used for the registration steps in ABS as well as in BME quantification (Chapter 3) are available at: http://elastix.bigr.nl/wiki/index.php/Par0030.

Atlas creation

For this work, three healthy individuals (no arthritis or arthralgia) were selected at random from a pool of 32 subjects (mix of healthy volunteers and RA patients). The coronal MR images (T1-weighted, pre-contrast) of the right wrist of these three subjects were used to create the atlas. Consideration was given to ensure the images of the three individuals differed somewhat in appearance. In the atlas images the carpal bones were manually segmented using ITK-SNAP[5] [39]. A rheumatologist experienced in reading RA MRI provided training. Care was taken to include both trabecular bone (marrow) as well as cortical bone (dense outer shell). The carpal bones are labelled 1-8, in order of the 'SLTPTTCH' mnemonic used by medical students. A corresponding colour scheme was generated and used throughout visualisation (see Table 2.1 and Figure 2.3). The colours were chosen to be as distinct as possible by maximising the distance between colours around the colour wheel.

Table 2.1: Segmentation labels for the carpal bones and corresponding colours.

Carpal bone	Label	RGB value			Colour
Scaphoid	1	255	0	0	
Lunate	2	0	255	0	
Triquetrum	3	0	0	255	
Pisiform	4	255	255	0	
Trapezium	5	0	255	255	
Trapezoid	6	255	0	255	
Capitate	7	255	157	0	
Hamate	8	0	170	127	

---

[4] www.mevislab.de

[5] www.itksnap.org

22
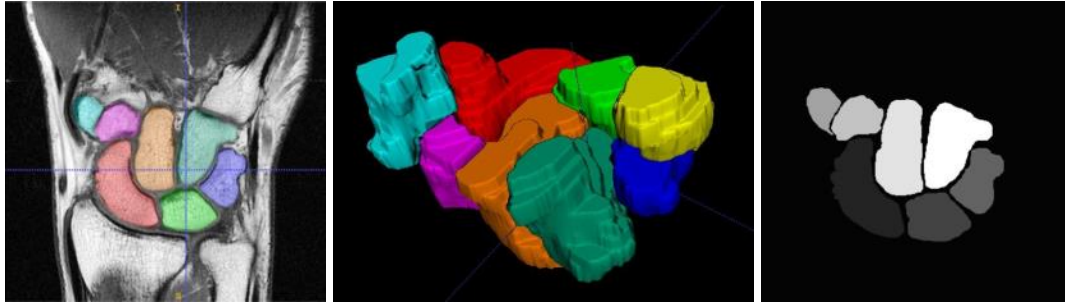
Figure 2.3: Manual carpal bone segmentation of T1-weighted eMRI of the right wrist. The MR image with segmentation overlay (left). A 3-D rendering of the segmentation (centre) and the grayscale label field (right).

Affine registration

Given the eMRI data is not free from inhomogeneities and bias field artefacts, NCC was used as the similarity metric for the affine registration.

For all registrations an adaptive stochastic gradient-descent optimiser in combination with a random coordinate sampler was used [40]. The number of iterations and the number of spatial samples were user-specified. For this optimiser and sampler combination, the `elastix` manual[6] suggests 3000 spatial samples per iteration, with new samples taken in every iteration. Extensive testing by the creators of `elstix` form the basis of these recommendations (see [35]). Similarly, the number of iterations was set at 1000, which according to the manual is a robust setting.

Linear first order B-spline interpolation was used during registration/optimisation. For the final deformation, this was increased to a third order B-spline for the best results.

The goal of the affine registration is to match global structures. Hence a Gaussian scale space was used, meaning smoothing is applied at each resolution, but no down-sampling. Four resolution levels were used in the smoothing image pyramid for both the fixed and moving image. By default, `elastix` applies equal smoothing in all directions. However, given the data suffers from highly anisotropic voxel sizes ($x, y, z \mapsto 1:1:10$), less smoothing was applied in the z-direction by a factor of ten. By tailoring the smoothing to the anisotropic nature of the data, a more evenly blurred volume results (see Figure 2.4 for details).

Table 2.2 summarises the key parameter settings for the affine registration, they remained unchanged throughout in order to independently test B-spline registration parameter settings.

---

[6] http://elastix.isi.uu.nl/download/elastix_manual_v4.7.pdf
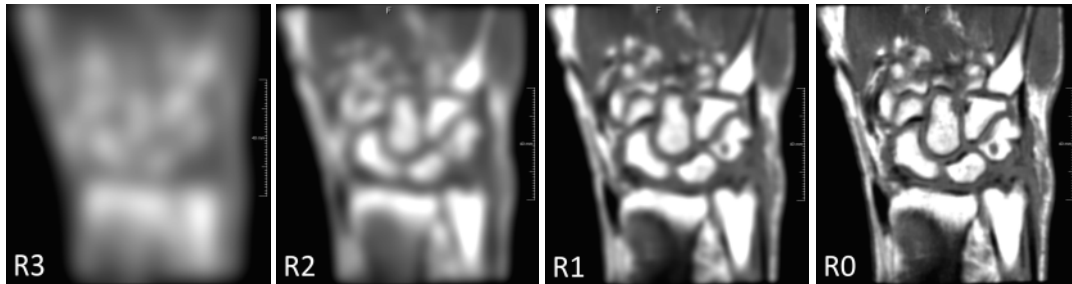
Figure 2.4: The effect of a Gaussian smoothing schedule, from highest to lowest (R3-R0) resolution level. It ensures global structures are matched before smaller ones. The amount of Gaussian smoothing (σ in voxels) in each direction (x,y,z) at four different scales is: (40 40 4, 20 20 2, 10 10 1, 5 5 1). In the z-direction, the applied smoothing is a factor ten less than in the x- and y-directions in the first two levels. No smoothing is applied (σ=1) in the z-direction in the last two levels.

Table 2.2: A summary of important `elastix` registration parameter settings. Affine registration parameters were fixed. For the B-spline registration, (the range of) different parameter settings that were tested are shown.

`elastix` **parameter**	**Registration**	
Transform	Affine	B-spline
Similarity metric	NCC	SSD, NCC, MMI
Optimiser	ASGD	ASGD
Number of iterations	1000	500-3000
Sampler	Random Coordinate	Random Coordinate
Number of samples	3000	500-5000
Image pyramid	Gaussian smoothing	Gaussian smoothing
Number of resolutions	4	3
Final interpolator	3rd order B-spline	3rd order B-spline
Final grid spacing (mm)	N/A	Custom, 4, 8, 16

SSD = Sum of Squared Differences, NCC = Normalized Cross Correlation,
MMI = Mattes Mutual Information, ASGD = Adaptive Stochastic Gradient Descent

24

B-spline registration

The non-rigid registration is modelled by non-linear transformation, parameterised by a cubic B-spline polynomial. It takes the result from the affine registration as a starting point and assumes global alignment between the two wrists is achieved. As in the affine registration, NCC was used although other similarity metrics were also tested in parameter optimisation. The same optimiser, random sampler and interpolation settings were used. Instead of four-levels, a three-level pyramid was used. The number of iterations as well as the number of spatial samples was set to 2000. The final control point grid spacing was set to 8mm, giving a schedule of 32mm-16mm-8mm, for all directions. This crucial parameter was also optimised with respect to segmentation accuracy.

Label fusion

Once all three atlas label fields had been deformed into that patient coordinate space, they were fused using MV. Equal weights were assigned to each voter. The weight refers to the probability of being true (namely $p = 0.\overline{33}$). A majority of three voters implies a vote threshold of 0.66. In other words, when two or more of the voters agree that a certain voxel is bone/non-bone, then said label is assigned.

Post-processing

As a way to refine the fused ABS result, a post-processing method was developed based on Otsu's method [41]. Otsu's method can be used to automatically segment an image into foreground and background on the assumption its intensity distribution is roughly bi-modal. The optimal threshold value is found by iteratively minimising the intra-class variance of the two pixel classes. In pre-contrast MR images of the wrist, Otsu's method is a very simple yet effective way of detecting bone marrow and other fatty tissues, which display a high intensity (see Figure 2.5). The fused ABS result can be restricted to regions that coincide with the labels derived from Otsu's method. The MeVisLab network used to implement this post-processing step is given in Figure 2.6.
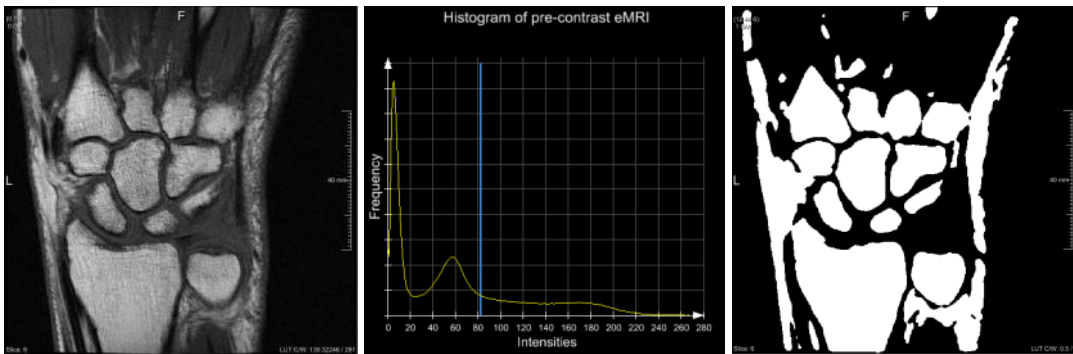
Figure 2.5: A pre-contrast T1 eMRI of the wrist before (left) and after (right) applying Otsu's method. The intensity histogram is shown (middle) with the threshold value (82) in blue.
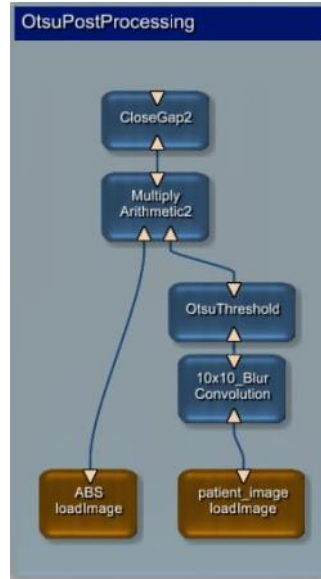
Figure 2.6: MeVisLab network (read from bottom to top); the post-processing method takes the original patient image (pre-contrast) and applies a 10x10 averaging filter. Then the Otsu threshold image is calculated resulting in a binary image. This segmentation is multiplied by the fused result from the ABS. Finally a morphological CloseGap filter is used to close any holes.

Parameter selection experiments

Key model parameters for the affine registration were determined empirically, using a number of atlases and patient image pairs. Once suitable settings were found they were fixed to allow the isolated testing of parameters in the B-spline registration.

The B-spline transformation model has a large number of parameters that precisely define and guide the registration, allowing for complex deformations. Whilst giving a high amount of flexibility and control over the process, it also creates an impractical amount of possible combinations of settings. The parameter set (defined as the vector $\boldsymbol{\mu}$) can be varied giving a multi-dimensional parameter space to optimised. Seeing as the number of permutations is too large to conceivably test, an effort was made to test a number of key parameters for a range of settings. These were, in order: the (maximum) number of iterations per computation level, the similarity measure, the final B-spline grid spacing and the number of randomly selected spatial samples. The optimal parameter setting from the first experiment was fixed and used in subsequent tests etc.

Performing these searches did not result in a global optimal setting, but helped decide if parameters have a significant influence on the registration accuracy, and subsequently, segmentation success. All parameter testing was fully automated using Python scripts to enable batch computations. The high-level pseudo code is given in Figure 2.7 as well as a flow-chart in Figure 2.8. Note for purposes of parameter testing, the Otsu post-processing was not used as this method was introduced after parameter testing to try and refine segmentation results.

Measuring segmentation success

At the time of testing, eight training images were manually labelled (coronal T1 scan of the right wrist, pre-contrast). These served as the benchmarks or 'ground-truth' allowing the abovementioned B-spline registration parameters to be optimised. The final ABS label field L_{ABS} was compared to the manually segmented benchmark L_{BM} using the Dice Similarity Coefficient:

$$DSC(L_{ABS}, L_{BM}) = \frac{2|L_{ABS} \cap L_{BM}|}{|L_{ABS}| + |L_{BM}|},$$

where $\cap$ represent the overlap between the two label fields. The Dice overlap score, or similarity index, is the most widely used measure for evaluating the performance of a segmentation algorithm [36]. DSC has a value of 0 when there is no overlap and 1 when they perfectly agree. It is calculated per label class (i.e. per carpal bone) excluding the background class '0'. An average DSC can then be calculated per patient or per carpal bone.

```python
for param in param_list :
    for patient in patient_database :
        for atlas in atlas_database :
            elastix_affine.compute()
            elastix_bspline.compute()
            transformix.compute()
        computeMajorityVote(L_A1, L_A2, L_A3)
        computeOverlap(L_ABS, L_BM)
```

Figure 2.7: Pseudo code for the high-level Python script used to automatically test a range of parameter settings for the B-spline registration. A single parameter is varied at a time, and the effect on the segmentation accuracy is measured using the degree of overlap between the ABS label field and the manually segmented benchmark.
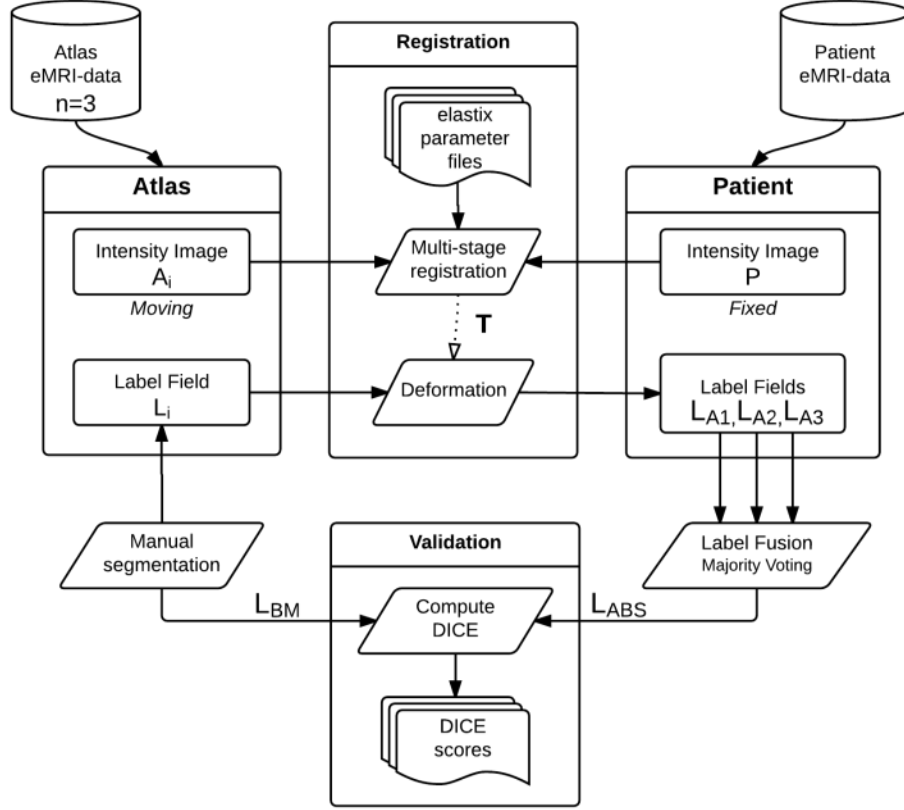
Figure 2.8: Automated parameter testing in a multi-ABS framework. A 3-subject atlas is used to segment the carpal bones from each patient. The three resulting segmentation estimates L_{A1}, L_{A2}, L_{A3} are fused using majority voting. Finally the overlap between L_{ABS} and L_{BM} is computed per label class using Dice similarity coefficient. `elastix` parameter files are varied at the input and the resulting segmentation accuracy, in terms of DSC, is recorded at the output.

2.3 Optimisation

This section reveals the methods and results of optimising key parameters in the B-spline registration. To enable isolated testing, the atlas, affine registration parameters and label fusion method remained constant.

Maximum number of iterations

The number of iterations per resolution level was tested, ranging from 500-3000 in increments of 500. The parameter specifies the *maximum* number of iterations, however using the gradient descent optimiser, this is also the minimum number, since there is no other stopping condition. A scatter plot of the average DSC versus the number of iterations is shown in Figure 2.9. The plot reveals there was some spread among the patients with the average DSC ranging from approximately 0.79 to 0.89. From this spread, we can conclude that ABS performed better on some patients than on others. The order of segmentation performance of patients remained consistent as the number of iterations was varied (i.e. patient 8 always performed best and patient 4 the worst, Figure 2.9). The varying segmentation success can be due to some patients' images being more similar in appearance to one or more of the atlases, leading to a higher level of agreement and thus a more accurate segmentation.
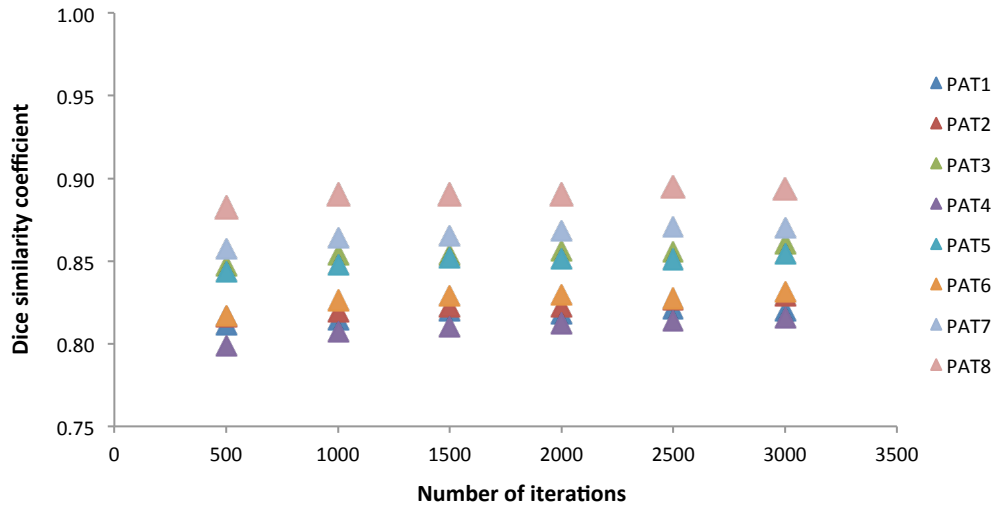
Figure 2.9: The average DSC vs. number of iterations for different patients, averaged across carpal bones. The number of iterations refers to the optimiser setting in the B-spline registration.

To gain an understanding of the individual carpal bones' segmentation performance, an average across the patients is shown in Figure 2.10. In general, the higher the number of iterations, the greater the registration success. However it is clear from the results that varying the number of iterations had no significant effect on ABS success (which in turn depended on registration success) as measured by the DSC. Interestingly, the hamate and capi-

tate were consistently the best-segmented carpal bones using this method. Possible explanations for this fact are given in the discussion of this chapter (section 0).
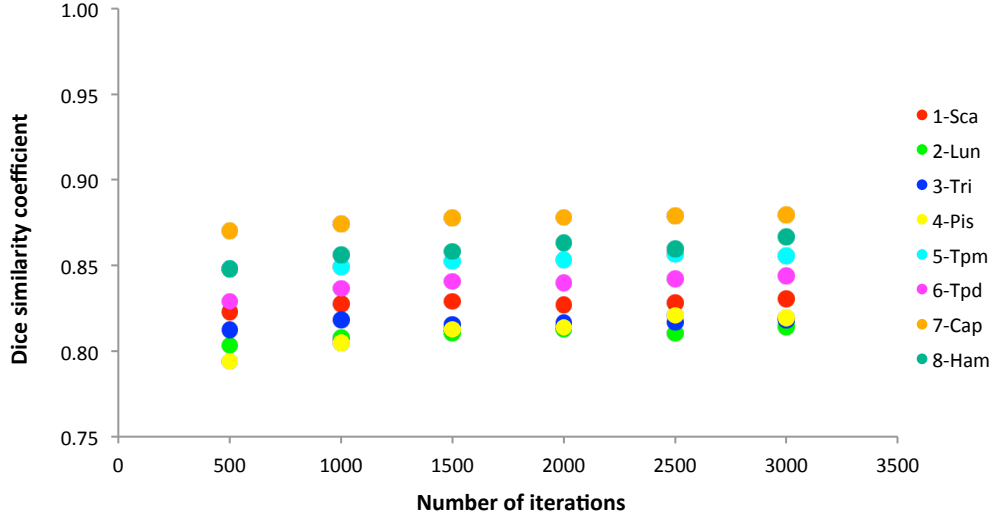
Figure 2.10: Average DSC vs. number of iterations for different carpal bones, averaged across patients. Little effect on segmentation accuracy, as measured by the DSC, was observed. In the interest of saving computation time, a lower setting is therefore recommended (e.g. 1500-2000 iterations).

It is clear that, for the group of eight patients, varying the allowable number of iterations per resolution level in the B-spline registration has little effect on the segmentation success as measured by the DSC. At 500 iterations, the average segmentation success is poorest. For the other settings there is no significant variation. Given the segmentation is quite robust to changes in this parameter setting, we can consider the computation time, which increases from 30 seconds to about four minutes with an increasing number of iterations. Hence, we conclude that 1500-2000 iterations is a suitable parameter setting to satisfy the trade-off between computation time and segmentation accuracy.

Similarity Measure

Three similarity measures (built-in to `elastix`) were tested, namely: Sum of Squared Differences (SSD), Normalized Cross Correlation (NCC) and Mattes Mutual Information (MMI). Using eight test-subjects, the SSD metric yielded one gross segmentation failure or outlier, achieving a DSC of just 0.11 (11%), see Figure 2.11. Based on this alone, the alternatives are preferred, as they showed no significant outliers. Between NCC and MMI, NCC shows slightly less variation and a slightly higher median. Given the images registered are from the same modality and with the same MR sequence, it was expected for NCC to work well. Though usually employed for registering multi-modal data, the results show that MMI can also be used for mono-modal data. Based on these results, the NCC metric is employed in both registration stages of ABS.
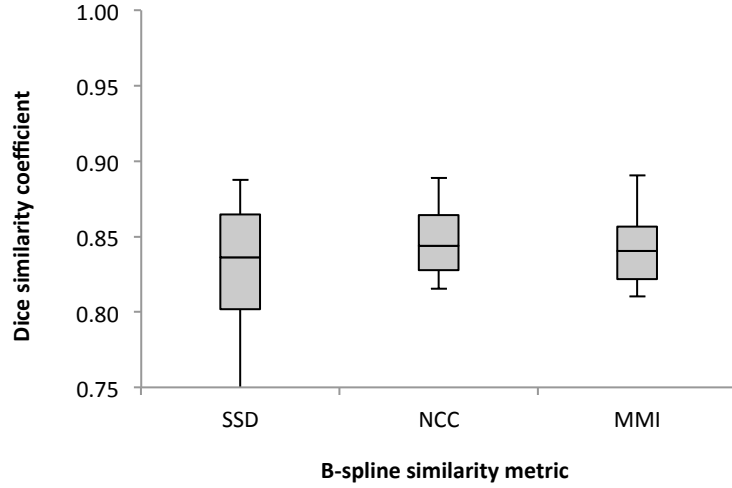
Figure 2.11: Average Dice similarity score vs. B-spline similarity metric. Sum of Squared Differences (SSD) similarity transform showed an outlier at 0.11 DSC (not shown due to scale). Normalised Cross Correlation (NCC) achieves the most segmentation overlap with benchmarks on average, but Mattes Mutual Information (MMI) is also suitable for mono-modal registration.

B-spline grid spacing

A crucial aspect of the B-spline transformation model is the spacing (σ) between the control points x_k which are defined on a grid over the fixed image. The parameter `FinalGrid-Spacing` stipulates the separation, in millimetres, of the control points in the final resolution level. During preceding tests, the final grid spacing was set to 8mm. It was decided to halve and double this value, and collect the DSC values accordingly. From Figure 2.12 it can be seen that 4mm isotropic final grid spacing yields the best segmentation results with a median DSC of just under 0.85 (dotted line).
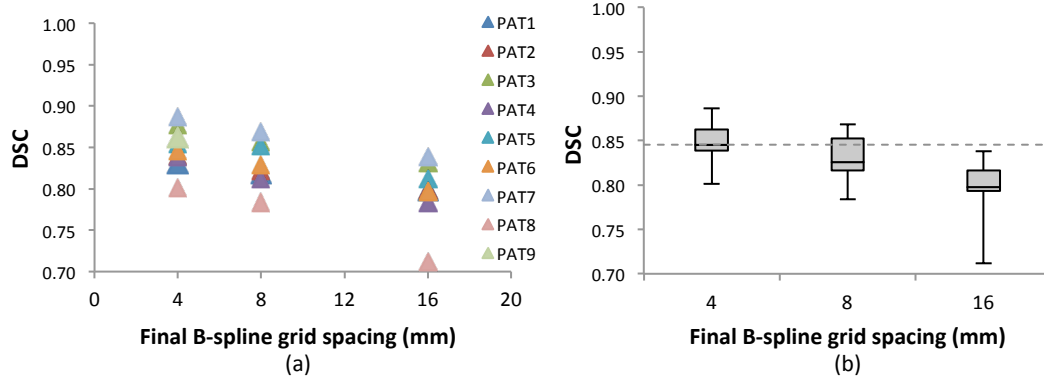
Figure 2.12: Average DSC vs. Final B-spline grid spacing. (a) shows the segmentation success for different patients, averaged across carpal bones. (b) shows the boxplot average of the same data. It is clear that on average (dotted line is the median), the segmentation overlap between ABS and benchmark labels is greatest using a 4mm final grid point spacing in the B-spline registration.

By default, `elastix` isometrically doubles the grid spacing for each resolution level counting back from the final level. Meaning, when a final B-spline grid spacing of 4mm is set, a three-level pyramid will have the following control point spacing for the respective levels, equal in all three directions: 16mm, 8mm, 4mm. Given the MR images of the wrist are highly anisotropic, we can specify a custom grid spacing schedule that is better suited to the image dimensions. The coronal scans have the following dimensions in terms of voxels and physical units:

Table 2.3: Image dimensions of coronal T1-w eMRI of the wrist.

Dimension	Image size (voxels)	Voxel spacing (mm)	Image size (mm)
x	512	0.195	100.00
y	512	0.195	100.00
z	18	2.20	39.6

Anatomically, carpal bones vary in size, and due to the slice selection of the MRI acquisition, only a small part of the bone may be visible in any one slice. By inspection, the minimum cross-sectional distance of the bones as they appear *in*-plane, ranges from 30-100 voxels, or 5-20mm in physical units. Hence, we expect an *in*-plane grid spacing of 4mm (every 20 voxels) to be able to match these fine structures. Going lower than 4mm we approach the voxel size and risks giving the registration too much freedom, leading to irregular transformations. This can be corrected by adding a regularisation term to the cost function which penalises excessive deformation. For simplicity, this penalty term was not added to the cost function. In order to determine the coarsest control point spacing, it was decided

32

to aim for 4 control points across the length and width of the image. This allows the major structures in the image to be deformed first, before halving the spacing and deforming more locally. Taking into account anisotropic voxel sizes and target object sizes, the custom B-spline control point spacing was set as follows:

Table 2.4: Custom B-spline control point spacing in mm for a three-level pyramid.

Computation level	Control point spacing (mm)		
	x	y	z
1	24	24	10
2	12	12	5
3	4	4	1

When compared to a naïve setting of the control point spacing, the custom grid achieved greater segmentation accuracy on average as measured by the Dice overlap score (Figure 2.13) as well as less spread overall.
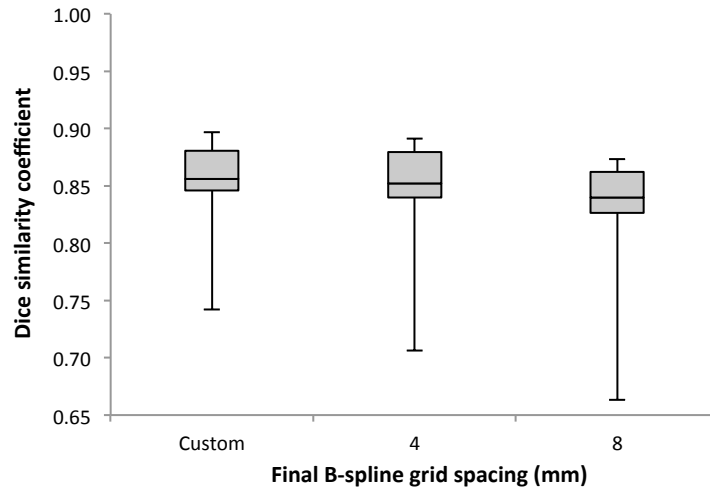
Figure 2.13: Average DSC vs. Final B-spline grid spacing; the segmentation accuracy, as measured by the Dice overlap score improved with smaller final grid-size spacing (see Figure 2.12). In comparison, the custom grid spacing performs even better, with a median DSC of 0.86.

Number of spatial samples

The number of spatial samples in the previous experiments was set at 2000. In order to investigate the effect of this parameter on the segmentation success, its setting was varied from 2000 to 5000 in increments of 1000. Increasing the number of spatial samples, when tested on ten subjects, yielded no significant change in the average dice similarity rating (see Figure 2.14. In the interest of saving computation time, it is therefore recommended for 2000 spatial samples to be taken at each resolution level. A lower setting is not recommend-

ed by the `elastix` manual (and [35]), which states 2000 samples is a minimum when combining the Adaptive Stochastic Gradient Descent Optimiser with the Random Coordinate sampler. Out of curiosity, the segmentation quality was observed using only 500 and 1000 samples. Though still fairly robust, the minimum DSC drops compared to using 2000 samples or higher. Hence, the results support the recommended minimum setting of 2000 (new) random samples per iteration.
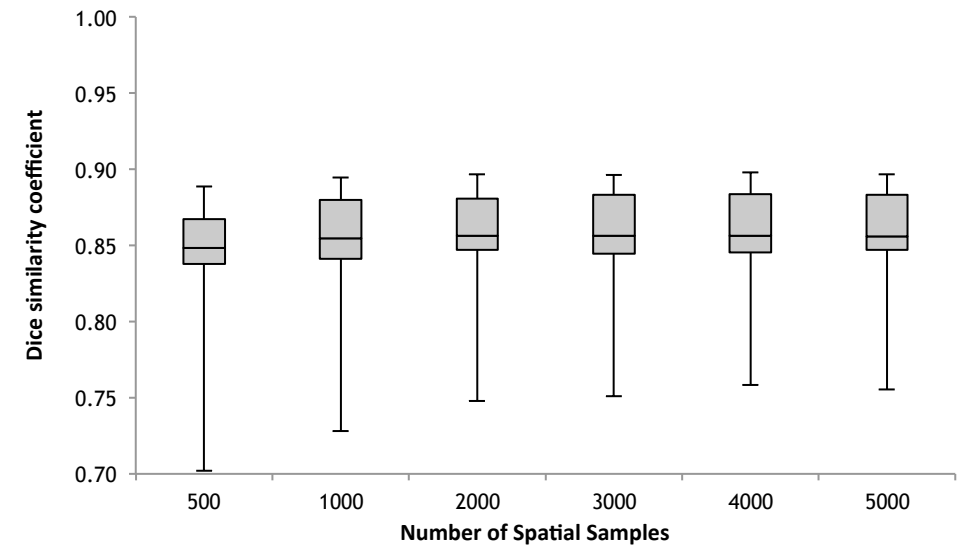
Figure 2.14: Average carpal bone overlap scores (Dice) when varying the number of spatial samples taken (using Random Coordinate sampling) in each iteration of the optimiser. Varying this parameter setting had little to no effect on the segmentation accuracy.

Optimisation results

Post-optimisation, the settings that yielded the highest Dice overlap were fixed for the remainder of the work. By the end of the thesis, 13 manual segmentations had been created. Performing ABS on this set using the optimal parameter settings yielded overall mean DSC of 0.85 (see Figure 2.15).
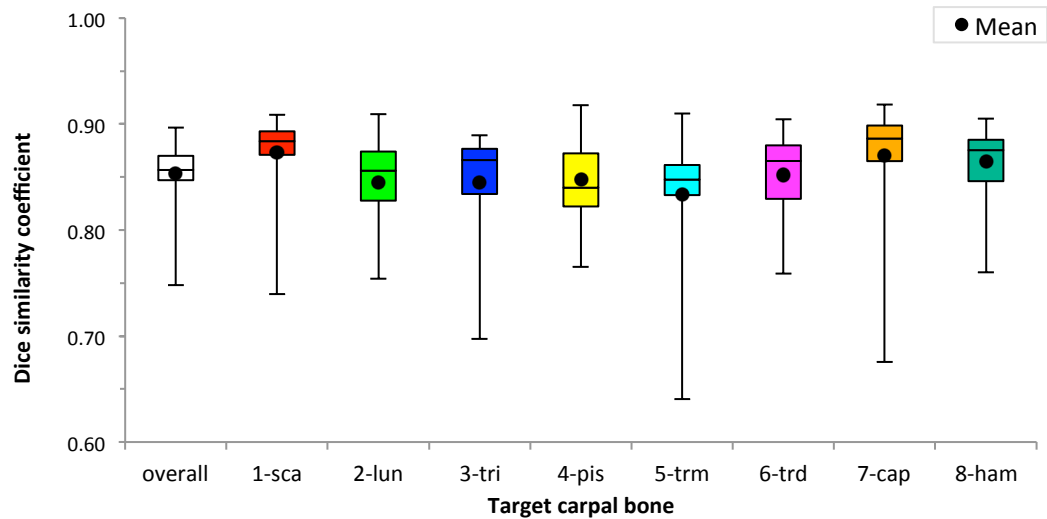
Figure 2.15: Box-and-whisker plots of carpal bone segmentation accuracy (DSC) using 13 manually segmented subjects as benchmarks. The results are shown per carpal bone, with all bones combined on the far left; a mean DSC of 0.85 was achieved across all carpal bones.

Post-processing

Post-processing of the ABS labels confines the segmentation to (normal) bone marrow, reducing false positive segmentation error (see Figure 2.16). Segmentation accuracy using Otsu post-processing was seen to be slightly lower overall, compared to stand-alone ABS. For comparison see Figure 2.17. The overall mean and median DSC are given in Table 2.5.
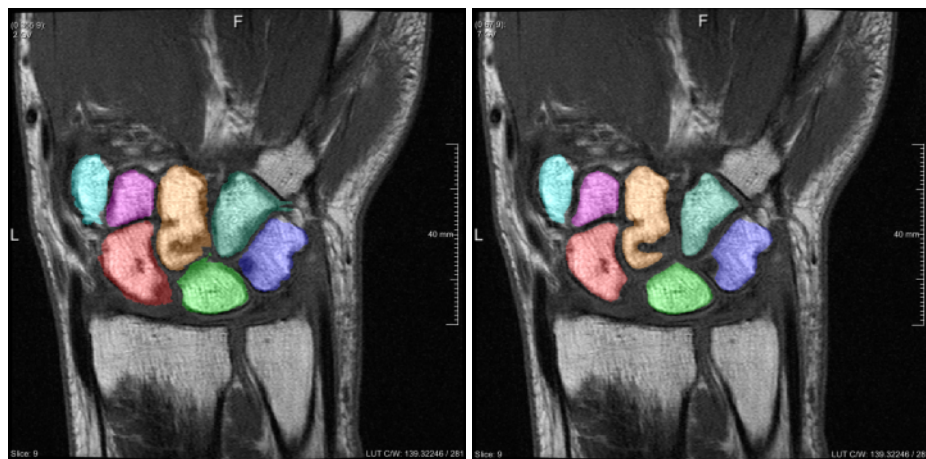
Figure 2.16: Refining the ABS label field (left) by multiplying it by an Otsu threshold image. The post-processed result is a more conservative estimate of the carpal bone location (right).

35
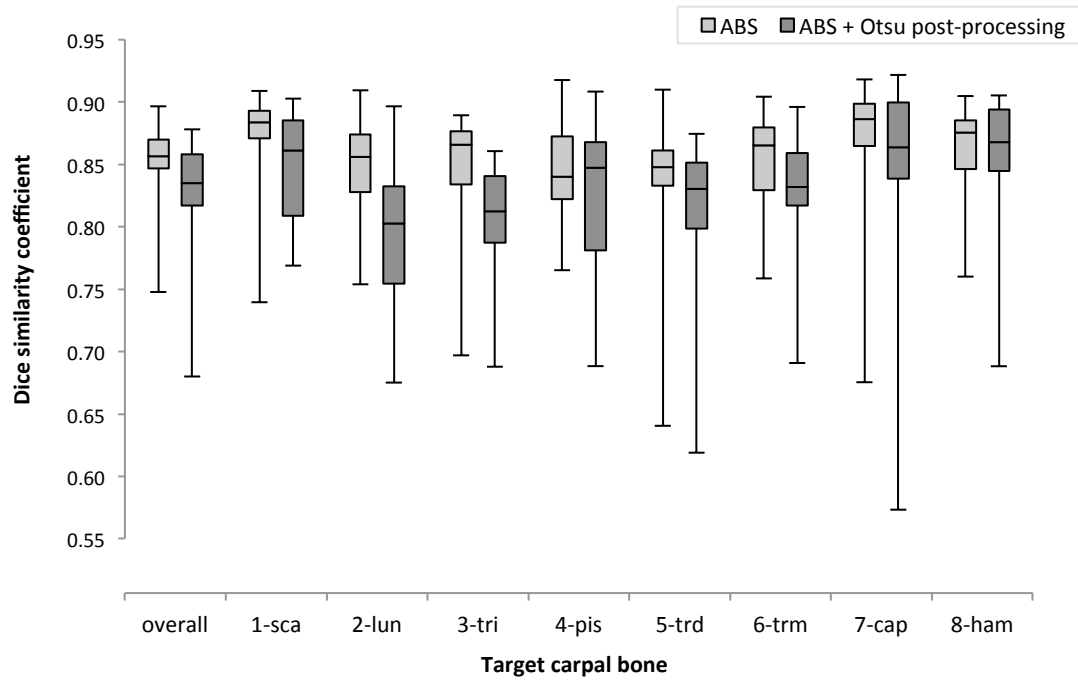
Figure 2.17: Box-and-whisker plots of carpal bone segmentation accuracy (DSC) for the training set (n=13) using ABS with and without Otsu post-processing. On the far left the results for all carpal bones are combined with the remainder showing the results per label class or carpal bone.

Table 2.5: A summary of the overall Dice overlap scores for all carpal bones for all training subjects

Structure	Experiment	DSC	
		Median	Mean
Overall	ABS	0.86	0.85
	ABS + Otsu post-processing	0.83	0.83

2.4 Discussion

The proposed ABS segmentation of the carpal bones is able to automatically and correctly locate the joints with, on average, a segmentation accuracy (Dice) of 85%. Further improvements to the algorithm are needed to account for the 15% of the bony voxels incorrectly labelled. Segmentation quality using a registration-based approach is limited by the accuracy of each pairwise registration, and (thus) the (anatomical) similarity between the patient and atlas subjects. Moreover, in a multi-atlas framework, label fusion also affects segmentation quality. In light of these three areas, a discussion follows regarding the atlas-based carpal bone segmentation method used in this study.

Atlas selection

Three healthy subjects were randomly selected from the training set to form the atlas. Atlas images with high co-similarity do not capture as much inter-subject anatomical variability, making for a poorer model. Hence, care was taken to select images that were dissimilar in appearance, though this was not done objectively. The term "healthy" indicates subjects have had no prior RA-related symptoms such as joint pain or stiffness. Without signs of advanced (destructive) RA, they form better models of what bones "should" look like and are generally easier to manually segment. That said, healthy subjects were not entirely free from lesions and abnormalities (e.g. erosions).

A range of atlas selection strategies exist (see section 11.4 in T. Rohlfing's review [29]). This thesis proposes a classifier approach to multi-atlas segmentation, using pairwise registrations between each atlas subject and the unseen patient image followed by label fusion to obtain the final segmentation estimate. Another option for multiple atlases, as described by Sabuncu et al., is to co-register all training subject to a common coordinate space [38]. One then computes a probability map that describes the likelihood of observing a particular label for a given voxel. The idea of fuzzy atlas labels in this case is appealing, especially given the ambiguity of manual labelling. A template image is generated, typically by shape-averaging all atlas images. This would need to be well defined, using statistical shape models for example, in order to maintain anatomical feasibility. Only one registration is then performed to propagate the atlas labels. Similarly, from a pool of atlas images, the best or "most similar" atlas may be selected for registration, to maximise segmentation success. This approach requires the similarities to be known *a priori*. The two alternative multi-atlas implementations fuse the atlas information prior to registration and thus rely on a single registration. This increases the chance of segmentation error due to occasional registration failure. Though computationally more demanding, doing multiple pairwise registrations followed by label fusion is more robust against these failures.

The proposed method uses three subjects to form the atlas. This is the minimum number for a multi-atlas approach to force a majority vote. Increasing the atlas to include five subjects or more can easily be achieved by adding more manually segmented training subjects. Although the computation time will increase with each added atlas subject, the overall segmentation performance will stand to benefit as the atlas' ability to generalise increases.

Atlas creation

The definition of the ground truth, in both atlas images and creating benchmarks for valida-
tion, remains a major underlying challenge in obtaining a satisfactory and validated segmen-
tation algorithm. The ground truth, or benchmark, is manually delineated, and therefore
limited by the data and the person performing the labelling. The very visual limitations this
thesis aims to address (inherent in the human visual system) as well as the many pitfalls that
RAMRIS readers encounter (as mentioned in 1.2 – Limitations of visual scoring) also affect
atlas and benchmark creation. Significant partial volume effects (PVEs) are present due to
the elongated voxels $(0.2\colon 0.2\colon 2.2mm)$. This decreases the ability to accurately resolve bone
boundaries and calls on intuition and anatomical knowledge to complete labelling. In some
cases it is near impossible to place a definitive boundary. This uncertainty factor should be
taken into account to reduce the impact of human judgement in atlas creation.

Lastly, all manual segmentations were performed by the author, who was only briefly
instructed by a rheumatologist on classifying bone in eMRI. Ideally, an expert reader per-
forms all segmentations, or at least verifies them.

Registration

The registration algorithm has to account for differences in appearance between patient and
atlas images. These can arise due to anatomical differences, (random) scanner variation, as
well as global positional and/or postural changes. One source of scanner related variability
is the orientation of the sampling planes. Usually aligned manually by the radiologist prior
to each scan, the sampling planes will "cut" through the wrist at different angles. In combi-
nation with highly anisotropic voxel sizes (i.e. thick slices), appearance differences are ex-
acerbated. In fact, it was observed on multiple occasions that two scans of the same patient
may present entirely differently appearing slices. This was noticeable when trying to propa-
gate the segmentation from one of the atlas images to a follow-up scan of the same subject
(perfect anatomical correspondence) using an affine registration. Though thought to be a
straightforward example, due to misaligned sampling planes, mutual similarity was affected
to the point that no satisfactory segmentation was achievable. This raises a significant limi-
tation inherent in the data and challenges the underlying assumption that if (a section of) a
bone is visible in one scan, then it should be visible in another. Hence, the registration-based
segmentation approach is an ill-posed problem.

In testing the effect of the number of optimiser iterations in the B-spline registration,
the capitate and hamate consistently achieved the greatest segmentation overlap (see Figure
2.10 and Figure 2.15). Coincidentally, these are anatomically the largest carpal bones. Being
a spatial overlap measure, the DSC favours larger object, as they are harder to miss. Hence,
the DSC, though a useful measure, is inherently biased towards large objects. We should
consider this as we may be measuring effects of the segmentation accuracy metric itself, and
not just ABS success. Another explanation may be that larger bones span more slices of the
data volume, meaning the overall impact of spatial variation due to PVEs is decreased. For
larger bones this implies greater atlas-patient co-similarity, leading to improved registration
and thus segmentation quality.

For this work, a Random Coordinate sampler is used in all registrations. It works with a
user-defined number of samples. Hence, there are no benefits (in terms of data reduction)

from down sampling in the multi-resolution pyramid, as only a small subset of voxels are selected in each iteration. Furthermore, the data volumes being registered are less than 10MB in size, so there is no big incentive to reduce the data size, as storage is not an issue. As for computation time, there is added cost due to having the full-sized images at each level of the pyramid, but at the same time, there is time saved in avoiding the down sampling. Using a Random Coordinate sampler, samples are also taken from non-voxel locations, leading to a smoother cost-function and avoiding what is known as the 'grid-effect' (explained by Thévenaz et al. [42]).

Label fusion

Majority voting (MV) is a simple yet effective way of fusing segmentation labels. In this process each voter (the transformed label image, which is a combination of the atlas image and the registration algorithm) is assigned equal weight (one third). However, one can also adopt a strategy that lets more similar atlases carry more weight during label fusion. Based on the image as a whole, this can be a global weight, or, when computed on a fixed neighbourhood, a local-weighted fusion. Exploring more advanced methods of label fusion can help compensate for registration error and improve overall segmentation results. Investigating this topic has been left for future work.

Post-processing

The method of combining the ABS labels with an Otsu threshold image results in a higher certainty of selecting normal bone marrow voxels, thus restricting the segmentation to the sub-cortical bone and out of the synovium. With cleaned-up boundaries, false positive segmentation error and instances where label classes overlap are reduced. However, there are some downsides. Namely, the Otsu segmentation incorporates healthy bone marrow, but ignores cortical bone, which has a low intensity (black) as well as key bone lesions such as erosions and BME (both of a lower intensity) (see Figure 2.18). Each bone is effectively under segmented. Multiplying it by the ABS label field then further restricts the segmentation. We would like to include cortical bone as this has been taken into account in forming the benchmark segmentations. Furthermore, ignoring lesions such as BME directly contradicts the purpose of this work, which is to detect and quantify them. In short, using this post-processing method, false positive error is reduced, but false negative error is introduced.
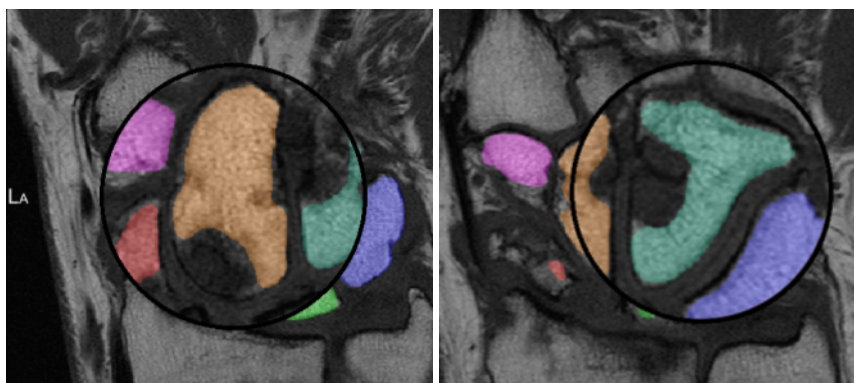
Figure 2.18: The result of merging ABS labels with an Otsu threshold eMRI of the right wrist. Systematic under segmentation and the exclusion of important bone lesions are visible. The capitate (left) and the hamate (right) are magnified. In both cases, large lesions in the bone marrow are not included in the segmentation. Cortical bone is also not correctly labelled due to its low intensity.

2.5 Conclusion

ABS segmentation of the carpal bones using multi-atlas label fusion incorporates prior knowledge including anatomical labels. Although atlas creation is a one-off exercise, it was found to be tedious and time-consuming, and non-trivial. Three training subjects were included to form the atlas, forcing a majority vote decision in fusing the segmentation results. The registration framework provided by the `elastix` toolbox was easily adapted to the wrist images, although objective parameter optimisation is time-consuming and limited to the training set. Four key `elastix` parameters were tested in the B-spline registration step to determine their optimal setting in terms of segmentation quality as well as computation time. The resulting recommended parameter settings are summarised in Table 2.6.

The ABS-algorithm has been shown to automatically detect the carpal bones with a mean segmentation accuracy in terms of the Dice overlap of 0.85. Segmentation errors are predominantly made on the boundaries and will affect subsequent image analysis. Fusing the ABS label field with an Otsu threshold image saw reductions in false positive segmentation error but introduced false negative error when compared to the ground truth; overall segmentation accuracy is reduced slightly with a mean DSC of 0.83. The Otsu post-processing systematically avoids bone lesions, such as BME, which have a dark appearance in pre-contrast T1 MRI. On this basis alone it was deemed inadequate for the purposes of this work.

Table 2.6: Optimal parameter settings, in terms of segmentation accuracy, for the B-spline registration in multi-ABS of the carpal bones. Based on a training set of 8-10 subjects.

`elastix` **parameter** **(B-spline registration)**	**Setting**
Maximum number of iterations	1500-2000
Similarity Measure	Normalized Cross Correlation
B-spline grid spacing	Custom*
Number of spatial samples	2000

*see Table 2.4

3. BME Quantification

3.1 Introduction

This chapter explains the methods used to generate a quantitative measure of BME by means of comparative analysis of pre- and post-contrast MRI. The technique is implemented for the carpal bones, but could easily be extended to other joints or bones, provided they are segmented. In general, the approach taken in quantifying BME largely depends on the type of study undertaken and the available data. For our purposes the data is treated as it would be in a cross-sectional study; looking at a range of patients at a particular point in time. Static data is used (a single scan per visit per sequence) as opposed to dynamic (4-D) data where time is the added dimension (as in DCE-MRI). A brief look at related works in the literature on capturing BME from MRI is provided, before discussing the proposed automated method.

The different techniques used to characterise BME can be categorised based on the type of MRI data used. Most methods using static data focus on capturing BME using morphology (i.e. shape and size). Given the volume of BME is known it can be expressed as a percentage of the total bone volume in which it resides. In the case of dynamic contrast-enhanced MRI (DCE-MRI), perfusion and enhancement rates are observed over a number of time points. These parameters can be used to assess the level of inflammatory activity. First, a number of volumetric approaches to capturing BME are highlighted. Finally, one study is mentioned for comparison that looks at BME using relative enhancement (perfusion).

Roemer et al. approximate the BME volume in MR images of the knee using a simple volumetric measure [43]. By manually drawing three orthogonal lines spanning the lesion's maximal peripheral margins a cubic approximation of BME volume is calculated. This technique relies on human perception and manual drawing to determine BME's often fuzzy boundaries and is not likely to be very reproducible nor accurate. MR systems and sequences were not held constant throughout this study.

Mayerhoefer et al. present a computer-assisted method for the quantification of BME in STIR MRI of the knee [30]. Their aim was to address the lack of reproducibility in prior work on calculating BME volume. For each patient, a region-of-interest (ROI) is placed in three areas of healthy (or normal-appearing) bone marrow. Using the arithmetic mean from each ROI a threshold-value is calculated to label BME voxels for that particular examination. This is necessary as the intensity value for bone marrow fluctuates between patients and even in the same scan. The overall knee joint is manually segmented to measure the total bone volume.

Li et al. [44] (2008) also focused on quantifying BME-like lesions in MR of the knee in patients with osteoarthritis. Their methods are quite similar to that of the Mayerhoefer paper, relying on manual delineation and ROI placement in order to calculate a threshold value that selects BME lesions. For validation a "gold-standard" threshold was created based on the consensus of two radiologists who manually identified the best threshold value. A follow-up paper from the same lead author, published in 2012, looked at BME in the wrist in DCE-MRI. Again, contours are manually placed in normal bone marrow to calculate the standard deviation of intensities. Five times the standard deviation is taken as the threshold

to segment BME. The problem with aforementioned ROI-based approaches is that they rely on manual placement (subjective) and they assume said region is representative of the whole bone marrow, which is not necessarily true.

Leung et al., present their work on bone lesion quantification using serial T1 MR images of the talus bone in rats [45]. Joint inflammation was locally induced in one ankle and compared to the left ankle as a control. They selected potential lesion voxels by applying Otsu's thresholding method to a difference image of the baseline and follow-up scan to detect high-intensity voxels. These voxels were generated across the time series of five images and summed. A bone lesion was only counted if it was found in two or more instances across the series. This method constitutes a true 3-D assessment of pathology and includes a registration framework for segmentation propagation and comparative analysis.

Kubassova et al. made use of DCE-MRI (4-D data) to capture inflammatory lesions in the MCP joints [46]. The degree of inflammation is not measured by lesion volume but by the rate of contrast-agent uptake. Heuristics such as maximum rate of enhancement, initial rate of enhancement, and time of onset of enhancement were used to estimate the total number of enhancing voxels.

For the purposes of this work only static data (one contrast-enhanced image) is available, ruling out the methods used by Leung and Kubassova. Given the other articles lack automation and reproducibility due to their ROI-based approach, we seek a solution of our own.

3.2 Method

The approach of attaining a quantitative BME score per carpal bone is broken down into four steps (see Figure 3.1): 1) the pre- and post-contrast images need to be co-registered. The resultant rigid transformation is used to propagate the carpal bone labels. 2) By examining which intensity values co-occur in the pre- and post-contrast images, three joint-intensity groups were identified, which correspond to the predominant tissue types. Edematous bone is segmented using knowledge-based fuzzy clustering of these joint-intensities. 3) To reduce the effect of segmentation error, a post-processing step was included to remove artefacts. 4) the volume of BME per carpal bone is computed. The high-level process is illustrated in Figure 3.1. A number of optimisation experiments were performed to evaluate the influence of segmentation error on the BME score and its correlation strength with RAMRIS.
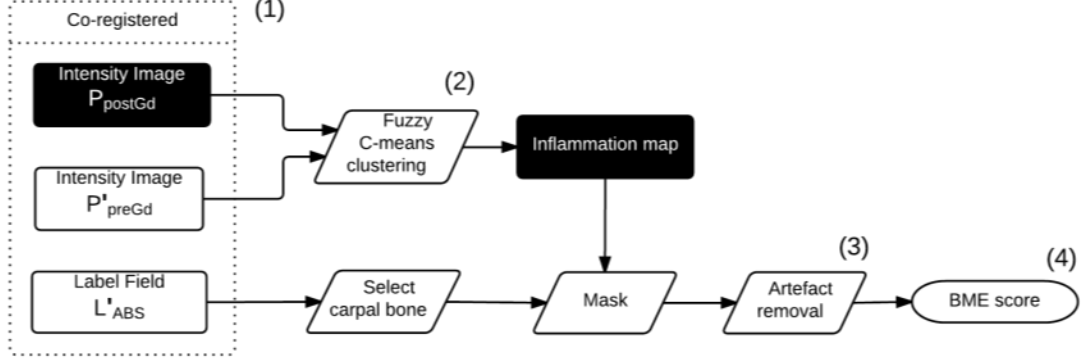
Figure 3.1: The framework for carpal bone BME quantification using co-registered (contrast-enhanced) images of the wrist. Knowledge-based clustering of the joint-intensities generates a probability map that can automatically distinguish BME from normal bone marrow. The ABS is propagated and used to select individual carpal bones and calculate the percentage volume of BME. Artefact removal is added to reduce the effects of segmentation error at the bone boundary.

Registration & segmentation

To enable comparison between pre- and post-contrast images, they must first be registered to the same spatial domain. Computing this registration also allows the ABS to be propagated to the post-contrast image domain. Figure 3.2 shows the high-level registration framework. The registration is different with respect to the ABS in a number of ways, which in turn changes the parameter setting requirements. Firstly, the registration is intra-patient instead of inter-patient. Secondly, given different MR sequences are used (Table 1.2), this registration is akin to having multi-modal data. Therefore MI is used (Advanced Mattes mutual information in `elastix`). Given both wrist images come from the same patient (scanned only minutes apart) we expect to only have to account for positional and postural changes (rotation and translation). This can be corrected using a single rigid registration. In `elastix` this is known as the Euler transform. To compensate for differences in scale as well, a Similarity transform can be chosen. Unless specified otherwise, the geometric centre is automatically chosen by `elastix` as the centre of rotation.

Remaining parameter settings were identical to the rigid registration used in ABS; the adaptive stochastic gradient descent optimiser was used in combination with a random coordinate sampler: 3000 spatial samples were selected for 1000 iterations in a 4-level pyramid with custom Gaussian smoothing as described in 2.2.
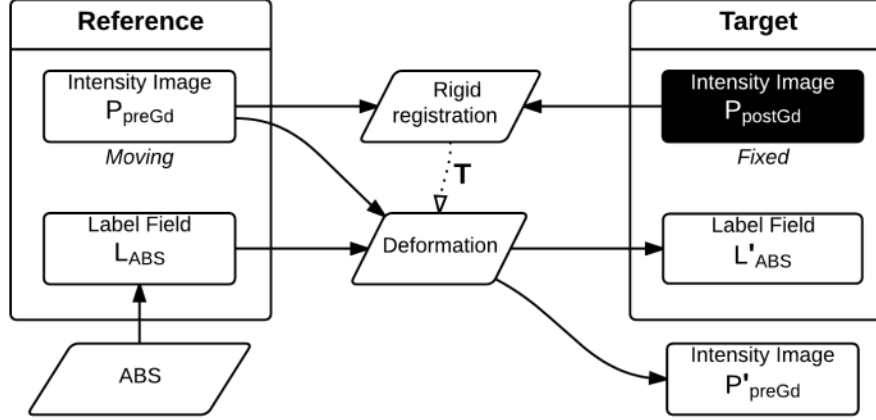
Figure 3.2: Rigid registration framework prior to BME quantification. The post-contrast age P_{PostGd} which is to be segmented is defined as the fixed image. Once the coordinate transformation T is known the ABS labels L_{ABS} as well as the pre-contrast image P_{preGd} are deformed into the coordinate space of the post-contrast image, indicated by L'_{ABS} and P'_{preGd}. Segmentation is therefore obtained by performing one additional transformation (label propagation).

Joint histogram analysis

An automatic tissue classification algorithm is proposed based on joint histogram analysis of pre and post-contrast eMRI of the wrist. A knowledge-based fuzzy clustering method assumes three joint-intensity classes and is able to distinguish normal from edematous bone marrow. Carpal bone voxels classified as BME are counted and expressed as a percentage of total bone volume.

The 2-D joint histogram plots the frequency of co-occurring image intensity values. In the joint histogram of pre- and post-contrast image pair, three clusters are visible (see Figure 3.3). Assuming no artefacts or abnormalities, the main tissue types belong to the following clusters:

1. Air, tendons and cortical bone
2. Muscle and synovia
3. Fat and bone marrow

What characterises these three groups are their joint-intensities. Air, tendons and cortical bone have low intensities in both images, and belong to the cluster 1, near the origin. Muscle and synovial tissue occupy low to mid-grey values on the intensity scale in both images (cluster 2). The third cluster represents normal bone marrow and other fatty tissue, characterised by a bright signal in the pre-contrast image, and a low signal in the fat-suppressed image. Where healthy bone marrow belongs to cluster 3, BME displays a reversal of the expected joint-intensities; it displays a low-intensity signal in the pre-contrast image, and a high-intensity signal in the post-contrast image. Thus BME voxels lie closer to cluster 2 in the joint histogram. In other words, the co-occurrence of intensities of BME behaves more like that of (inflamed) synovial tissue. It is this knowledge that helps us detect BME voxels, given the carpal bone segmentation is known.
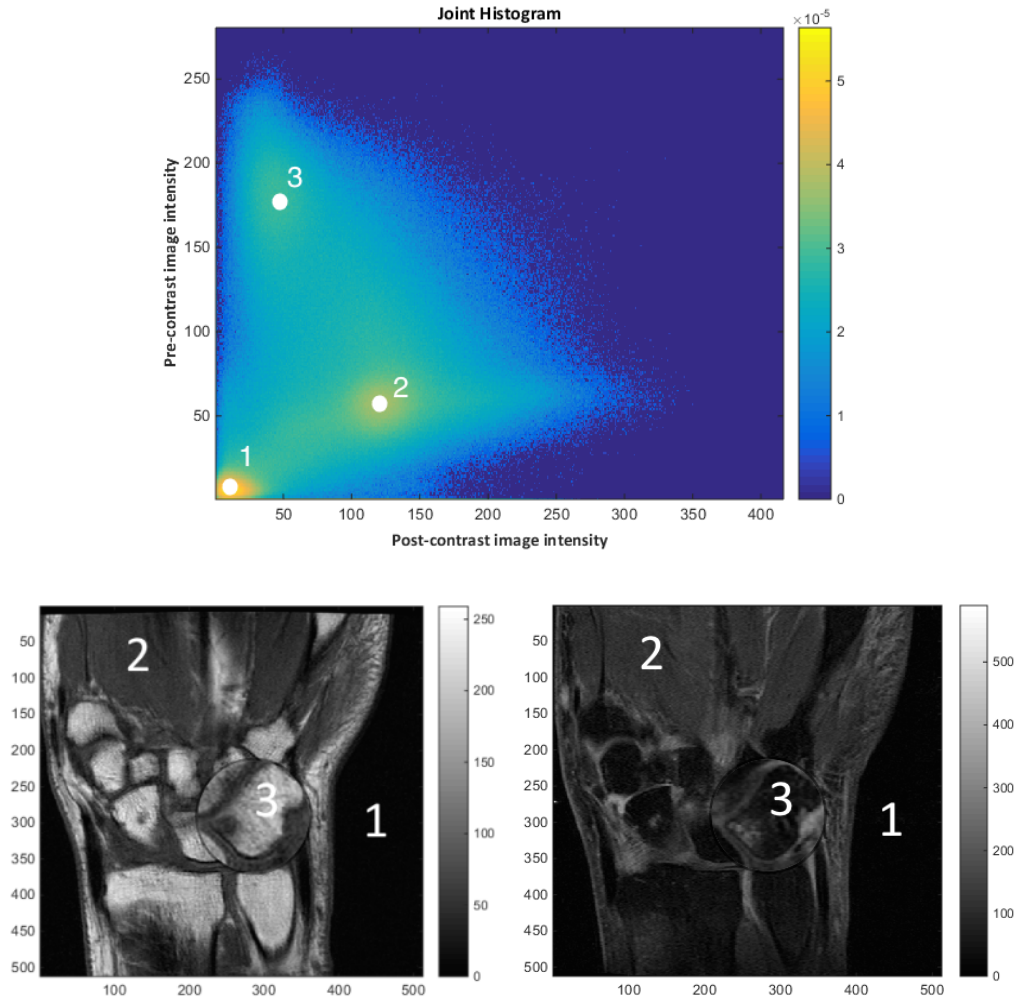
Figure 3.3: Top: joint histogram of pre- and post-contrast eMRI of the wrist (shown bottom left and right respectively). White dots show approximate cluster centres of the three intensity classes: 1 (air, tendons and cortical bone), 2 (muscle and synovia), 3 (fat and bone marrow). BME is present in the triquetrum (magnification) and has a joint-intensity closer to that of cluster 2.

Fuzzy C-means (FCM) clustering was performed in MATLAB (available in version 2014b as part of the Fuzzy Logic Toolbox) based on a three-cluster model. The intensities of the post-contrast image together with the intensities of the co-registered pre-contrast image are provided as inputs. It was empirically determined that 30 iterations were sufficient to generate (and reproduce) the (same) cluster centres. This method generates three membership probability maps, one for each cluster. The map for which the synovium/muscle tissue voxels have the highest probabilities is automatically extracted based on cluster two's expected centre location. In terms of pre-contrast image intensity (y-axis of joint histogram), cluster two is always the middle value. Hence, based on the median value the cluster centre

y-values, the correct cluster is automatically detected. The corresponding probability map to cluster two acts as the inflammation map and is read back from MATLAB for further processing (see Figure 3.4). Membership threshold was set at 0.6 to select potential lesion voxels. Using the inflammation map and the carpal bone segmentation, each bone is masked and then post-processed on an individual basis before the volume of BME is computed.
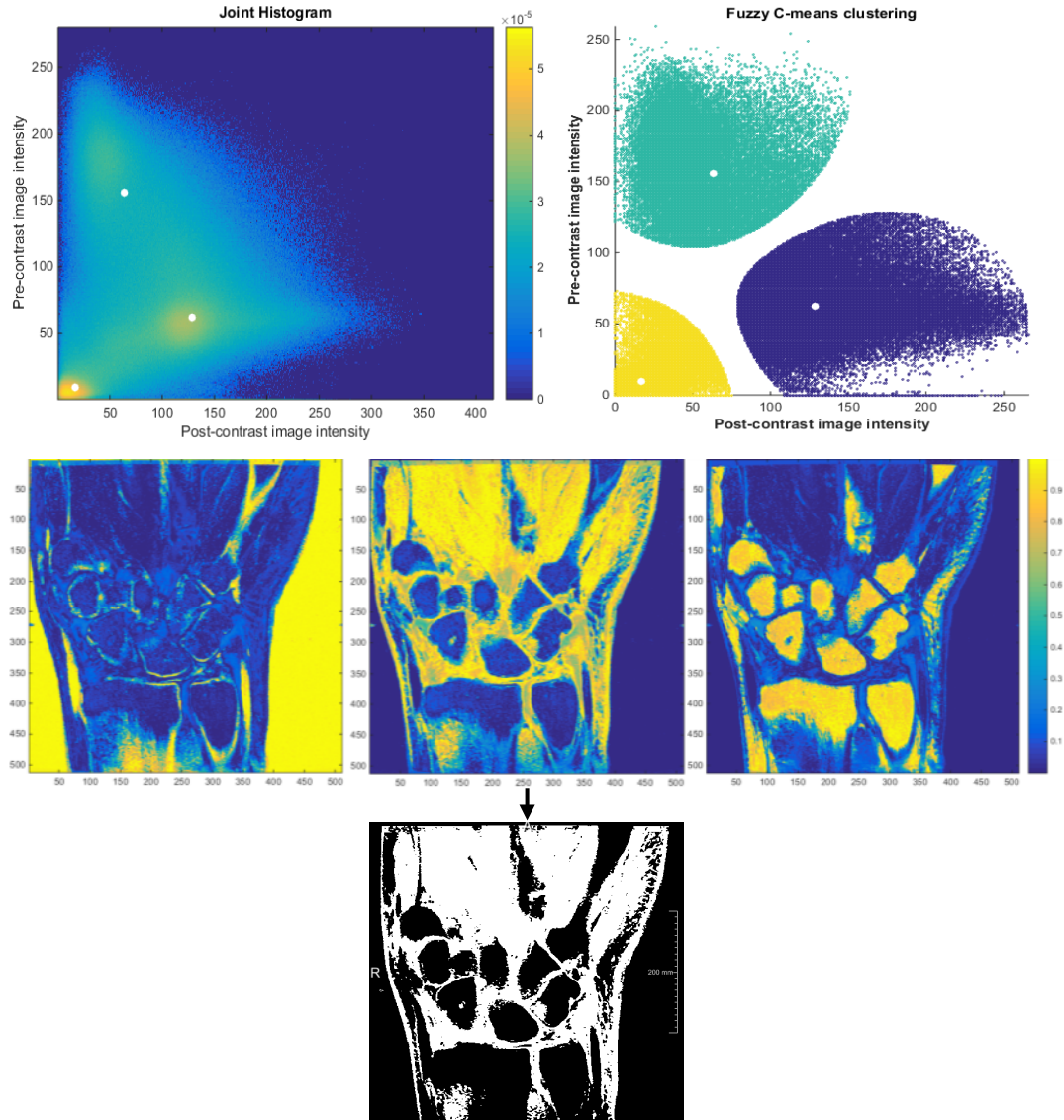
Figure 3.4: Top-left: joint histogram with actual cluster centres from FCM. Top-right: Three clusters with membership threshold of 0.6. Middle: three probability maps resulting from FCM clustering of using three classes. Each channel shows the probability of membership of a certain class: voxels in the image on the left show the probability of belonging to cluster 1 (air, cortical bone, tendons). Correspondingly the middle image displays the likelihood of cluster 2 membership and the image on the

right, cluster 3 membership probability. The middle image is defined as the inflammation map. A threshold is set at 0.6 to segment the normal from abnormal bone-marrow (bottom).

Post-processing

Post-processing steps were implemented to overcome confounding effects due to segmentation error at the bone boundary. Namely, over-segmentation causes part of the synovia to be included, which is of a high intensity, from the inflammation map. They typically appear as long elongated structures along the bone boundary (see Figure 3.5), but are not necessarily continuous. Using an ITK[7] flatness filter, these objects could be removed. The flatness filter is insensitive to the size of the object and is calculated based on the principal moments from the moment-of-inertia tensor (or, equivalently, the covariance matrix of the voxel distribution within each object). This matrix has three eigenvalues, which indicate the size of the major axes of the object. It is defined as the square root of the ratio of the first two eigenvalues:

$$Flatness = \sqrt{e2/e1}.$$

As the name suggest, this filter is sensitive to flat objects, but by incorporating the square root it becomes more sensitive to elongated, ribbon-like structures. This filter operates on a slice-by-slice basis. A threshold value (λ) of 2 was used (a value of 1 removed everything and higher values did not remove enough). After removing the confounding edges, some small fragments remained. Hence, a connected components filter was applied to remove small specks (of 20 voxels or less) using a 3D 6-voxel neighbourhood. An example of the post-processing technique for one carpal bone (hamate) is shown in Figure 3.6.
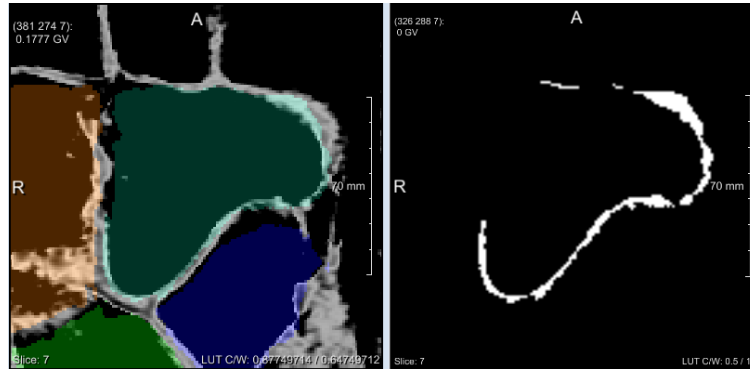
Figure 3.5: The segmentation mask for the hamate extends into the synovium (left). Using ITK Flatness filter, the confounding edge is detected (right) and subsequently removed.

---

[7] Insight Segmentation and Registration Toolkit - www.itk.org

49
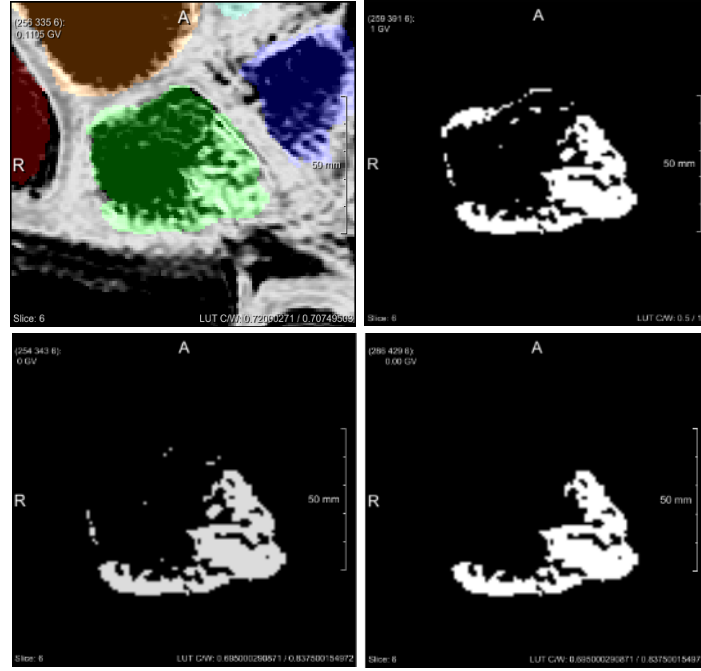
Figure 3.6: From inflammation map to final BME segmentation in the lunate, reading from top-left to bottom-right. The inflammation map with segmentation mask overlayed (top-left). Masked bone with 0.6 membership to the synovium class (2) (top-right). After flatness filter (bottom-left) and after connected components filter (bottom-right). Non-BME voxels along the top edge of the bone are correctly discarded.

BME score

Finally, the BME score is calculated as the volume of BME contained within the bone, divided by the total bone volume (known from the segmentation mask), expressed as a percentage:

$$BME\ Score = \frac{BME\ volume}{Carpal\ bone\ volume} \times 100\% \ .$$

Optimisation

A subset of 15 subjects was selected from the original training set of 29 used to optimise ABS. The added requirement was the presence of both the pre-contrast and post-gadolinium fat-suppressed T1 coronal images of the right wrist as in Table 1.2. Two subjects were excluded due to a different voxel spacing (poor resolution), leaving 13 subjects (10 unique). All 13 patients were from the Early Arthritis cohort (EAC) at the LUMC [47]. A trained RAMRIS reader had scored all subjects for the presence of BME in the carpal bones. These scores served as the gold-standard in evaluating and optimising the automated method.

Manual segmentation of the carpal bones was performed for each subject in the pre-contrast image, using the same technique as described in 2.2 – Atlas creation. These label fields served as benchmarks (BMs), allowing for the ABS to be bypassed and the effect of ABS error on the BME scores to be investigated.

The post-processing (PP) step described above was also bypassed using both ABS and BM segmentation, giving four experiments. See Figure 3.7 for experimental set-up.

All BME scores were automatically computed per carpal bone and averaged per patient. The average percentage BME score per patient was correlated to the average RA MRI score per patient. Given that the RAMRIS is measured on an ordinal scale, the Spearman's rank correlation coefficient was used. This is a non-parametric test that ranks the data in order to test the statistical dependence between two variables.
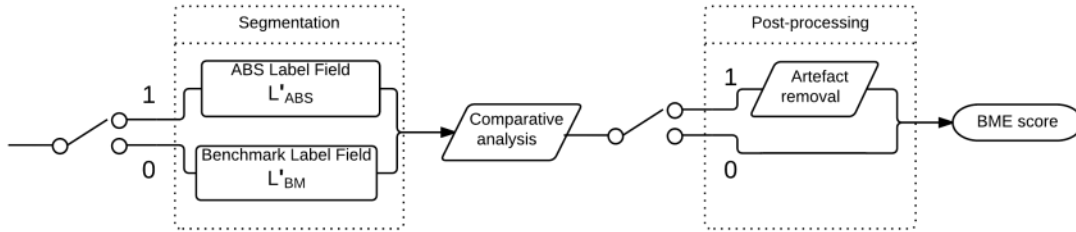
Figure 3.7: Optimisation experiments set-up. For the segmentation mask, either the ABS labels (1) or the manually segmented benchmark (BM) labels (0) can be used (both propagated via a rigid transformation). After comparative analysis, artefact removal (using ITK flatness and Connected components filters) is selected (1) or bypassed (0). This gives a total of four combinations of segmentation and post-processing settings $(00, 01, 10, 11)$, thus four experiments.

3.3 Results

The average scores over all carpal bones, scored both visually (by a trained reader using the RAMRIS system) as well as quantitatively (the proposed method) under four testing conditions, are given in Table 3.1. Scatter plots for each test are provided in Figure 3.8. Corresponding Spearman's rank correlations were computed and are shown in Table 3.2. For more exhaustive results, such as the RAMRIS and quantitative scores per carpal bone, see Appendix B:

Table 3.1: Mean BME scores for all carpal bones in the training set (13 subjects) for 4 test scenarios

Subject	mean RAMRIS	mean % BME			
		test 1 BM	test 2 BM+PP	test 3 ABS	test 4 ABS+PP
EAC01	0.25	11%	3%	18%	4%
EAC02	0.75	15%	3%	15%	3%
EAC03	1	17%	2%	21%	5%
EAC04	0.375	20%	9%	24%	12%
EAC05	0	12%	3%	11%	2%
EAC06	0.625	15%	5%	25%	10%
EAC07	0.125	11%	3%	18%	4%
EAC08	0.625	13%	2%	16%	1%
EAC09	0.25	11%	2%	14%	3%
EAC10	0	11%	4%	18%	5%
EAC11	0	25%	7%	24%	7%
EAC12	0	14%	4%	14%	4%
EAC13	0	6%	0%	5%	0%

Table 3.2: Correlations between RAMRIS and the quantitative BME score for the 4 test scenarios.

			mean % BME			
			test 1 BM	test 2 BM+PP	test 3 ABS	test 4 ABS+PP
Spearman's ρ	mean RAMRIS	Correlation Coefficient	.402	-.224	-.304	.156
		Significance (2-tailed)	.174	.462	.312	.611
		n	13	13	13	13

No significant correlation was obtained at the 95% confidence level. The settings in test 1 (using BM segmentations) rendered BME scores that showed the strongest correlation with the mean RAMRIS, followed by those from test 4. This suggests post-processing is only warranted when ABS labels are used, which was expected, as that is what it was intended for.
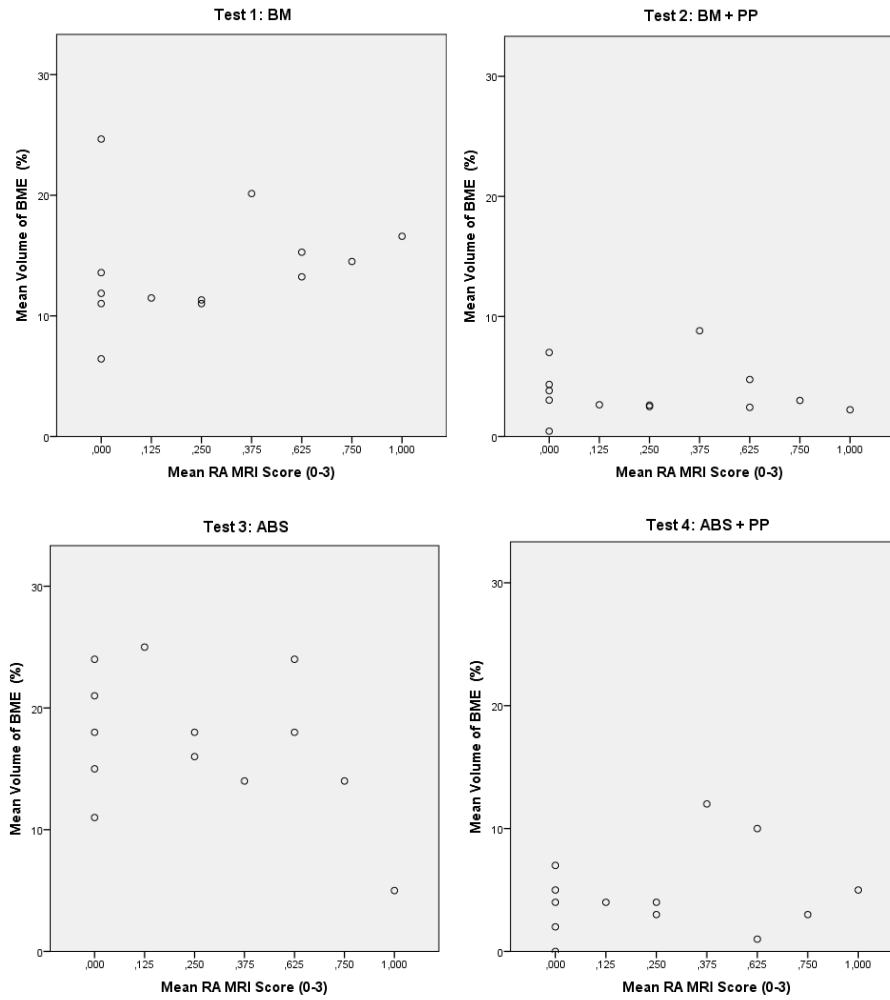
Figure 3.8: Scatter plots for the four experiments using either BM or ABS labels and with or without post-processing (PP).

3.4 Discussion

No significant correlation was found between RAMRIS and the quantitative score for BME in the carpalia. However, the highest correlation with the lowest p-values were found using test 1 and test 4. Respectively, these tests used BM segmentation with no post-processing, or the ABS with post-processing.

Segmentation error

Prior to calculating the percentage of bone volume containing BME, segmentation error is introduced at two points in the algorithm; first, in the ABS process and second, in label propagation to the contrast-enhanced image. By using manual segmentations as bench-

marks, the ABS routine is effectively bypassed. However, segmentation error is still introduced in propagating the labels to the contrast-enhanced image (rigid registration error). The BME score is particularly sensitive to oversegmentation, as this includes voxels belonging to the neighbouring synovium which displays a high intensity, like BME, in the inflammation map. As such, over segmentation places an upward bias on the BME score which is why not many scores of 0% are recorded. It was observed that a small confounding edge in a single slice resulted in a score as high as 15-20% per bone. Whilst the post-processing step performed quite well to reduce this type of error, it is not 100% robust and only treats the symptoms instead of addressing the root cause. Much like in the case of the post-processing method used in ABS, with the removal of false positive error comes the introduction of false negative error. In other words, depending on the morphology of the potential lesion voxels, part of true BME lesion voxels may be removed along with confounding synovium edges.

BME quantification

The presence of noise due to segmentation error overestimates the presence of BME. However, once a BME lesion is correctly identified, it contains many small holes after thresholding. This because BME has a rough texture (high variance) displaying a range of intensity values. An expert reader is likely to round of at the observed edges of the lesion, and count the whole content as one lesion. The proposed automated score however, displays holes within the lesion that are not counted (see Figure 3.6). This results in an underestimation when compared to RAMRIS. Some morphological closing may be needed to count the overall lesion, not just the highest intensity voxels inside it.

In capturing BME a number of assumptions were made: all voxels that are included in the carpal bone segmentation that display cluster 2 joint-intensities are counted as BME voxels. Confounding lesions that show enhancement such as erosions filled with synovial tissue or nutrient vessels are not accounted for. There are other complicated cases where lesions occur contiguously. For example, an erosion is filled by inflamed synovial tissue which lies adjacent to a BME lesion. These cases, though rare, are difficult to assess, even for an expert reader.

Parameter selection

The total BME quantification pipeline has many parameters that are interdependent. Most settings were determined empirically, however, a more formal testing framework is needed to objectively determine optimal settings. Correlation to RAMRIS can be used as an outcome measure for optimisation for example. Better still, by improving segmentation, the final artefact removal steps could be omitted, which greatly reduces the amount of parameters to tune.

Comparing to RAMRIS

Given the ordinal scale of the RAMRIS system, comparison to quantitative scores is limited to non-parametric measures of correlation. Important to consider is that the proposed quantitative approach bases the score entirely on one orientation, namely the coronal view of the wrist. RAMRIS stipulates that lesions should be cross-verified in an opposing plane, offering more data or information to examine. For a fairer comparison, the quantitative method

should be extended to include the transverse plane where available. Interesting to note is that none of the quantitative scores exceed 30% which suggests either RAMRIS overestimated the total carpal bone volume affected by BME, or conversely, the quantitative score underestimated it.

Training set

The training set consisted of 13 subjects, which is a considerably small sample. All subjects were from the Early Arthritis cohort (EAC). This is a select subset of patients that have been diagnosed with early RA. In order to better train the algorithm, healthy subjects and patients in later stages of the disease should be included. For statistical analysis, a larger training set is also desirable as it would lead to more useful insights regarding the true relationship between RAMRIS and the automated score.

3.5 Conclusion

Using the training data (13 subjects), no significant correlation was found between the quantitative BME score and the semi-quantitative RAMRIS even when bypassing the ABS. Segmentation error along the bone boundary causes parts of the synovium to be included, placing an upward bias on the BME score. This error could be corrected using the proposed post-processing technique. Having optimised the BME quantification process with respect to the training set, validation is sought using an independent test set. This experiment is described in the next chapter.

4. Validation Study

4.1 Method

An independent test set of 29 subjects was compiled using persons from three different cohorts: 6 healthy volunteers (ATLAS cohort), 7 patients with Clinically Suspected Arthralgia (CSA) and 16 from the Early Arthritis Cohort (EAC). Subjects were selected ensuring the wrist images had been RAMRIS evaluated for BME in the carpal bones to allow for validation. Furthermore, the algorithm demands coronal images of the right wrist with the same pulse sequences as in Table 1.2. The images of the patients were placed in a structured database to facilitate automated analysis and image processing. The proposed algorithm from ABS to BME quantification was applied to all 29 subjects.

To answer the question of whether a relationship exists between RAMRIS and the quantitative scores, the mean BME scores per patient were correlated. Furthermore, to see if either scoring system can determine cohort membership the means of each patient group were compared. Finally, to see which carpal bones were most affected by BME, the average scores per bone were examined.

4.2 Results

Comparing mean scores per patient

Complete results can be found in Appendix B: Supplementary Results (Table B.6 - page 98 and Table B.7 page 99 for RAMRIS and quantitative scores respectively). In Table 4.1 the Spearman's correlation coefficient is shown, computed on the mean scores per patient. The corresponding scatter plot is given in Figure 4.1.

Table 4.1: Spearman's rank coefficient between the mean RAMRIS and mean % BME scores.

			mean % BME
Spearman's ρ	mean RAMRIS	Correlation Coefficient	.618[**]
		Significance (2-tailed)	.000
		n	29

[**]Correlation is significant at the 0.01 level (2-tailed).
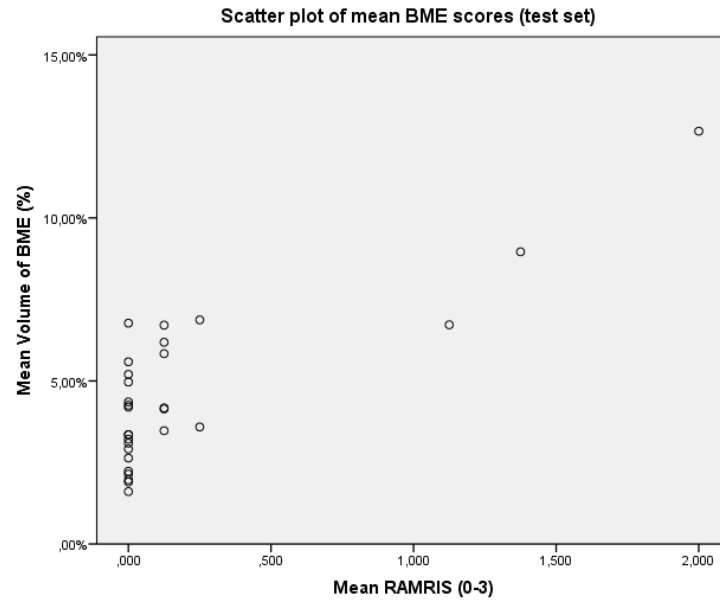
Figure 4.1: Scatter plot of the mean scores (RAMRIS vs. volume of BME)

Comparing mean scores per cohort

Table 4.2 states the mean scores per patient for both scoring systems, when grouped by co-hort. The corresponding box-and-whisker plots are given in Figure 4.2 and Figure 4.3.

Table 4.2: A summary of mean scores per cohort for RAMRIS and the percentage BME score.

Cohort		mean RAMRIS	mean % BME
ATL	mean	.0833	3.49%
	n	6	6
	Std. Deviation	.102	0.74%
CSA	mean	.0179	3.47%
	n	7	7
	Std. Deviation	.047	1.70%
EAC	mean	.320	5.49%
	n	16	16
	Std. Deviation	.612	2.72%
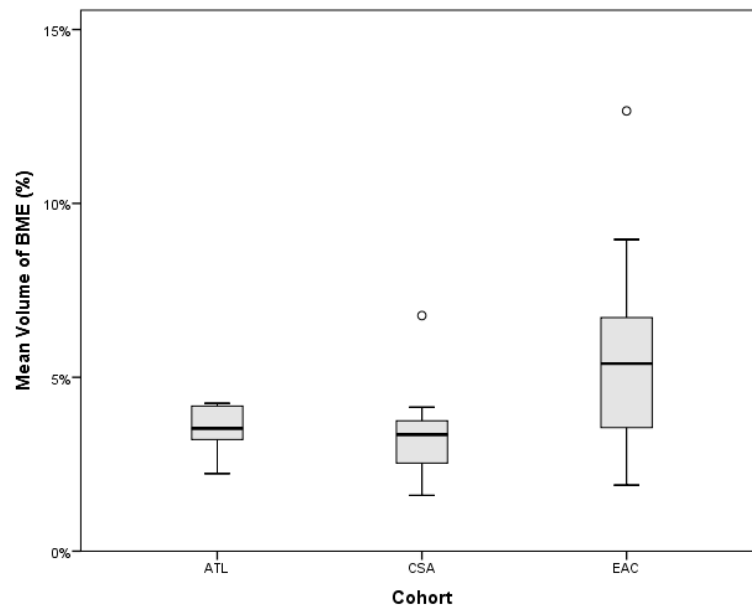
Figure 4.2: A box-and-whisker plot for the automated BME score in the carpal bones. Average scores per patient in the test set (29 subjects) are given and grouped by cohort.
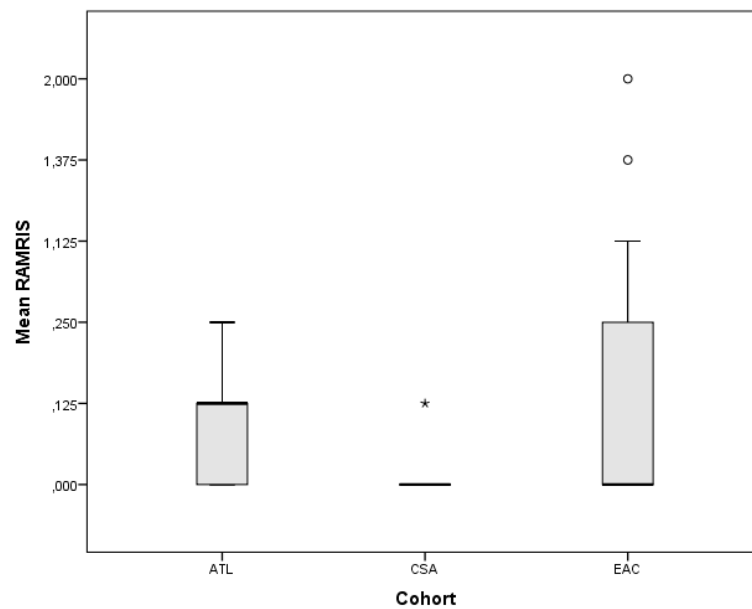
Figure 4.3: A box-and-whisker plot for BME RAMRIS in the carpal bones. Average scores per patient in the test set (29 subjects) are given and grouped by cohort.

Comparing mean scores per carpal bone

The mean scores per carpal bone, using both scoring systems, are shown in Table 4.3.

Table 4.3: Mean scores per carpal bones (RAMRIS and percentage BME volume)

Carpal bone	mean RAMRIS	mean % BME
1-Sca	.172	6%
2-Lun	.345	8%
3-Tri	.276	5%
4-Pis	.172	4%
5-Tpm	.069	6%
6-Tpd	.172	3%
7-Cap	.276	3%
8-Ham	.103	2%
Total	.198	5%

Qualitative results

In order to gain more insight into the algorithm's behaviour, the following qualitative observations were made:

- Oversegmentation leads to erroneous inclusion of synovium leading to an upward bias in the BME score
- Undersegmentation is observed, wheres (large) lesions near the bone edge are missed as their intensities mimic that of synovium, leading to a downward bias in the BME score
- In most cases the post-processing significantly suppresses the effect of segmentation error, but some (small) areas remain.
- Parts of BME lesions are also removed by the flatness filter and connected component filter. In the worst case, "flat" BME-like lesions are removed entirely. Hence, the filter is not specific enough in combatting oversegmentation and will also reduce the true BME score.

4.3 Discussion

Comparing mean scores per patient

A positive monotonic correlation (Spearman's rho = 0.618, n=29, $P < 0.01$) is observed between RAMRIS and proposed quantitative measure of BME in the carpal bones. This indicates the scores, on average, are in good agreement. In its present form, the quantitative score is affected by noise due to oversegmentation. Using the post-processing technique,

most of this noise can be suppressed. For most carpal bones in the test set, a RAMRIS score of 0 was given (see Table B.7 in Appendix B); the corresponding quantitative score calculates a BME scores between 0-16%. It is important for the algorithm to be able to distinguish between healthy and diseased carpal bones. From this validation experiment however, no clear threshold value was visible. Some bones with a RAMRIS of 1, scored 3%, yet another with the same RAMRIS scored 26%. The quantitative approach is more sensitive, but there is also more variability in the scores.

Comparing mean scores per cohort

Generally speaking, we expect subjects in the Atlas cohort to display less RA-related features such as BME. The CSA group would be more at a risk and the EAC the most likely to display BME in the carpal bones. Comparing the means of the two scoring systems, this expectation is not met as CSA subjects displayed less BME in the carpal bones than subjects from the Atlas cohort. Given only 6 and 7 subjects were examined respectively, this could simply be due to the (random) sampling of subjects. Ranking the cohorts by the mean BME score, from least to most affected, the two scoring systems show agreement. With an unequal number of subjects from each cohort and few samples in general, further conclusions are difficult to make from this data.

Comparing mean scores per carpal bone

In order to gain an understanding of which carpal bones were most affected by BME, the mean scores per carpal bone for both RAMRIS and the automated score are given in Table 4.3. With only eight joint sites, there are insufficient data points to perform a statistical correlation. The average RAMRIS scores of each bone varied between .069-.345 with a mean of .198, whereas the quantitative scores varied between 2-8% with a mean of 5%. On average, the lunate was the most highly scored bone for BME in both scoring systems. Two studies in the literature give us an indication of what bones are typically most affected by BME. Peterfy et al. provide a review of MRI findings from four multicentre randomised controlled trials, using data from 522 RA patients [27]. Results showed that BME (referred to as osteitis) was most frequently scored in the navicular (scaphoid), lunate and capitate (in that order). In a study of 42 individuals with early RA (median symptom duration of 4 months) by McQueen et al. the lunate, triquetrum and capitate were the most common joint sites for BME. If we take these four bones (scaphoid, lunate, triquetrum and capitate) as highest scoring bones (for BME), we see that the RAMRIS scores for the test set agree with all four and the quantitative scores agree with three of the four.

4.4 Conclusion

The proposed automated BME score shows a strong positive correlation to the visual scores according to the OMERACT RAMRIS standard (Table 4.1: Spearman's rho = 0.618, n=29, $P < 0.01$). Furthermore, the scoring systems show good agreement on which cohorts and carpal bones were most affected by BME. We expect the two systems to agree, as they measure the (proportional) volume of the bone affected by BME. However, the quantitative approach does not make use of an additional viewing plane (i.e. the transversal wrist data).

5. General Discussion and Recommendations

This thesis has presented an image processing algorithm that automatically detects the carpal bones from eMRI of the wrist and measures the amount of BME present in each joint as a percentage of total bone volume. Validation using an independent test set shows a definitive correlation to RAMRIS. In this chapter the main findings are summarised, key strengths and limitations are identified and recommendations are given for future work.

5.1 Strengths & Weaknesses

This study is novel in the sense that it provides fully automated analysis, aside from the preparatory manual labelling involved in atlas creation. A set of unseen patient images can be easily processed in batches, provided they are of the same MR sequences and stored in a database with the same predefined structure. The absence of human interaction makes this method an objective one, overcoming the reproducibility and variability issues inherent in other computer-assisted techniques. Furthermore, this algorithm provides a truly quantitative measure of MRI BME.

Using the `elastix` registration framework, optimised parameter files have been devised for the coronal wrist images, maximising registration and in turn, segmentation accuracy. The feasibility of ABS for the carpal bones has been demonstrated. Initial results are encouraging and it has been observed that limited spatial resolution in the underlying data ultimately restricts the segmentation and quantification accuracy. This finding has changed the course of the encompassing early RA project (coined ESMIRA) where efforts are being made to fuse transversal and coronal data. The intention is to obtain high resolution data with isotropic voxels for more reliable quantification.

The technique of capturing BME maximises the available information by comparing both the pre- and post-contrast images. A knowledge-based fuzzy clustering technique over all image intensities robustly counts the voxels that belong to BME. This is an improvement on the ROI-based methods seen in related works, as they require manual placement and work on the underlying assumption that the ROI is representative of normal bone marrow.

The algorithm has shown to be robust and adaptive to signal intensity variance between and across patients, even without bias field correction or intensity normalisation. This is indicated by the fact that there were no significant segmentation failures or outliers.

Overall the proposed algorithm makes an important step toward fully automatic BME quantification for the early detection of RA.

The main limitation of the system is that the segmentation error significantly confounds the BME metric. This error arises from the large spatial variation between atlas and patient largely owing to thick slices in the MRI data. Segmentation error is carried forward through to the final score. Registration can only do so much to correct for spatial variation and cannot account for missing information.

Related to the above is that the scope of this project was very narrow; the decision was made to focus solely on BME in the carpal bones. Based on its superior resolution, the coronal scans were selected for segmentation and subsequent quantification. This ignores the available contrast-enhanced transverse images from the same patients. More is said on this matter in section 5.4 - Recommendations.

This technique may introduce some atlas bias to the BME measurement. Although three subjects are included in the atlas, they all model normal wrist anatomy. Therefore the system may not be as well suited to segmenting advanced RA subjects that have more extensive joint damage. However the overall algorithm would stand to benefit if lesions and other parts of the wrist anatomy were included in the model. For example, by adding a label class for the synovium, or learning the background structures from the data.

In atlas creation only the bone marrow and cortical bone belonging to the carpal bones are delineated. This is to determine the correct size and location of each bone. Bone size is used to calculate the final BME score, which is a percentage of total bone volume. However a trade-off exists in including cortical bone or not in the manual labelling. By including it, a more accurate estimate of the total bone volume is made. However, by ignoring cortical bone, the likelihood of the segmentation encroaching onto the synovium, which confounds the BME score, would be reduced. Considering OMERACT defines BME to be "a lesion within the trabecular bone", it could be argued that cortical bone can be omitted from the atlas for the purposes of detecting BME.

Only a small number (13) and a select set (Early Arthritis Cohort) of subjects were used to train the ABS and the BME quantification sub-parts. Hence, some overfitting is expected and this decreases the model's ability to generalise beyond the training set.

5.2 Characterising BME

This thesis, as well as other works that use static MRI, essentially treat BME as a lump. However given its inflammatory nature, functional information from DCE-MRI may provide more insight and sensitivity to BME. The more vascularised a certain tissue, the more contrast-agent is absorbed. Bone marrow vascularity, instead of the proportion of bone affected, might constitute a more sensitive characterisation of this lesion. In a study by Hodgson et al., there were indications that the relative enhancement rate is more responsive to treatment than RAMRIS BME scores [48]. These findings were echoed by Cimmino et al. who in a small study of ten RA subjects found no change in RAMRIS after treatment, whereas the clinical Disease Activity Score (DAS) as well as the DCE-MRI derived synovitis scores both decreased and correlated well with each other [49]. Unfortunately the ONI MSK Extreme is not capable of DCE-MRI.

5.3 Clinical Implications

Although a lot of advancements have been made, "the concept of using MRI as an outcome measure in RA remains 'young' " [50]. The clinical relevance of RAMRIS is yet to be fully established and more research is needed to reveal the exact relationship with other disease measures. Until such time the exact role of BME as a measure of RA disease activity also remains unclear. Although undoubtedly related to RA, its specificity may be limited as patients in remission still show BME (sub-clinical inflammation) [51]. In patient assessment it is likely that an aggregation of factors need to be gauged, including bio-clinical as well as imaging disease markers. Eventually, if fully quantitative methods are to replace visual scoring, longitudinal studies will be needed to evaluate their ability to predict disease progression and their responsiveness to therapeutic regimens.

On the commercial side of things, time is a crucial factor in moving hand scanning into clinical practice. In a move towards fast and accessible assessment of inflammation in RA patients, Hemics BV[8] has introduced a new type of optical hand scanner. Using a light source and two cameras, the speed and magnitude of blood pooling in the hands and wrists are recorded. In less than two minutes, a 2-D image is generated to distinguish healthy from inflamed tissue and an optical score is assigned to each joint. This small and affordable unit that allows for wider deployment and more frequent scanning to monitor disease progression, without the need for a rheumatologist. Although a promising alternative, no studies could be found in the literature that can validate its efficacy.

5.4 Recommendations

- The most significant factor holding back the success of this algorithm is a material one; the eMRI data is highly anisotropic with so-called matchstick-shaped voxels. Including the transverse scan would increase the amount of available information and make for a fairer comparison to RAMRIS. For example, using super-resolution reconstruction (SSR) the coronal and transversal volumes can be fused to give high-resolution data with isotropic voxels. Given the imaging protocol at the LUMC however, this is only possible using the contrast-enhanced images, meaning the quantification of BME using joint-histogram clustering could not be used.

- Segmentation, using the proposed framework, is essentially a label fusion problem. Seeing as registration errors often occur locally, a locally-weighted fusion technique would be worth investigating, whereby the weights allocated to each voter depend on local registration performance (i.e. similarity).

- Given the limitations of the Dice overlap score, an additional/alternative metric to evaluate segmentation success with respect to a ground truth could provide a more complete view on segmentation performance. A mean surface distance measure is one example. Alternatively, additional overlap metrics such as precision and recall (sensitivity) can give more insight.

- In an application to organ segmentation in MR, Berendsen et al. described the benefits of integrating statistical shape models into the cost-function as a penalty term in non-rigid registration [52]. Including statistical shape models of the carpal bones (see van de Giessen et al. [53]) could improve the amount of prior knowledge captured by the model and thus better predict the expected deformations.

- The proposed method of capturing BME from the image data could be extended, but not before segmentation accuracy is addressed. Local statistical measures such as the variance could be used as a measure of texture or fuzziness which is typical of BME.

---

[8] http://www.hemics.com/technology

5.5 Conclusion

In conclusion, early diagnosis and aggressive treatment is the best chance of achieving re-mission in RA patients. MRI is capable of capturing key pathophysiological RA features at an early stage. Quantitative characterisation of MRI BME aims to increase sensitivity and reduce the subjectivity that currently undermines visual RA MRI evaluation. This thesis has described an automatic method that captures bone marrow edema in the carpal bones using eMRI. Partial-volume effects due to anisotropic voxel spacing significantly reduce similari-ty between patient images. This limits the registration quality and in turn leads to segmenta-tion error. This error is carried forward and confounds the sensitive BME score. Post-processing was able to curb the impact of this effect, and a promising correlation was found with the existing RAMRIS system. For future work: the inclusion of transversal data would increase the available information, make for a fairer comparison to RAMRIS and improve the overall accuracy of the algorithm.

References

[1]   M. Østergaard and M. Szkudlarek, "Imaging in rheumatoid arthritis – why MRI and ultrasonography can no longer be ignored," *Scand. J. Rheumatol.*, vol. 32, no. 2, pp. 63–73, Jan. 2003.

[2]   I. Blankers, A. Notenboom, and C. Schrijvershof, "Zorgkosten en maatschappelijke baten van reumatherapieën bij reumatoïde artritis," Den Haag, 2012.

[3]   J. M. C. Albers, H. H. Kuper, P. L. C. M. Van Riel, M. L. L. Prevoo, M. A. Van 'T Hof, A. M. Van Gestel, and J. L. Severens, "Socio-economic consequences of rheumatoid arthritis in the first years of the disease," *Rheumatology*, vol. 38, pp. 423–430, 1999.

[4]   P. M. J. Welsing, M. Bijl, A. A. van Bodegraven, W. F. Lems, E. Prens, and J. W. J. H. Bijlsma, "Cost effectiveness of biologicals: high costs are the other face of success," *Ned. Tijdschr. Geneeskd.*, vol. 155, p. A3026, 2011.

[5]   S. F. Kengetallen, "Data en feiten 2011 gereed," *Pharm. Weekbl.*, vol. 146, p. 13, 2011.

[6]   M. P. M. van der Linden, S. Le Cessie, K. Raza, D. Van Der Woude, R. Knevel, T. W. J. Huizinga, and A. H. M. van der Helm-van Mil, "Long-term impact of delay in assessment of early arthritis patients.," *Arthritis Rheum.*, pp. 1–28, 2010.

[7]   J. Narváez, J. Narváez, E. De Lama, and M. De Albert, "MR imaging of early rheumatoid arthritis," *Radiographics*, vol. 30, no. 1, p. 143, 2010.

[8]   J. van Aken, L. R. Lard, S. le Cessie, J. M. W. Hazes, F. C. Breedveld, and T. W. J. Huizinga, "Radiological outcome after four years of early versus delayed treatment strategy in patients with recent onset rheumatoid arthritis.," 2004.

[9]   M. A. Quinn and S. Cox, "The evidence for early intervention.," *Rheum. Dis. Clin. North Am.*, vol. 31, pp. 575–589, 2005.

[10]   J. a B. van Nies, A. Krabben, J. W. Schoones, T. W. J. Huizinga, M. Kloppenburg, and a H. M. van der Helm-van Mil, "What is the evidence for the presence of a therapeutic window of opportunity in rheumatoid arthritis? A systematic literature review.," *Ann. Rheum. Dis.*, pp. 861–870, 2013.

[11]   M. P. M. van der Linden, S. Le Cessie, K. Raza, D. Van Der Woude, R. Knevel, T. W. J. Huizinga, and A. H. M. van der Helm-van Mil, "Long-term impact of delay in assessment of early arthritis patients.," *Arthritis Rheum.*, pp. 1–28, 2010.

[12]   L. C. Vasanth, H. Pavlov, and V. Bykerk, "Imaging of rheumatoid arthritis.," *Rheum. Dis. Clin. North Am.*, vol. 39, no. 3, pp. 547–66, Aug. 2013.

[13] F. M. McQueen, N. Stewart, J. Crabbe, E. Robinson, S. Yeoman, P. L. Tan, and L. McLean, "Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset.," 1998.

[14] L. Casimiro, L. Brosseau, V. Robinson, S. Milne, M. Judd, G. Well, P. Tugwell, and B. Shea, "Therapeutic ultrasound for the treatment of rheumatoid arthritis.," *Cochrane Database Syst. Rev.*, p. CD003787, 2002.

[15] P. Bøyesen, E. A. Haavardsholm, M. Østergaard, D. van der Heijde, S. Sesseng, and T. K. Kvien, "MRI in early rheumatoid arthritis: synovitis and bone marrow oedema are independent predictors of subsequent radiographic progression," *Ann. Rheum. Dis.*, vol. 70, no. 3, pp. 428–433, 2011.

[16] A. Duer-Jensen, K. Hoørslev-Petersen, M. L. Hetland, L. Bak, B. J. Ejbjerg, M. S. Hansen, J. S. Johansen, H. M. Lindegaard, H. Vinterberg, J. M. Moøller, and M. Oøstergaard, "Bone edema on magnetic resonance imaging is an independent predictor of rheumatoid arthritis development in patients with early undifferentiated arthritis," *Arthritis Rheum.*, vol. 63, pp. 2192–2202, 2011.

[17] E. A. Haavardsholm, M. Østergaard, H. B. Hammer, P. Bøyesen, A. Boonen, D. van der Heijde, and T. K. Kvien, "Monitoring anti-TNFalpha treatment in rheumatoid arthritis: responsiveness of magnetic resonance imaging and ultrasonography of the dominant wrist joint compared with conventional measures of disease activity and structural damage.," *Ann. Rheum. Dis.*, vol. 68, pp. 1572–1579, 2009.

[18] M. Østergaard, C. Peterfy, P. Conaghan, F. McQueen, P. Bird, B. Ejbjerg, R. Shnier, P. O'Connor, M. Klarlund, P. Emery, H. Genant, M. Lassere, and J. Edmonds, "OMERACT rheumatoid arthritis magnetic resonance imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system," in *Journal of Rheumatology*, 2003, vol. 30, pp. 1385–1386.

[19] M. Østergaard, J. Edmonds, F. McQueen, C. Peterfy, M. Lassere, B. Ejbjerg, P. Bird, P. Emery, H. Genant, and P. Conaghan, "An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas.," *Ann. Rheum. Dis.*, vol. 64 Suppl 1, pp. i3–7, Feb. 2005.

[20] B. Ejbjerg, "The EULAR-OMERACT rheumatoid arthritis MRI reference image atlas: the wrist joint," *Annals of the Rheumatic Diseases*, vol. 64. pp. i23–i47, 2005.

[21] F. McQueen, M. Østergaard, C. Peterfy, M. Lassere, B. Ejbjerg, P. Bird, P. O'Connor, H. Genant, R. Shnier, P. Emery, J. Edmonds, and P. Conaghan, "Pitfalls in scoring MR images of rheumatoid arthritis wrist and metacarpophalangeal joints.," *Ann. Rheum. Dis.*, vol. 64 Suppl 1, pp. i48–55, Feb. 2005.

[22] E. A. Haavardsholm, M. Østergaard, B. J. Ejbjerg, N. P. Kvan, T. A. Uhlig, F. G. Lilleås, and T. K. Kvien, "Reliability and sensitivity to change of the OMERACT rheumatoid arthritis

magnetic resonance imaging score in a multireader, longitudinal setting," *Arthritis Rheum.*, vol. 52, pp. 3860–3867, 2005.

[23]   M. Fiona, "Bone marrow edema and osteitis in rheumatoid arthritis: the imaging perspective.," *Arthritis Res. Ther.*, vol. 14, no. 5, 2012.

[24]   R. J. Hodgson, P. O'Connor, and R. Moots, "MRI of rheumatoid arthritis - Image quantitation for the assessment of disease activity, progression and response to therapy," *Rheumatology*, vol. 47. pp. 13–21, 2008.

[25]   M. Hetland, B. Ejbjerg, K. Hørslev-Petersen, S. Jacobsen, A. Vestergaard, A. Jurik, K. Stengaard-Pedersen, P. Junker, T. Lottenburger, I. Hansen, L. Andersen, U. Tarp, H. Skjødt, J. Pedersen, O. Majgaard, A. Svendsen, T. Ellingsen, H. Lindegaard, A. Christensen, J. Vallø, T. Torfing, E. Narvestad, H. Thomsen, M. Ostergaard, and C. study group, "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.*, vol. 68, no. 3, pp. 384–390, 2009.

[26]   E. Haavardsholm, P. Bøyesen, M. Østergaard, A. Schildvold, and T. Kvien, "Magnetic resonance imaging findings in 84 patients with early rheumatoid arthritis: bone marrow oedema predicts erosive progression," *Ann. Rheum. Dis.*, vol. 67, no. 6, pp. 794–800, 2008.

[27]   C. G. Peterfy, P. Countryman, A. Gabriele, T. Shaw, A. Anisfeld, W. Tsuji, E. Olech, N. B. Gaylis, P. G. Conaghan, V. Strand, and J. Dicarlo, "Magnetic resonance imaging in rheumatoid arthritis clinical trials: emerging patterns based on recent experience.," *J. Rheumatol.*, vol. 38, pp. 2023–2030, 2011.

[28]   W. Stomp, A. Krabben, D. van der Heijde, T. W. J. Huizinga, J. L. Bloem, A. H. M. van der Helm-van Mil, and M. Reijnierse, "Aiming for a shorter rheumatoid arthritis MRI protocol: can contrast-enhanced MRI replace T2 for the detection of bone marrow oedema?," *European Radiology*, 2014.

[29]   T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer, "Quo Vadis , Atlas-Based Segmentation ?"

[30]   M. E. Mayerhoefer, M. Breitenseher, S. Hofmann, N. Aigner, R. Meizer, H. Siedentop, and J. Kramer, "Computer-assisted quantitative analysis of bone marrow edema of the knee: initial experience with a new method.," *AJR. Am. J. Roentgenol.*, vol. 182, pp. 1399–1403, 2004.

[31]   F. Bergeest, JP, Jäger, "A Comparison of Five Methods for Signal Intensity Standardization in MRI," *Methods*, pp. 2–6, 2008.

[32]   S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "elastix: a toolbox for intensity-based medical image registration.," *IEEE Trans. Med. Imaging*, vol. 29, pp. 196–205, 2010.

[33]   S. Klein, U. a. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Med. Phys.*, vol. 35, no. 4, p. 1407, 2008.

[34]   P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration.," *IEEE Trans. Image Process.*, vol. 9, pp. 2083–2099, 2000.

[35]   S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines.," *IEEE Trans. Image Process.*, vol. 16, pp. 2879–2890, 2007.

[36]   X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz-de-Solórzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imaging*, vol. 28, pp. 1266–1277, 2009.

[37]   T. Rohlfing, D. B. Russakoff, and C. R. Maurer, "Expectation maximization strategies for multi-atlas multi-label segmentation.," *Inf. Process. Med. Imaging*, vol. 18, pp. 210–221, 2003.

[38]   M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Trans. Med. Imaging*, vol. 29, pp. 1714–1729, 2010.

[39]   P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, pp. 1116–1128, 2006.

[40]   S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *Int. J. Comput. Vis.*, vol. 81, pp. 227–239, 2009.

[41]   N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[42]   B. M. U. M. Thévenaz P., "Halton sampling for image registration based on mutual information," *Sampl. Theory Signal Image Process.*, vol. 7, pp. 141–171, 2008.

[43]   F. W. Roemer and K. Bohndorf, "Long-term osseous sequelae after acute trauma of the knee joint evaluated by MRI," *Skeletal Radiol.*, vol. 31, pp. 615–623, 2002.

[44]   X. Li, B. C. Ma, R. I. Bolbos, R. Stahl, J. Lozano, J. Zuo, K. Lin, T. M. Link, M. Safran, and S. Majumdar, "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*, vol. 28, pp. 453–461, 2008.

[45]   K. K. Leung, M. Holden, N. Saeed, K. J. Brooks, J. B. Buckton, A. A. Williams, S. P. Campbell, K. Changani, D. G. Reid, and Y. Zhao, "Automatic quantification of changes in

bone in serial MR images of joints," *Med. Imaging, IEEE Trans.*, vol. 25, no. 12, pp. 1617–1626, 2006.

[46]   O. Kubassova, R. Boyle, and A. Radjenovic, "Quantitative analysis of dynamic contrast-enhanced {MRI} datasets of the metacarpophalangeal joints," 2007.

[47]   D. P. C. de Rooy, M. P. M. van der Linden, R. Knevel, T. W. J. Huizinga, and A. H. M. van der Helm-van, "Predicting arthritis outcomes-what can be learned from the Leiden Early Arthritis Clinic?," *Rheumatology*, vol. 50, pp. 93–100, 2011.

[48]   R. Hodgson, A. Grainger, P. O'Connor, T. Barnes, S. Connolly, and R. Moots, "Dynamic contrast enhanced MRI of bone marrow oedema in rheumatoid arthritis.," *Ann. Rheum. Dis.*, vol. 67, pp. 270–272, 2008.

[49]   M. A. Cimmino, M. Parodi, G. Zampogna, M. Boesen, O. Kubassova, F. Barbieri, F. Paparo, G. Garlaschi, and M. Cutolo, "Dynamic contrast-enhanced, extremity-dedicated MRI identifies synovitis changes in the follow-up of rheumatoid arthritis patients treated with rituximab.," *Clin. Exp. Rheumatol.*, vol. 32, no. 5, pp. 647–652, 2014.

[50]   A. H. M. van der Helm-van Mil, "The use of MRI as an outcome measure in clinical trials.," *Nat. Rev. Rheumatol.*, vol. 8, pp. 643–644, 2012.

[51]   X. Li, A. Yu, W. Virayavanich, S. Noworolski, T. Link, and J. Imboden, "Quantitative characterization of bone marrow edema pattern in rheumatoid arthritis using 3 Tesla MRI," *J Magn Reson Imaging*, vol. 35, no. 1, pp. 211–217, 2012.

[52]   F. F. Berendsen, U. A. Van Der Heide, T. R. Langerak, A. N. T. J. Kotte, and J. P. W. Pluim, "Free-form image registration regularized by a statistical shape model: Application to organ segmentation in cervical MR," *Comput. Vis. Image Underst.*, vol. 117, pp. 1119–1127, 2013.

[53]   M. van de Giessen, M. Foumani, G. J. Streekstra, S. D. Strackee, M. Maas, L. J. van Vliet, K. a Grimbergen, and F. M. Vos, "Statistical descriptions of scaphoid and lunate bone shapes.," *J. Biomech.*, vol. 43, no. 8, pp. 1463–9, May 2010.

Appendix A: Literature Review

Early Detection of Rheumatoid Arthritis using extremity MRI:
A literature review of bone segmentation techniques

Edgar A. H. Roex

ulteit Werktuigbouwkunde, Maritieme Techniek
& Technische Materiaalwetenschappen
Delft University of Technology
E.A.H.Roex@student.tudelft.nl
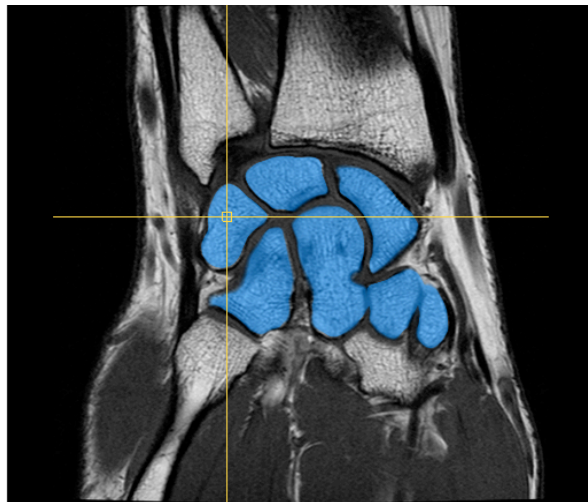
Created:	05-Nov-2013
Last updated:	10-Jan-2014
Supervisors:	Prof. B. P. F. Lelieveldt
	Dr. B. C. Stoel

Abstract—Visual scoring of magnetic resonance images for the early detection of rheumatoid arthritis is prone to human subjectivity and reliability issues. In a bid to develop an objective and quantitative alternative using digital image processing, accurate bone segmentation is required. In this literature review the suitability of existing bone segmentation techniques is evaluated for the purpose of automatically delineating the carpal bones from MRI data. Furthermore, the requirements of the segmentation technique are stipulated, as well as the search strategy used to peruse the literature. Following a systematic search, a range of methods was uncovered. By studying seven key articles in-depth, their strengths and weaknesses are discussed. From the analysis, it is argued that atlas-based segmentation is the preferred method, given it is automatic, incorporates prior knowledge and is implemented directly in 3D.

Contents

Introduction

The aim of this literature study is to reveal and summarise the current state of knowledge with regard to the early detection of rheumatoid arthritis (RA) using magnetic resonance (MR) images of extremities (hands, wrists and feet). Specifically it focusses on segmentation of bone tissue in MR images. Segmentation, in image processing, refers to the process of subdividing an image into its constituents or regions of interest [1]. By examining the literature, one is in a better position to place and direct their research (i.e. formulate a research problem), which ultimately aims to address a gap in knowledge. Of course the literature should be consulted on an ongoing basis, as the project evolves and as new articles are published. This review will establish the context and rationale of the master's thesis project and confirm the choice of research focus. Furthermore, by evaluating the literature, eventual findings can be more easily compared to those of others, and their implications assessed.

Rheumatoid arthritis (RA) is an autoimmune disease typified by chronic inflammation and eventual destruction of the joints. The exact pathogenesis is still unknown. Modern therapeutics can suppress the progress of the disease and even force it into remission, preventing joint deformities and subsequent disability. To maximise the chance of success, this treatment strategy must be initiated very early, thus early detection is critical. Accordingly, the focus of medical imaging has shifted to detecting pre-erosive inflammatory features, which act as prognostic markers for RA. These include synovitis, tenosynovitis and bone marrow edema (BME). Magnetic resonance imaging (MRI), with its superior soft tissue contrast, is a vital tool that captures all relevant structures and changes in arthritic disease [2][3]. Current scoring of MRI findings for RA is done semi-quantitatively, using the RA MRI scoring system (RAMRIS) developed by the Outcomes Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group [4][5]. It visually scores the carpal (wrist) and metacarpophalangeal (MCP) joints for synovitis, bone marrow edema and erosion and requires a trained reader. This approach, though a welcome one, is inherently limited by the human visual system and prone to observer variability and reliability issues. Objective and quantitative assessment is preferable but requires further development in order to ultimately enhance the predictive power of the aforementioned imaging biomarkers.

Devised from MRI studies, the term BME is defined by OMERACT as: "a lesion within the trabecular bone, with ill-defined margins and signal characteristics consistent with increased water content" [4]. In other words, BME describes the replacement of bone marrow fat by fluid (containing H^+ ions), a change readily detected by MRI. Edematous bone is more vascularised and therefore can be visually enhanced with the aid of gadolinium, an intravenous contrast agent, in dynamic contrast enhanced (DCE) MRI [6]. After a dose of intravenous contrast, BME is seen as increased signal (see Figure 2). In patients with early RA, BME is reported to be an independent predictor of radiographic progression [7] and in the largest cohort study to date it is hailed the strongest imaging predictor of future erosions [8]. With eighty-four patients however, the sample size of this study is fairly small. Visually, BME has an indistinct, "feathery" appearance, thus making it challenging to delineate. The lesion is unique to MRI and has been described as "intriguing" by Peterfy [9], all together making it an interesting and worthwhile feature to study. MRI BME is reported to be most frequently scored in the carpal (wrist) bones, namely the lunate, triquetrum and capitate [10] (see Figure 1 for bone anatomy of the hand). This is supported Li et al. [11], who report the carpal bones contain significantly higher total volume of BME, compared to other joint sites of the hand. These reasons suggest the carpalia form a warranted joint site to study for the characterisation of BME. In addition, the available extremity MRI data is acquired in separate stations, so a choice of joint site (either MCPs or carpal bones) is

needed to avoid image stitching. Accordingly, the scope of this graduation project is narrowed; the focus will be on detecting and quantifying MRI BME in the carpal bones using T1-weighted pre- and post-contrast images generated by extremity MRI.
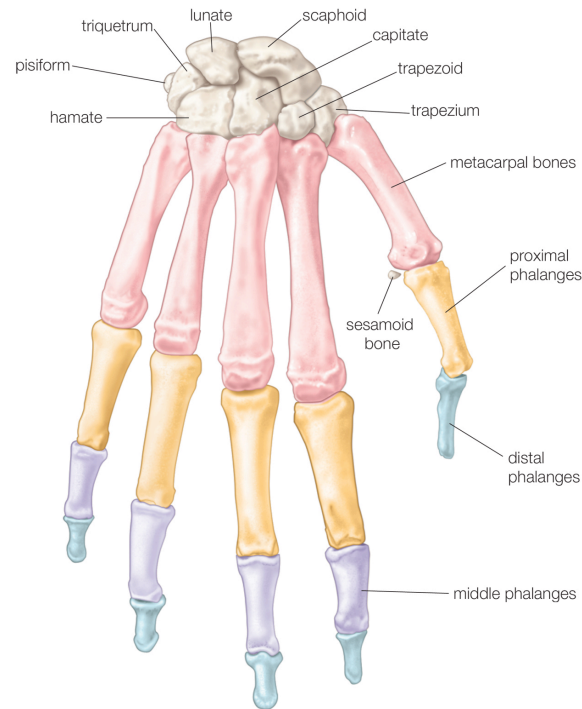
Figure 1. Bones of the hand, showing the carpal bones (wrist bones), metacarpal bones (bones of the hand proper), and phalanges (finger bones) —*Encyclopædia Britannica, Inc.* (2011)

Figure 2. Example of a bone marrow edema lesion (B) on an MR image of the knee. This lesion was given a score of 2 on a scale of 0 to 3. [12]

To recap, current visual scoring methods of extremity MR images for early detection of RA lack objectivity and reliability and there is a call for a quantitative approach. Hence, the goal of this project is to develop and validate image-processing tools that can quantitatively assess pre-erosive inflammatory biomarkers for the early detection of RA. The first step of the proposed quantitative characterisation of MRI BME in the carpal bones is the segmentation of the carpal bones. Following which, the presence of BME is measured using image-processing techniques. These quantitative findings will be compared with a RAMRIS evaluation of the same data. This literature review addresses the first step, namely: the segmentation of bone tissue in MR images. Accurate segmentation is the goal, and will determine the eventual success or failure of subsequent computerised analysis. Being a crucial step in the image-processing pipeline, the purpose of this literature review is to reveal the current state of knowledge with respect to bone segmentation techniques for MR images. This will help form a decision on what segmentation method is most appropriate for segmenting carpals in MRI data or whether a new method needs to be developed. This review is organised into three sections. The first is a discussion of the methods used to survey the literature. Secondly, the results of the literature review are presented and selected works critically analysed. Finally, the implications of these results are discussed and a way forward is suggested.

Methods

This section outlines the methods used to retrieve, organise and analyse the literature deemed relevant to answering the research question. In addition, the tools, key search terms, and inclusion and exclusion criteria used are discussed. The literature search was performed between 4-Nov-2013 and 5-Dec-2013. A number of papers [13]–[15] were provided by my supervisor as a starting point, however these focussed more on the broader scope of the project, rather than on bone segmentation specifically. Serving a similar purpose and containing many useful references, a grant application on the wider project of early detection of RA was also provided [16].

Search terms—Atlas, automatic, bone, bone marrow (o)edema, carpal(s), computed tomography (CT), computer-aided, computer-assisted, image analysis, joint(s), knowledge-based, magnetic resonance imaging (MRI), quantification, radiography, registration, rheumatoid arthritis (RA), segmentation, synovitis, wrist.

Databases—Google Scholar, IEEE Xplore, PubMed, Science Direct, Web of Knowledge, Wiley Online Library.

A number of requirements were set to guide the literature search. Namely, bone segmentation ought to be automatic, or with as little user interaction as possible, seeing as the aim is to make the interpretation of RA MR images objective and computerised. The segmentation technique should be fit for MRI data, both pre- and post-contrast T1-weighted images. Mimicking a human reader, the segmentation process should include prior knowledge of the carpals where possible. In other words, a naïve approach to segmentation is discarded, as it is neither accurate nor robust enough for our needs.

An initial search was performed broadly to avoid missing important articles. Later searches were refined by adding and combining different search terms. Some search engines, such as Thomson Reuters' Web of Knowledge allow search refinement by categories such as radiology and medical imaging, rheumatology and biomedical engineering. By trying several logical and relevant permutations, more literature could be captured. In trying to ascertain the state of the art as well as to avoid duplicates, the most recent publication on the same work was included whenever possible. There were no limits put on publication year however, as image processing is a fairly young field, and there are many important publications from previous decennia that modern techniques are founded on. Leading authors were identified and citations were traced on Google Scholar and in reference lists to find related works. Articles were downloaded and managed using *Mendeley Desktop* reference manager software. This software also supports cite while you write (CWYW) referencing with word processors but also exports to BibTeX for use with LaTeX. Useful added features include cloud-based file storage and the ability to make annotations and highlights within the program itself, enabling a paperless literature review.

After collection, a number of key articles were read to further familiarise oneself with the topic. Following which, a selection process took place. Key criteria were based on the imaging modality used. A preference was given to MRI related works, especially those including DCE-MRI. However CT and conventional radiography (CR) based techniques were also included as they can form a source of inspiration as well. Similarly, a preference was given to fully automated segmentation techniques. The broad searches were targeted to bone segmentation, whereas later search terms pertaining to

(rheumatoid) arthritis were included to see if disease-specific solutions already existed. Papers that did not focus on the segmentation of bone tissue in the broadest sense were not included. The purpose of article selection was, despite a specific application area, to get a wide range of techniques, to draw inspiration from, and critically judge them on their suitability to the research problem. Often it was sufficient to read the abstract, and image-processing sections of articles to judge whether it was unique or novel in terms of the segmentation technique it presented. A table was used to enter the most important publications and organise them based on the segmentation technique, imaging modality, target site, level of automation and whether the work was validated or not (see Table 1). This also provided a classification scheme for incoming articles to help maintain oversight.

Results

In this section the collected literature is presented and critically evaluated. Key trends and commonalities will be identified. Furthermore, the suitability of existing methods to the research problem (automatic segmentation of carpal bones in MR images) will be assessed. Image segmentation is the division of an image into meaningful structures [17]. A myriad of different approaches, methods and techniques to achieve this are presented in the literature, as well as extensive reviews thereof. However, there is no single approach that can solve all segmentation tasks. It is very application dependent and segmentation algorithms need to be tuned to the target data. A review of segmentation techniques would be superfluous; instead the most relevant and applicable techniques found in the literature will be highlighted. For the most relevant segmentation approaches their suitability is critically analysed in light of the requirements - segmentation should be automatic, fit for MRI data and incorporate prior knowledge.

At the time of writing, some ninety publications have been collected, forty of which are related to or contain details on bone segmentation. Selecting only papers that focus on the technical aspects of bone segmentation further reduced this list. Seven papers were chosen for detailed evaluation. They were read in-depth, and in most cases, several times. They are listed in Table 1 (page 13) and are classified on key aspects such as segmentation technique, bone site, and level of automation. In the literature review, the common segmentation techniques encountered can be roughly categorised as follows:

- Threshold-based and region growing
- Active contours and level sets
- Classifiers
- Atlas-based

Important to note is that no segmentation method belonged exclusively to one category, and in all cases, a combination of image processing and segmentation techniques is used. Manual outlining or corrections, though often inevitable, is not included as it is not automated. Following is a discussion of the seven selected papers (with reference to a few others), presented in order of the segmentation categories listed above. For each paper, a description of the segmentation method used is given, followed by an evaluation of its merits and weaknesses in light of the research problem.

Kubassova et al. [18] present their work on the automatic segmentation of the MCP joints in DCE-MRI acquired images. Their approach is primarily threshold-based and is quite creative, hence its inclusion. The authors reduce the 3D segmentation problem to 2D and even 1D as will be explained. To align pre- and post-contrast images, registration is performed akin to Mysling et al. [19] (discussed later). Segmentation of the MCP joints is realised in two stages: first, regions that represent bone interiors are detected, then, the boundaries of the regions are refined using an adaptive segmentation technique. The preliminary stage involves background removal followed by morphological opening to remove outliers. A location classifier tells the algorithm where to "expect" bone based on training data. Note the coordinate system needs to be normalised for this. Should bone for whatever reason fall outside this expected region, then it will fail to be correctly classified. The preliminary segmentation assumes the bone to be elliptical in shape and free of holes. Although blood vessels also satisfy these assumptions, they are distinguished based on their smaller size and location near the joint edges. The adaptive segmentation using the preliminary segmentation mask is used as an initial

bone boundary estimate and aims to improve its conformity to the true bone boundary. As shown in Figure 3, a bounding box is placed around each bone site. The centroid is located and half the diagonal is used as the radius, which is then rotated step-wise. At each step the intensity profile along the radius is retrieved and the point of steepest change is determined to be the optimum boundary pixel.

Kubassova's segmentation algorithm, though creative, is likely to breakdown in a number of areas. Firstly, in case of extreme curvature, using the adaptive segmentation method, the radius could cross more than one boundary pixel. Another limitation in this work is that the segmentation is performed in 2D, on a per-slice basis. This way no information from neighbouring slices is able to contribute to a better bone boundary estimate. Furthermore, the initial bone segmentation mask is derived from the superposition of twenty scans obtained in the DCE-MRI sequence. This seems excessive and depends greatly on the success of interim registrations to control for patient motion and normalise the coordinate system. In the preliminary segmentation, the bone is assumed to contain no holes. This assumption is likely to be violated, especially with RA patient data, where erosions are prevalent. The paper lacks details on the removal of small blood vessels and given they satisfy the same assumptions as bone (elliptical, no holes), accidental exclusion of bone voxels could occur (false negative). Although Kubassova et al. achieve a reasonable segmentation rate (80% mutual overlap) by combining relatively basic image processing techniques, the total image processing chain contains many sub-methods, meaning any small error is carried forward and multiplied. Should all assumptions hold, this algorithm is quite efficient. Yet, as the authors themselves are aware, "the procedure is not perfect".
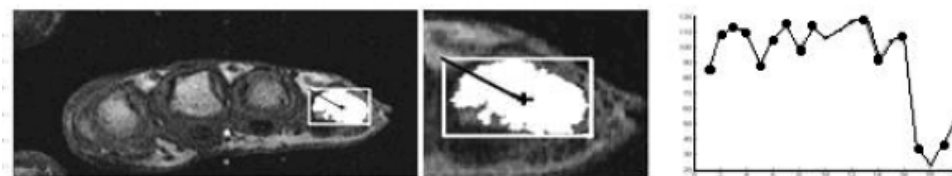
Figure 3. Adapative segmentation of MCP. Boundary pixel refinement using 1D signal intensity thresholding. (Kubassova, 2005) [18]

Duryea et al. [20] present "semi automated three-dimensional segmentation software to quantify carpal bone volume changes on wrist CT scans for arthritis assessment". Even though the segmentation is semi-automatic and CT-based, this paper was included as it closely aligns with the research at hand. User interaction is needed to place a ROI over each carpal in the most central slice. A seed-point is then manually placed on the bone boundary to initialise an edge-tracking algorithm (see Figure 4.a). From the seed point, the carpal's edge is traced using the maximum greyscale gradient. Even at CT resolution, the carpals are not always separable based on intensity gradient alone, and so the user must correct the edge-tracing if it mistakenly jumps to an adjacent carpal. This is seen in Figure 4.c below and highlights the need to bestow prior knowledge upon the algorithm. After the central slice has been outlined, adjacent slices are segmented using active contours; delineations from the previous slice serve as initialisations for the next. The biggest drawback of Duryea's segmentation approach is the lack of automation. In addition, a gradient-based cost function is simple yet does not incorporate other useful information available in the data.
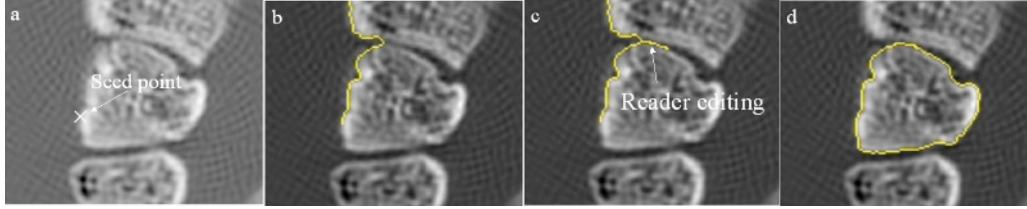
Figure 4. User-guided edge tracing for active contour initialization. (Duryea, 2008) [21]

In contrast to Duryea, Lorigo et al. [22] incorporate texture (in the form of local variance), as well as intensity into the weighting function of the geodesic snakes to segment knee bones in MRI. Exploiting the fact that trabecular bone (situated near joints) has fairly uniform texture, the authors report successful segmentation. Use of the balloon model allows initialisation anywhere within the bone. After convergence, due to windowing effect, the final contour has to be pushed outward by a constant equal to the window size. As in Duryea's work, the 2D segmented contours are stacked to arrive at the 3D segmentation. Intensity-only and variance-only energy functions were also tested, where interestingly, it was found the variance-only model rivalled the two-value approach. Although this technique is shown to be successful on T1 (anatomical) scans of the knee, it remains to be seen how it works on contrast enhanced MR images, where the intensity varies more greatly and a different parameterisation and cost function may be called for. As is the case in this publication, but also true for others, the absence of a ground truth makes it difficult to assess segmentation accuracy. Different radiologists will make different interpretations on where bone boundaries are situated. This inherent variability makes validation difficult to perform. At best we can say: "the segmentation result is supported by that of a radiologist". Comparing the approaches from Duryea and Lorigo and given the complexity of MR images, it is clear gradient information alone is not sufficient nor reliable enough to achieve satisfactory results (automatically). Another difficulty in active contour based approaches is that they segment 2D contours which are then stacked to form a final 3D segmentation. When computing an optimal boundary pixel, this limits the available information to 2D neighbouring pixels. When performed directly in 3D, boundary pixels from adjacent slices can contribute additional information to form a more informed, and thus accurate segmentation. As reported by Duryea et al. partial volume effects and bone surfaces that are oblique to the imaging plane cause disruptions in the image boundary. Resulting from volume averaging, these challenges are important to note, and are an inherent artefact of tomographic MR images.

Where a location classifier is a helper method in Kubassova's work, Folkesson et al. [23] base their entire segmentation method of articular cartilage in knee MRI on a three class k-nearest-neighbour (kNN) classification scheme (see Figure 5). Applied directly in 3D, this fully automatic segmentation technique could be adapted for carpal bone segmentation. Using a range of features, the authors are able to include prior knowledge such as position in the image, and Gaussian smoothed intensity. By approximating the cartilage as a curved disc in 3D, the eigenvalues and eigenvectors of the Hessian are taken as features; these describe the size and direction of the principal curvatures. Although this method is quick and reportedly successful, the feature sets would need to be tailored to extracting carpal bones in place of knee cartilage (a disk shaped assumption does not hold for the carpals). Given the carpals are somewhat rounded objects, a potential shape descriptor for the carpals is eccentricity, which describes how much a shape deviates from a perfect circle (eccentricity of 0). This is explained by Zhang et al. [24] in their work on bone age assessment using radiographs of the carpals in children. Their acceptance bounds seem rather wide (.1<eccentricity<.9), meaning few objects are eliminated. Having said that, these bounds could be enough to reject all objects that deviate too much from being a circle/ellipse. On the classification of carpal bones based on their shape, van de Giessen

et al. [25] report no evidence of distinct shape classes in their study, which generates statistical shape models (SSM) of the lunate and scaphoid. This suggests, shape descriptors alone are insufficient to anatomically differentiate carpal using classifier-based segmentation.
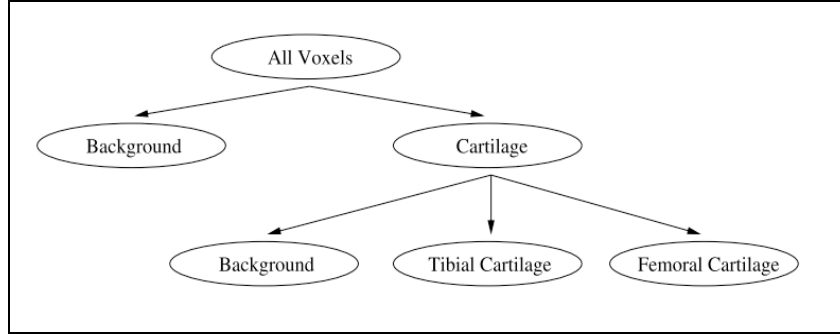
Figure 5. Tree representation of hierarchical three class classification scheme. (Folkesson, 2005) [23]

The most recent and relevant publication is by Mysling et al. (2013) [19]. The authors present an atlas-based segmentation (ABS) technique of the wrist synovium. If this paper targeted BME of the wrist instead, it would precisely address our research problem. ABS is commonly used in medical image processing, and involves a clinical expert manually annotating an image or data set, forming a reference image or atlas. The atlas is registered[1] (deformed) to the patient scan, allowing the label information to be transferred to the target data via spatial inference or voting. In constructing their atlas, Mysling et al. segment the wrist synovium into three regions by subtracting manual delineations of the bones from manual delineations of the overall wrist region of interest (ROI) (see Figure 6). Registration occurs in two stages. First, a global, rigid registration using only translation, rotation and scale is performed. This process is also called an affine transformation. Subsequently, several iterations of local, non-rigid registration based on and B-spline basis functions are performed. The success of each registration is determined by a similarity measure. Normalised mutual information (NMI) was found to be the best performing similarity measure in performing registration to correct for patient motion during contrast administration. Almost every paper on ABS echoes this result, suggesting NMI is useful in optimising multi-modal registration. The hierarchical and multi-stage approach to registration is also commonly seen in the literature. Therefore Mysling et al. do not present a novel segmentation technique per se, but the paper is of interest as it looks at quantifying in-flammatory arthritis, in the same anatomical region, using a similar MR protocol (including DCE-MRI) and thus aligns closely to our research. Only the choice of inflammatory RA biomarker is different. Seeing as the carpals are part of the manual segmentation step, the atlas construction can be simplified and adapted to suit our needs. The fact that the authors achieve segmentation and subsequent quantification is a proof of concept that there is an alternative to the labour-intensive visual scoring of MR images for RA. Although it could be argued that ABS, in the manual annotations required for atlas construction, is also laborious, this process is a one-off. Another thing to note is that (freeform) registration, particularly when performed in 3D, is expensive to compute and therefore time-consuming. ABS is applied to the automatic segmentation of the pelvic bones in Ehrhardt's [26] work. The authors note that strong anatomical variations in soft-tissue prevents accurate registration in most cases, and as such they limit their ABS to bone. Extending this idea to the carpal region, segmentation has to handle not only anatomical variations, but also postural changes, as the wrist is a complex joint.

---

[1] Registration is the process of matching or aligning of one image to another

12
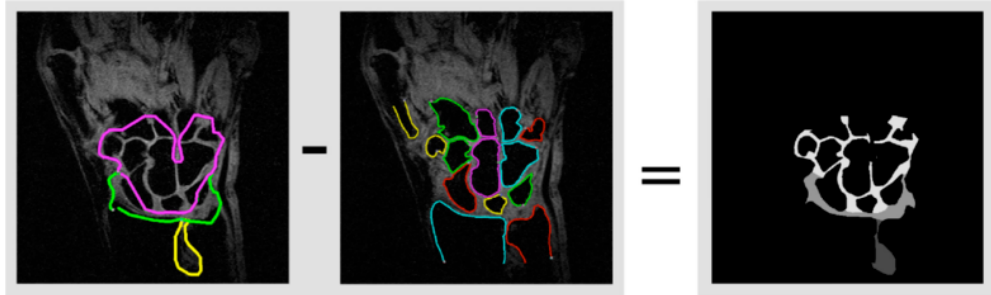
Figure 6. Synovium atlas construction. Left: manual segmentation of synovial regions and wrist bones. Right: resultant synovial atlas. (Mysling, 2013) [19]

Leung et al. [15] describe methods to segment the talus bone of a Lewis rat from serial MR images based on multi-stage registration. This paper is also situated along the RA line of research. The rat is chosen to achieve in-vivo characterisation of bone changes as well as comparison to histological findings, making this publication unique. Just how close a rat's talus is to a human carpal anatomically and physiologically remains insure, but nevertheless, the methods presented can be extrapolated to other bones or multiple bones. Only one randomly selected image after being manually segmented, served as the atlas. Using a single atlas is the simplest approach in ABS. It is understandable given there were only 11 subjects, but it does limit the variability the atlas is able to capture. The multi-stage registration process is similar to that described in Mysling et al. [19]. Leung et al. detail various parameter settings and how long it took to achieve accurate delineation (2-4 hours for a bone). They made use of cross-correlation (CC) as well as NMI as similarity measures. Optimising these similarity measures was performed using simulated annealing and steepest gradient descent (more research is needed to understand how these function exactly). Instead of registering the whole image, the authors chose to register only voxels contained within a ROI around the talus bone. How this is ROI is found is not stated however and may require user interaction. Although the registration is made computationally easier, ROI size and position has a great impact on the registration accuracy. By registering the whole image, more mutual pixel information is available and could lead to better segmentation results. Leung et al. highlight the importance of accurate segmentation, stating that in their case, bone lesions constituted as little as 1% of the volume of the talus bone.

Table 1. Summary of papers related to bone segmentation

First author, year publication [ref]	Modality	Target site	RA Biomarker	T1w + DCE	Segmentation technique	2D/3D	Level of automation	Validation
Kubassova, 2005	MRI	MCP	-	Y	Knowledge-based, adaptive thresholding	2D	Automatic	Clinical judge-ment, manual segmentation
Duryea, 2008 [20]	CT	Carpals (*in-vivo* and phantom)	Erosion (volume)	-	Edge-tracking, active contours	2D→3D	Semi-Aut.	Reproducibility only
Lorigo, 1998 [22]	MRI	Knee	-	N	Level sets	2D→3D	Automatic	-
Folkesson, 2005 [23]	MRI	Knee cartilage	-	N	Classifiers	3D	Automatic	Manual segmen-tation
Ehrhardt, 2001 [26]	CT	Pelvis	-	-	Atlas-based	3D	Automatic	Future work
Leung, 2006 [27]	MRI	Talus (rat)	Erosion	N	Atlas-based	3D	Automatic	Manual segmen-tation, RAMRIS, his-tology
Mysling, 2013 [19]	MRI	Carpals	Synovitis	Y	Atlas-based	2D	Automatic	RAMRIS (2 radiologists)

Table 2. Pros and Cons of segmentation methods surveyed

Segmentation method [ref]	Pros	Cons
Threshold-based and region growing [18]	▪ Simplicity (easy to implement and largely free of parameter selection) ▪ Low computational cost	▪ Lacks automation ▪ Does not incorporate prior knowledge ▪ Many assumptions leads to over-generalisation
Active contours and level sets [20], [22]	▪ Automatic ▪ Reduces problem to 2D segmentation	▪ 2D→3D; lack of 3D spatial information ▪ Data-driven, not so much knowledge-driven
Classifiers [18], [23], [24]	▪ Automatic ▪ Training classifiers incorporates prior knowledge	▪ Requires additional image-processing tools ▪ Feature generation and selection is arduous
Atlas-based [15], [19], [26]	▪ Automatic ▪ 3D model of anatomy; maximise prior knowledge ▪ Registration frameworks useful for patient motion correction in DCE-MRI	▪ Manual annotation needed initially ▪ Computational cost of (accurate) registration ▪ Success limited by postural and anatomical variations

Conclusion

The purpose of this review was to reveal the current state of knowledge with respect to bone segmentation image processing techniques for MR images. The literature survey did not uncover an "off-the-shelf" solution to the challenge of automatically segmenting carpal bones from DCE-MRI data. In this sense, there is a gap in knowledge that needs to be addressed. In fact, there are many ways to achieve segmentation and establishing a ground truth for verification remains non-trivial. We can however, from critically analysing a range of existing methods, come to our own conclusions as to which approach can be best adapted to suit our needs.

Table 2 summarises the pros and cons of the three major segmentation techniques encountered in the literature. Most threshold-based and region growing approaches to segmentation require user intervention or correction, therefore lacking automation, which is one of the requirements. Furthermore, the ability to include prior knowledge is limited. Although implementation is rather straightforward, thresholding requires many assumptions to achieve satisfactory results. The use of level sets can be made automatic, however the cost function definition remains data-driven, as opposed to knowledge-driven. Furthermore, most active contours perform segmentation in 2D, thereby sacrificing valuable 3D information from adjacent slices which can contribute to the vote in voxel classification. For our purposes, ABS is seen as a suitable method, as it is automatic, is knowledge-based, and performed directly in 3D. The atlas is, as its name suggests, a map or spatial record about what we know about a region, in our case the wrist. In medical imaging, it is a probabilistic model of a population of spatial data. In essence, an ABS approach mimics that of the human reader in a RAMRIS-based evaluation, except the atlas is digitalised and uses MRI (modality specific). ABS fits into the framework of this project, as registration is required to correct positional differences between pre- and post-contrast acquisitions. A possible downside is the computational cost of achieving accurate registration. The freeform registration, depending on the allowable degrees of freedom and other settings, will take the longest to compute. However, given this work is not aimed at clinical implementation, the issue of computing time can be put aside. In a number of publications [18][24], ROIs are placed around joint sites of interest, with the motivation of discarding unwanted information and improving the run time of segmentation and subsequent image processing analysis. In the case of ABS however, registration benefits from more information in order to achieve alignment. Hence, for our purposes, ROI will not be used until at least after affine segmentation has been achieved.

Now that the segmentation technique has been chosen. Further research is needed in how to construct the atlas, and how to perform the registration specifically. Then comes the question of how to automatically capture BME and quantify its severity within each carpal bone. A choice of atlas construction is necessary; for example, should a single atlas be used, or multiple atlases, or an average atlas? The goal is to define an atlas that best describes the population and allows for accurate registration. Hence, averaging may lead to loss of anatomical detail and feasibility. The patient-atlas registration process is multipart and complex and its success will determine the accuracy of final segmentation. It is therefore important to carefully select and tune the desired parameters to achieve close registration. A choice of similarity measure is needed too, although it is clear NMI will be useful in registering pre- and post-contrast images. It is intended to use the `elastix` toolbox, which is an intensity based medical image registration software and is freely available [28]. Further consultation of literature and empirical testing will determine which settings lead to accurate and satisfactory bone segmentation.

Abbreviations and acronyms

ABS	Atlas-Based Segmentation
BME	Bone Marrow Edema
CR	Conventional Radiography
CT	Computed Tomography
DCE-MRI	Dynamic Contrast Enhanced MRI
kNN	k-Nearest Neigbours
NMI	Normalised Mutual Information
MCP	Metacarpophalangeal (joints)
MRI	Magnetic Resonance Imaging
OMERACT	Outcome Measures in Rheumatoid Arthritis Clinical Trials
RA	Rheumatoid Arthritis
RAMRIS	Rheumatoid Arthritis Magnetic Resonance Imaging Scoring system
ROI	Region Of Interest

Glossary

Articular	Of or relating to a joint or joints
Arthralgia	Joint pain
Edema	Swelling caused by accumulation of fluid
Geodesic	Relating to or denoting the shortest possible line between two points on a sphere or other curved surface.
Osteitis	Inflammation of bony tissue
Pannus	A membrane of granulation tissue, causes cartilage destruction and bone erosion.
Prognosis	A prediction of the probable course and outcome of a disease
Registration	The process of matching or aligning of one image to another
Segmentation	The process of subdividing an image into its constituent objects or regions of interest

References

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Third. Pearson Education, 2008, p. 689.

[2] J. Narváez, J. Narváez, E. De Lama, and M. De Albert, "MR imaging of early rheumatoid arthritis," *Radiographics*, vol. 30, no. 1, p. 143, 2010.

[3] M. Østergaard and M. Szkudlarek, "Imaging in rheumatoid arthritis – why MRI and ultrasonography can no longer be ignored," *Scand. J. Rheumatol.*, vol. 32, no. 2, pp. 63–73, Jan. 2003.

[4] M. Østergaard, J. Edmonds, F. McQueen, C. Peterfy, M. Lassere, B. Ejbjerg, P. Bird, P. Emery, H. Genant, and P. Conaghan, "An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas.," *Ann. Rheum. Dis.*, vol. 64 Suppl 1, pp. i3–7, Feb. 2005.

[5] P. Conaghan, M. Lassere, M. Østergaard, C. Peterfy, F. McQueen, P. O'Connor, P. Bird, B. Ejbjerg, M. Klarlund, R. Shnier, H. Genant, P. Emery, and J. Edmonds, "OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 4: an international multicenter longitudinal study using the RA-MRI Score.," *J. Rheumatol.*, vol. 30, pp. 1376–1379, 2003.

[6] R. Hodgson, A. Grainger, P. O'Connor, T. Barnes, S. Connolly, and R. Moots, "Dynamic contrast enhanced MRI of bone marrow oedema in rheumatoid arthritis.," *Ann. Rheum. Dis.*, vol. 67, pp. 270–272, 2008.

[7] P. Bøyesen, E. A. Haavardsholm, D. van der Heijde, M. Østergaard, H. B. Hammer, S. Sesseng, and T. K. Kvien, "Prediction of MRI erosive progression: a comparison of modern imaging modalities in early rheumatoid arthritis patients," *Ann. Rheum. Dis.*, vol. 70, no. 1, pp. 176–179, 2011.

[8] M. Hetland, B. Ejbjerg, K. Hørslev-Petersen, S. Jacobsen, A. Vestergaard, A. Jurik, K. Stengaard-Pedersen, P. Junker, T. Lottenburger, I. Hansen, L. Andersen, U. Tarp, H. Skjødt, J. Pedersen, O. Majgaard, A. Svendsen, T. Ellingsen, H. Lindegaard, A. Christensen, J. Vallø, T. Torfing, E. Narvestad, H. Thomsen, M. Ostergaard, and C. study group, "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.*, vol. 68, no. 3, pp. 384–390, 2009.

[9] C. G. Peterfy, "MRI of the wrist in early rheumatoid arthritis.," *Ann. Rheum. Dis.*, vol. 63, no. 5, pp. 473–7, May 2004.

[10] M. Fiona, "Bone marrow edema and osteitis in rheumatoid arthritis: the imaging perspective.," *Arthritis Res. Ther.*, vol. 14, no. 5, 2012.

[11]   X. Li, A. Yu, W. Virayavanich, S. Noworolski, T. Link, and J. Imboden, "Quantitative characterization of bone marrow edema pattern in rheumatoid arthritis using 3 Tesla MRI," *J Magn Reson Imaging*, vol. 35, no. 1, pp. 211–217, 2012.

[12]   D. T. Felson, S. McLaughlin, J. Goggins, M. P. LaValley, M. E. Gale, S. Totterman, W. Li, C. Hill, and D. Gale, "Bone marrow edema and its relation to progression of knee osteoarthritis.," *Ann. Intern. Med.*, vol. 139, pp. 330–336, 2003.

[13]   A. S. Chand, A. McHaffie, A. W. Clarke, Q. Reeves, Y. M. Tan, N. Dalbeth, M. Williams, and F. McQueen, "Quantifying synovitis in rheumatoid arthritis using computer-assisted manual segmentation with 3 tesla MRI scanning," *J. Magn. Reson. Imaging*, vol. 33, no. 5, pp. 1106–1113, 2011.

[14]   A. R. Crowley, J. Dong, A. McHaffie, A. W. Clarke, Q. Reeves, M. Williams, E. Robinson, N. Dalbeth, and F. M. McQueen, "Measuring bone erosion and edema in rheumatoid arthritis: a comparison of manual segmentation and RAMRIS methods," *J Magn Reson Imaging*, vol. 33, no. 2, pp. 364–371, 2011.

[15]   K. K. Leung, M. Holden, N. Saeed, K. J. Brooks, J. B. Buckton, A. A. Williams, S. P. Campbell, K. Changani, D. G. Reid, and Y. Zhao, "Automatic quantification of changes in bone in serial MR images of joints," *Med. Imaging, IEEE Trans.*, vol. 25, no. 12, pp. 1617–1626, 2006.

[16]   B. C. Stoel, "HTSM 2013, Atlas-based Image Analysis of Inflammatory Markers on Extremity MRI for Early Identification of Rheumatoid Arthritis," *STW Grant Appl.*, pp. 1–23, 2013.

[17]   T. Maintz, "Chapter 10. Segmentation," in in *Digital and Medical Image Processing*, Course reader for Introduction to Image Processing, University of Utrecht, 2005.

[18]   O. Kubassova, R. D. Boyle, and M. Pyatnizkiy, "Bone segmentation in metacarpophalangeal MR data," *Springer*, pp. 726–735, 2005.

[19]   P. Mysling, S. Darkner, J. Sporring, E. Dam, and M. Lillholm, "Automatic measurement of wrist synovitis from contrast-enhanced MRI: a registration-centered approach," *Int. Soc. Opt. Photonics*, 2013.

[20]   J. Duryea, M. Magalnick, S. Alli, L. Yao, M. Wilson, and R. Goldbach-Mansky, "Semiautomated three-dimensional segmentation software to quantify carpal bone volume changes on wrist CT scans for arthritis assessment," *Med. Phys.*, vol. 35, no. 6, pp. 2321–2330, 2008.

[21]   J. Duryea, M. Magalnick, S. Alli, L. Yao, M. Wilson, and R. Goldbach-Mansky, "Semiautomated three-dimensional segmentation software to quantify carpal bone volume changes on wrist CT scans for arthritis assessment," *Med. Phys.*, vol. 35, no. 6, pp. 2321–2330, 2008.

[22]   L. M. Lorigo, O. Faugeras, W. E. L. Grimson, R. Keriven, and R. Kikinis, "Segmentation of bone in clinical knee MRI using texture-based geodesic active contours," *Springer*, pp. 1195–1204, 1998.

[23]   J. Folkesson, E. Dam, O. Olsen, P. Pettersen, and C. Christiansen, "Automatic segmentation of the articular cartilage in knee {MRI} using a hierarchical multi-class classification scheme.," *Med. image Comput. Comput. Interv. {MICCAI} ... Int. Conf. Med. Image Comput. {Computer-Assisted} Interv.*, vol. 8, no. Pt 1, pp. 327–334, 2005.

[24]   A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, and H. K. Huang, "Bone age assessment of children using a digital hand atlas.," *Comput. Med. Imaging Graph.*, vol. 31, no. 4–5, pp. 322–31, Jan. 2007.

[25]   M. van de Giessen, M. Foumani, G. Streekstra, S. Strackee, M. Maas, L. van Vliet, K. Grimbergen, and F. Vos, "Statistical descriptions of scaphoid and lunate bone shapes," *J. Biomech.*, vol. 43, no. 8, pp. 1463–1469, 2010.

[26]   J. Ehrhardt, H. Handels, T. Malina, B. Strathmann, W. Plötz, and S. Pöppl, "Atlas-based segmentation of bone structures to support the virtual planning of hip operations," *Int. J. Med. Inform.*, vol. 64, no. 2–3, pp. 439–447, 2001.

[27]   K. K. Leung, M. Holden, N. Saeed, K. J. Brooks, J. B. Buckton, A. A. Williams, S. P. Campbell, K. Changani, D. G. Reid, and Y. Zhao, "Automatic quantification of changes in bone in serial MR images of joints," *Med. Imaging, IEEE Trans.*, vol. 25, no. 12, pp. 1617–1626, 2006.

[28]   S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "elastix: a toolbox for intensity-based medical image registration.," *IEEE Trans. Med. Imaging*, vol. 29, pp. 196–205, 2010.

Appendix B: Supplementary results

Training set

Table B.1: RAMRIS BME scores for the carpal bones of training set subjects.

Subject	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
EAC01	0.25	1	0	1	0	0	0	0	0
EAC02	0.75	0	1	2	2	0	0	0	1
EAC03	1	1	1	2	1	0	1	1	1
EAC04	0.375	1	1	0	0	0	0	1	0
EAC05	0	0	0	0	0	0	0	0	0
EAC06	0.625	2	1	0	0	0	0	1	1
EAC07	0.125	0	0	1	0	0	0	0	0
EAC08	0.625	1	1	1	1	0	0	1	0
EAC09	0.25	0	1	1	0	0	0	0	0
EAC10	0	0	0	0	0	0	0	0	0
EAC11	0	0	0	0	0	0	0	0	0
EAC12	0	0	0	0	0	0	0	0	0
EAC13	0	0	0	0	0	0	0	0	0

Table B.2: Test 1: BM. Percentage BME scores for the carpal bones of training set subjects.

Subject	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
EAC01	11%	9%	13%	21%	12%	7%	9%	10%	9%
EAC02	15%	14%	19%	24%	21%	8%	9%	11%	11%
EAC03	17%	20%	17%	23%	18%	13%	15%	15%	11%
EAC04	20%	22%	42%	26%	6%	18%	18%	19%	11%
EAC05	12%	10%	15%	12%	13%	14%	12%	12%	7%
EAC06	15%	30%	15%	13%	11%	8%	12%	17%	16%
EAC07	11%	12%	10%	20%	8%	8%	10%	12%	12%
EAC08	13%	19%	11%	16%	9%	11%	14%	13%	13%
EAC09	11%	16%	13%	14%	14%	7%	8%	9%	7%
EAC10	11%	16%	13%	14%	14%	7%	8%	9%	7%
EAC11	25%	28%	30%	30%	20%	31%	26%	18%	15%
EAC12	14%	18%	17%	15%	19%	6%	12%	12%	9%
EAC13	6%	6%	8%	8%	8%	6%	7%	4%	5%

Table B.3: Test 2: BM + PP. Percentage BME scores for the carpal bones of training set subjects.

Subject	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
EAC01	3%	2%	8%	5%	0%	0%	1%	3%	2%
EAC02	3%	2%	9%	4%	3%	0%	1%	2%	2%
EAC03	2%	1%	4%	8%	1%	0%	1%	2%	1%
EAC04	9%	13%	25%	14%	0%	1%	7%	8%	2%
EAC05	3%	0%	2%	4%	7%	0%	3%	7%	0%
EAC06	5%	10%	4%	3%	2%	0%	1%	10%	8%
EAC07	3%	0%	0%	12%	0%	1%	2%	5%	0%
EAC08	2%	1%	1%	7%	2%	2%	4%	2%	1%
EAC09	2%	6%	4%	4%	4%	0%	0%	2%	1%
EAC10	4%	8%	5%	6%	0%	7%	4%	4%	1%
EAC11	7%	6%	6%	15%	4%	16%	2%	6%	1%
EAC12	4%	5%	11%	6%	0%	0%	2%	5%	0%
EAC13	0%	0%	1%	2%	0%	0%	0%	0%	0%

Table B.4: Test 3: ABS. Percentage BME scores for the carpal bones of training set subjects.

Subject	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
EAC01	18%	20%	15%	20%	15%	20%	15%	20%	15%
EAC02	15%	19%	11%	19%	11%	19%	11%	19%	11%
EAC03	21%	27%	15%	27%	15%	27%	15%	27%	15%
EAC04	24%	28%	20%	28%	20%	28%	20%	28%	20%
EAC05	11%	11%	11%	11%	11%	11%	11%	11%	11%
EAC06	25%	32%	18%	32%	18%	32%	18%	32%	18%
EAC07	18%	19%	17%	19%	17%	19%	17%	19%	17%
EAC08	16%	19%	13%	19%	13%	19%	13%	19%	13%
EAC09	14%	18%	9%	18%	9%	18%	9%	18%	9%
EAC10	18%	21%	14%	21%	14%	21%	14%	21%	14%
EAC11	24%	31%	17%	31%	17%	31%	17%	31%	17%
EAC12	14%	16%	11%	16%	11%	16%	11%	16%	11%
EAC13	5%	6%	4%	6%	4%	6%	4%	6%	4%

Table B.5: Test 4: ABS + PP. Percentage BME scores for the carpal bones of training set subjects.

Subject	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
EAC01	4%	3%	4%	3%	4%	3%	4%	3%	4%
EAC02	3%	4%	2%	4%	2%	4%	2%	4%	2%
EAC03	5%	9%	1%	9%	1%	9%	1%	9%	1%
EAC04	12%	16%	7%	16%	7%	16%	7%	16%	7%
EAC05	2%	2%	1%	2%	1%	2%	1%	2%	1%
EAC06	10%	16%	5%	16%	5%	16%	5%	16%	5%
EAC07	4%	4%	4%	4%	4%	4%	4%	4%	4%
EAC08	1%	1%	2%	1%	2%	1%	2%	1%	2%
EAC09	3%	6%	0%	6%	0%	6%	0%	6%	0%
EAC10	5%	6%	4%	6%	4%	6%	4%	6%	4%
EAC11	7%	10%	3%	10%	3%	10%	3%	10%	3%
EAC12	4%	8%	1%	8%	1%	8%	1%	8%	1%
EAC13	0%	0%	0%	0%	0%	0%	0%	0%	0%

Test set

Table B.6: RAMRIS, mean and per carpal bone, for test set subjects.

Subject	Cohort	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
1	ATL	0	0	0	0	0	0	0	0	0
2	ATL	0.125	0	1	0	0	0	0	0	0
3	ATL	0.125	0	0	0	0	0	0	1	0
4	ATL	0	0	0	0	0	0	0	0	0
5	ATL	0	0	0	0	0	0	0	0	0
6	ATL	0.25	0	0	0	0	0	0	2	0
7	CSA	0	0	0	0	0	0	0	0	0
8	CSA	0	0	0	0	0	0	0	0	0
9	CSA	0	0	0	0	0	0	0	0	0
10	CSA	0	0	0	0	0	0	0	0	0
11	CSA	0.125	0	1	0	0	0	0	0	0
12	CSA	0	0	0	0	0	0	0	0	0
13	CSA	0	0	0	0	0	0	0	0	0
14	EAC	0	0	0	0	0	0	0	0	0
15	EAC	0	0	0	0	0	0	0	0	0
16	EAC	1.125	1	3	3	2	0	0	0	0
17	EAC	0.125	0	1	0	0	0	0	0	0
18	EAC	0	0	0	0	0	0	0	0	0
19	EAC	0.250	0	1	1	0	0	0	0	0
20	EAC	0.125	0	0	1	0	0	0	0	0
21	EAC	2	3	2	1	2	1	3	3	1
22	EAC	0	0	0	0	0	0	0	0	0
23	EAC	0	0	0	0	0	0	0	0	0
24	EAC	0	0	0	0	0	0	0	0	0
25	EAC	0	0	0	0	0	0	0	0	0
26	EAC	0.125	0	0	0	0	1	0	0	0
27	EAC	1.375	1	1	2	1	0	2	2	2
28	EAC	0	0	0	0	0	0	0	0	0
29	EAC	0	0	0	0	0	0	0	0	0

Table B.7: Quantitative scores, mean and per carpal bone for test set subjects.

Subject	Cohort	mean	1 Sca	2 Lun	3 Tri	4 Pis	5 Tpm	6 Tpd	7 Cap	8 Ham
1	ATL	4%	14%	2%	4%	2%	10%	0%	1%	1%
2	ATL	4%	2%	10%	6%	8%	3%	1%	2%	1%
3	ATL	3%	5%	6%	4%	3%	0%	4%	5%	1%
4	ATL	3%	6%	1%	0%	3%	8%	5%	2%	1%
5	ATL	2%	4%	4%	2%	0%	2%	0%	2%	3%
6	ATL	4%	1%	0%	5%	4%	0%	0%	18%	0%
7	CSA	3%	5%	6%	2%	8%	6%	0%	0%	0%
8	CSA	3%	3%	3%	0%	4%	8%	3%	3%	1%
9	CSA	2%	2%	1%	0%	5%	3%	0%	3%	1%
10	CSA	3%	9%	6%	0%	7%	0%	0%	2%	3%
11	CSA	4%	5%	6%	5%	1%	5%	5%	3%	2%
12	CSA	2%	2%	2%	4%	1%	0%	0%	1%	3%
13	CSA	7%	7%	16%	7%	1%	9%	4%	4%	7%
14	EAC	6%	3%	13%	0%	13%	5%	4%	5%	1%
15	EAC	2%	1%	2%	5%	2%	1%	1%	2%	1%
16	EAC	7%	6%	18%	9%	3%	4%	8%	4%	1%
17	EAC	7%	10%	12%	6%	1%	7%	9%	1%	8%
18	EAC	4%	0%	6%	8%	8%	4%	3%	3%	3%
19	EAC	7%	11%	18%	5%	2%	13%	1%	2%	3%
20	EAC	6%	6%	9%	8%	8%	16%	0%	1%	1%
21	EAC	13%	11%	22%	26%	1%	13%	16%	10%	3%
22	EAC	5%	5%	9%	4%	10%	9%	2%	2%	1%
23	EAC	3%	2%	12%	0%	5%	3%	0%	1%	1%
24	EAC	4%	10%	4%	1%	8%	5%	0%	2%	3%
25	EAC	3%	4%	6%	4%	0%	4%	2%	0%	1%
26	EAC	6%	10%	7%	9%	6%	9%	0%	4%	1%
27	EAC	9%	15%	6%	12%	3%	13%	8%	6%	9%
28	EAC	5%	4%	7%	10%	3%	2%	6%	6%	2%
29	EAC	2%	2%	6%	0%	1%	2%	2%	1%	2%