Ioanna Chronaiou

Radiomics in psoriatic arthritis and breast cancer: Assessing disease burden and predicting survival through MR image analysis

Doctoral thesis

■ NTNU
Norwegian University of
Science and Technology

Ioanna Chronaiou

Radiomics in psoriatic arthritis and breast cancer:
Assessing disease burden and predicting survival through MR image analysis

Thesis for the Degree of Philosophiae Doctor

Trondheim, March 2022

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of
Science and Technology

Sammendrag

Radiomics i psoriasisartritt og brystkreft: Vurdering av sykdom og prediksjon av overlevelse gjennom MR-bildeanalyser

Psoriasisartritt er en kronisk inflammatorisk leddsykdom som utvikler seg hos pasienter med hudpsoriasis. Sykdommen innebærer betennelse i ett eller flere ledd og er heterogen med ulike kliniske mønstre. Behandling rettes mot betennelsene og avhenger av sykdommens aggressivitet. Tidlig diagnose og målrettet behandling er viktig for å forhindre progressiv leddskade, deformormasjoner og funksjonshemming som oppstår ved vedvarende betennelse. Brystkreft er den vanligste kreftformen hos kvinner i Norge. Mens kvinner diagnostisert med brystkreft i et tidlig stadium har høy overlevelsesrate, innebærer diagnosen avansert brystkreft en dårligere prognose. Tidlig prognostisk informasjon kan ha betydning for valg av behandling og veilede oppfølging etter behandling.

Magnetisk resonans (MR) avbildning kan bista i diagnose av psoriasisartritt, og brukes t rutinemessig for å vurdere behandlingsrespons i brystkreft. Bildene vurderes kvalitativt ved visuell inspeksjon. Radiologisk tolkning av bilder krever lang trening, er tidkrevende og kan vise høy inter-rater-variasjon. Radiomics, uthenting av kvantitative egenskaper fra medisinske bilder, kan forenkle tolkningav MR-bildene. Det er viktig å finne kvantitative mål fra MR som er pålitelige, sensitive og spesifikke for diagnostiske, prediktive og prognostiske formål.

Hensikten med dette prosjektet er å bidra til utvikling av kvantitative analysemetoder for MR-bilder og oppnå objektive og kvantitative MR-bildebaserte mål for diagnostiske og prognostiske formål. De spesifikke målene med dette prosjektet var å implementere et rammeverk for behandling og analyse av longitudinelle data innhentet med forskjellige skannere og protokoller, etablere kvantitative MR-bildebaserte mål for subtile beinmargsødem i ryggrad og iliosakralledd hos pasienter med psoriasisartritt, og vurdere prognostisk verdi av tekstur-egenskaper ekstrahert fra dynamisk kontrastforsterkede MR-bilder av lokalavansert brystkreft.

Den første artikkelen evaluerte terskling for kvantifisering av benmargsødem i ryggrad og iliosakralledd hos pasienter med psoriasisartritt, og sammenlignet de kvantitative målene fra terskling med et semi-kvantitativt scoringssystem etablert av spondyloarthritis research consortium of Canada (SPARCC). Kvantitative mål fra terskling viste en signifikant positiv

sammenheng med SPARCC for både ryggrad og iliosakralledd, med best resultat for ryggrad. Den andre artikkelen vurderte klassifisering og kvantifisering av benmargsødem hos pasienter med psoriasisartritt ved bruk av intensitet, gradient og teksturegenskaper fra MR-bilder av ryggrad som prediktorer. De beregnede størrelsene viste en signifikant korrelasjon med SPARCC. Klassifiseringen var mer sensitiv men mindre spesifikk enn terskling. Den tredje artikkelen vurderte prognostisk verdi av tekstur-egenskaper ekstrahert fra dynamisk kontrastforsterkede MR-bilder av brystkreft tatt opp før start av behandling. Det ble funnet en klar sammenheng mellom tekstur-egenskaper og overlevelse. I tillegg ga tekstur-egenskapene en merverdi til de kliniske prognostiske faktorene i prediksjon av overlevelse.

For å konkludere; Prosjektet viser at radiomics av MR bilder har potensiale til å bidra i diagnostisering og vurdering av prognose for pasienter med psoriasisartritt og brystkreft. Resultater fra studiene som inngår i dette prosjektet bør valideres i større kohorter.

Kandidat: Ioanna Chronaiou
Institutt: Institutt for Sirkulasjon og Bildediagnostikk
Veiledere: Beathe Sitter, Tone Frost Bathen og Else Marie Huuse-Røneid
Finansering: Høgskolen i Sør-Trøndelag (HiST)/Norges teknisk-naturvitenskapelige universitet (NTNU)

Πᾶσά ἐπιστήμη χωριζομένη τῆς ἀρετῆς πανουργία, οὐ σοφία φαίνεται.

(Knowledge, when separated from virtue, is seen to be cunning and not wisdom.)

Plato, 427-347 BC, Philosopher

Acknowledgements

for taking the time to proofread this thesis. Finally, thanks to Sotiris, who made me happy and kept me sane through the toughest part of this journey.

Summary

Radiomics in psoriatic arthritis and breast cancer: Assessing disease burden and predicting survival through MR image analysis

Psoriatic arthritis is a chronic inflammatory joint disease that develops in patients with skin psoriasis and manifests by inflammation in one or multiple joints, and highly heterogeneous distinct clinical patterns. Treatment strategies target inflammation and depend on disease aggressiveness. Early diagnosis and targeted treatment are important in order to prevent progressive joint damage, deformity and disabilities that occur because of persistent inflammation. Breast cancer is the most frequent type of cancer among women in Norway. While women with breast cancer diagnosed at an early stage have high survival rates, advanced breast cancers exhibit poorer prognosis. Early prognostic information can affect the choice of treatment and guide post-treatment follow-up.

Magnetic resonance (MR) imaging can assist in the diagnosis of psoriatic arthritis and has been routinely used in assessing treatment response in breast cancer. Clinicians rely mainly on qualitative MR findings, based on visual inspection of the images. Radiological interpretation of images requires long training, is time-consuming and can exhibit high inter-reader variance. Radiomics, the mining of quantitative features from medical images, can assist MR image interpretation. It is important to find quantitative MR image-derived measures that are reliable, sensitive and specific for diagnostic, predictive and prognostic purposes.

This project aims to assist in the development of quantitative analysis methods of MR images and obtain objective quantitative MR image-based measures for diagnostic and prognostic purposes. The specific aims of this project included implementing a framework for processing and analysis of data acquired longitudinally and with different scanners and protocols, establishing quantitative MR image-based measures of subtle bone marrow oedema in the spine and the sacroiliac joints (SI) of patients with psoriatic arthritis, and assessing the prognostic value of textural features extracted from dynamic contrast enhanced MR images of locally advanced breast cancer patients.

The first paper evaluated thresholding for the quantification of bone marrow oedema in the spine and the SI joints of patients with psoriatic arthritis and compared the quantitative measures provided by thresholding to a semi-quantitative scoring system established by the

spondyloarthritis research consortium of Canada (SPARCC). Quantitative measures by thresholding showed a significant positive correlation with the SPARCC scores for both the spine and SI joints, performing better in the spine. The second paper evaluated the effect of high intensity interval training (HIIT) in psoriatic arthritis patients by MR radiological assessment of the spinal bone marrow oedema at baseline and after intervention. In addition, the second paper explored the potential of MR image intensity, gradient and textural features to detect bone marrow oedema changes. Bone marrow oedema in the spine was not changed after HIIT and the features were not associated with HIIT in psoriatic arthritis patients. The third paper assessed the prognostic value of textural features extracted from pre-treatment dynamic contrast-enhanced MR images of breast cancer patients. A clear association between textural features and survival outcome was found. In addition, the textural features showed an added value to the clinical prognostic factors in predicting survival outcomes.

In conclusion, this project highlights the use of radiomics, by establishing diagnostic and prognostic quantitative MR image-based measures in both psoriatic arthritis and breast cancer. Results from the studies included in this project should be validated in larger cohorts.

Symbols and abbreviations

^{1}**H**	hydrogen-1
a	optimal percentage for threshold calculation
ACR	American college of radiology
ASDAS	Ankylosing Spondylitis Disease Activity Score
b	random effects coefficients vector (in linear mixed-effects models)
$\boldsymbol{B_0}$	external magnetic field
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
CASPAR	ClASsification criteria for Psoriatic ARthritis
CK	cytokeratin
CRP	C-reactive protein
d	distance
DAS	disease activity score
DCE	dynamic contrast-enhanced
DCIS	ductal carcinoma in situ
E,f	residuals (in partial least squares discriminant analysis)
EGFR	epidermal growth factor receptor
ER	estrogen receptor
$\boldsymbol{f_1, f_2, ..., f_{16}}$	GLCM textural features
G	magnetic field gradient
$\boldsymbol{g_1, g_2, ..., g_{10}}$	gradient features
$\boldsymbol{G_{FE}}$	frequency-encoding gradient
$\boldsymbol{G_{PE}}$	phase-encoding gradient
$\boldsymbol{G_{SS}}$	slice-selective gradient
$\boldsymbol{G_x, G_y}$	2-dimensional directional image gradients for x-axis and y-axis, respectively
$\boldsymbol{G_z}$	magnetic field gradient in the z-direction
GLCM	grey level co-occurrence matrix
HER2	human epidermal growth factor receptor 2
HIIT	high intensity interval training
I	image
$\boldsymbol{i_1, i_2, ..., i_7}$	intensity features

IDC	invasive ductal carcinoma
ILC	invasive lobular carcinoma
_L_	matrix of components (latent variables) (in partial least squares discriminant analysis)
LABC	locally advanced breast cancer
LCIS	lobular carcinoma in situ
LMM	linear mixed-effects model
LOPO	leave one patient out
_M_	net magnetization vector
M_0	net magnetization in equilibrium
M_{xy}	net transverse magnetization
M_z	net longitudinal magnetization
MR	magnetic resonance
N3	non-parametric non-uniform intensity normalization
NAC	neoadjuvant chemotherapy
NST	no special type
_P_	loadings matrix (in partial least squares discriminant analysis)
PCA	principal component analysis
PgR	progesterone receptor
PLS-DA	partial least squares discriminant analysis
RelHyper	relative hyper-intensity
RF	radiofrequency
ROI	region of interest
_S_	signal intensity
_SE_	signal enhancement
SI	sacroiliac
SPARCC	spondyloarthritis research consortium of Canada
STIR	short-tau inversion recovery
_t_	time
_T_	threshold
$T_{1,fat}$	intrinsic longitudinal relaxation time of fat
T_1	longitudinal relaxation constant
T_2^*	total transverse relaxation constant

T_2	transverse relaxation constant
TE	echo time
TI	inversion time
TR	repetition time
X	fixed effects matrix (in linear mixed-effects models)
	independent variable matrix (in partial least squares discriminant analysis)
Y	dependent variable matrix (in partial least squares discriminant analysis)
y	dependent variable vector (in linear mixed-effects models)
Z	random effects matrix (in linear mixed-effects models)
β	fixed effects coefficients vector (in linear mixed-effects models)
γ	gyromagnetic ratio
ε	observation error vector (in linear mixed-effects models)
θ	flip angle
μ	mean
σ	standard deviation
φ	direction of GLCM
ω	resonant frequency
ω_0	Larmor frequency

List of papers

Paper I

Quantifying bone marrow inflammatory oedema in the spine and sacroiliac joints with thresholding

Ioanna Chronaiou, Ruth S. Thomsen, Else M. Huuse, Leslie R. Euceda, Susanne J. Pedersen, Mari Hoff, Beathe Sitter

Paper II

Evaluating the impact of high intensity interval training on axial psoriatic arthritis based on MR images

Ioanna Chronaiou, Guro Fanneløb Giskeødegård, Ales Neubert, Tamara V. Hoffmann-Skjøstad, Ruth S. Thomsen, Mari Hoff, Tone F. Bathen, Beathe Sitter

Paper III

Feasibility of contrast-enhanced MRI derived textural features to predict overall survival in locally advanced breast cancer

Ioanna Chronaiou, Guro Fanneløb Giskeødegård, Pål Erik Goa, Jose Teruel, Roja Hedayati, Steinar Lundgren, Else Marie Huuse, Martin Pickles, Peter Gibbs, Beathe Sitter, Tone Frost Bathen

Table of Contents

1. Introduction

1.1 Psoriatic arthritis

Psoriatic arthritis is a chronic inflammatory joint disease that is associated with skin psoriasis (1). The prevalence of psoriatic arthritis in Norway has been estimated to 0.2%-0.7% (2, 3) and is the same for both sexes (4). Psoriatic arthritis is an autoimmune disease that manifests in five distinct clinical patterns - distal predominant pattern, oligo-articular asymmetrical, poly-articular rheumatoid arthritis-like, spondylitis, and arthritis mutilans (1). The most recent set of criteria for the classification of psoriatic arthritis patients is the Classification criteria for Psoriatic ARthritis (CASPAR) (5). The CASPAR comprises the presence of current or past psoriasis, the presence of current or past dactylitis, psoriatic nail dystrophy, a negative test result for rheumatoid factor, and radiological evidence of new bone formation. These criteria seem to have high sensitivity and specificity, and have been used in clinical practice as well as in the assessment of new treatments (5).

Patients with psoriatic arthritis experience a wide range of symptoms of varying severity, including stiffness, pain, swelling and tenderness of the joints, ligaments, and tendons. Patients may experience a different range and severity of symptoms, with inflammation in one or multiple joints. As psoriatic arthritis progresses over time, the disease can change to a poly-articular pattern, where more than five joints are affected, and the prognosis is poor (4). Sclerosis, erosions, fat deposition and bony bridges/ankyloses reflect previous inflammation and are considered chronic inflammatory signs. The course of the disease is unpredictable, with most patients having erosions and deformations within the first years of diagnosis (6). Enthesitis and synovial thickness were found predictive of future structural damage (7).

Treatment strategies for psoriatic arthritis vary and depend on the symptoms. Patients with mild psoriatic arthritis are usually treated with physical therapy, non-steroidal anti-inflammatory drugs, and local injections of corticosteroids. Patients with aggressive psoriatic arthritis receive additional treatment with synthetic (e.g. methotrexate, leflunomide, salazopyrin) and/or biologic disease-modifying anti-rheumatic drugs (e.g. tumour necrosis factor inhibitors, interleukin-17 inhibitors) that target inflammation (4). Early diagnosis and appropriate treatment are important in order to prevent progressive joint damage, deformity

and disabilities that occur when persistent inflammation is present. Psoriatic arthritis is also associated with decreased quality of life and increased mortality risk, with cardiovascular disease high on the list of psoriatic arthritis death causes (8).

Even though physical exercise is recommended for patients with psoriatic arthritis, the beneficial effects of this intervention have not been widely investigated (9). A previous study (9, 10) investigated the effect of high intensity interval training (HIIT) on psoriatic arthritis patients. HIIT is a method of physical activity consisting of repeated intervals of high aerobic activity at 80-95% of maximum heart rate separated by periods of lower intensity activity (11). It was shown that HIIT was beneficial for psoriatic arthritis patients in the prevention of cardiovascular disease by improving cardiorespiratory fitness and reducing abdominal fat (9), and in decreasing fatigue (10). However, no clear effects on disease activity markers were found (10).

1.2 Axial psoriatic arthritis

Psoriatic arthritis in the spine and the axial skeleton is called axial psoriatic arthritis and is defined as spondylitis. In patients with psoriatic arthritis, the prevalence of axial psoriatic arthritis varies from 25% to 75% depending on the criteria used (12, 13). Axial psoriatic arthritis can be clinically asymptomatic, with only subclinical inflammation present. Axial inflammation can be evaluated using the composite scores, the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (14) and the Ankylosing Spondylitis Disease Activity Score (ASDAS) (15). The BASDAI consists of five and the ASDAS of four questions, with patient reported answers. In addition, the ASDAS includes the level of the C-reactive protein (CRP) or the erythrocyte sedimentation rate. Both tools are useful in clinical practice as well as in research, when evaluating spondylitis disease activity (14, 15).

1.3 Breast cancer

Women's breast consists of glandular (lobes, lobules and ducts), fatty, connective tissue and a system of lymph nodes (Figure 1). Breast cancer is the most frequent cancer type in women in Norway, followed by lung and colon cancer (16).

There are several types of breast cancer based on the tissue it affects. Ductal carcinoma in situ (DCIS) is neoplastic proliferating epithelial cells in the ducts, with an intact layer of basement

membrane and myoepithelial cells (17). Left untreated, DCIS progresses to invasive breast cancer in 40% of the cases (17). Invasive ductal carcinoma (IDC, also known as invasive carcinoma of no special type (NST)) is the most common type of invasive breast cancer comprising 80% of breast cancer cases, beginning in the ducts and invading surrounding tissue (18). Lobular carcinoma in situ (LCIS) is a non-infiltrating proliferation of cells in the lobules, the milk-producing glands at the end of the ducts. Invasive lobular carcinoma (ILC) is a type of cancer that affects the lobules and their surrounding breast tissue. ILC is the second most common type of cancer with an estimated incidence of 10% of all breast cancers (18). Other, less common types of breast cancer include Paget's disease of the nipple and inflammatory breast cancer.
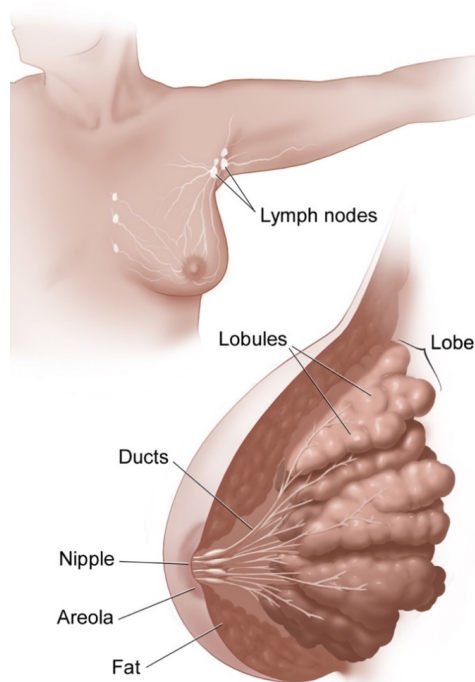
Figure 1. Drawing illustrating anatomy of breast and adjacent lymph nodes. Image from Don Bliss, National Cancer Institute.

Another way of subtyping breast cancer is based on gene expression. There are five intrinsic or molecular subtypes of breast cancer; luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-enriched, basal-like and normal-like breast cancer (19). Table 1 summarizes the immunohistochemical status for the five intrinsic subtypes for estrogen receptor (ER), progesterone receptor (PgR), HER2, protein Ki67 and basal markers. Ki67 is a marker of cell proliferation, and basal markers comprise of cytokeratin (CK) 5, CK14 and epidermal growth factor receptor (EGFR).

Table 1 Immunohistochemical status of intrinsic breast cancer subtypes

Intrinsic subtype	Immunohistochemical status
Luminal A	ER+ and/or PgR+, HER2-, Ki67-
Luminal B	ER+ and/or PgR+, Ki67+
HER2-enriched	ER- and PgR-, HER2+
Basal-like	ER- and PgR-, HER2-, basal marker +
Normal-like	ER+ and/or PgR+, HER2-, Ki67-
ER: estrogen receptor, PgR: progesterone receptor, HER2: human epidermal growth factor receptor 2	

1.3.1 Staging, biomarkers and prognosis

Breast cancer staging refers to identifying the burden of disease, which provides helpful information regarding prognosis and guides the choice of initial treatment (20). The most common staging system for breast cancer is anatomically based and uses a combination of primary tumour size (T), regional lymph node (N) spread and presence or absence of distant metastasis (M). A simplified version of the TNM classification system for breast cancer is presented in Table 2. Using the TNM status, breast tumours are grouped into five stages of increasing aggressiveness (0 to IV) and poorer prognosis.

The Nottingham grading system is used to divide breast cancers into three histological grades; low, intermediate and high grade, depending on how differentiated the cells are regarding their appearance and growth pattern compared to normal cells (21). In addition to the anatomically based TNM staging, histological type (IDC, ILC or other) and other biomarkers (ER, PgR, HER2) are important in staging and stratifying breast cancer patients to treatment (22).

Tumours positive for hormonal receptors are more likely to respond to hormonal therapy and thus have more available treatment options.

Table 2. Simplified version of the TNM classification system for staging breast cancer tumours (23).

Primary tumour size (T)	
TX	Primary tumour cannot be evaluated
T0	No evidence of primary tumour
Tis	Carcinoma in situ (DCIS, LCIS, Paget's disease of the nipple)
T1	Tumour is ≤ 20mm in greatest dimension
T2	Tumour is > 20mm, but ≤50mm in greatest dimension
T3	Tumour is > 50mm in greatest dimension
T4	Tumour has grown into the chest wall and/or the skin, or tumour is inflammatory breast cancer
Regional lymph nodes (N)	
NX	Regional lymph nodes cannot be evaluated
N0	No regional lymph node metastasis
N1	Metastasis to movable ipsilateral level I, II axillary lymph node(s)
N2	Metastasis to fixed ipsilateral level I, II axillary lymph node(s)
N3	Metastasis in ipsilateral infraclavicular (level III axillary) lymph node(s)
Distant metastasis (M)	
M0	No clinical or radiographic evidence of distant metastasis
M1	Distant detectable metastasis
DCIS: ductal carcinoma in situ, LCIS: lobular carcinoma in situ	

1.3.2 Treatment strategies

Treatment strategies in breast cancer patients depend on the disease burden, the tumour location and biology and the age and health status of the patient (24). Stage I and II, except for IIB (T3N0M0), are considered as early stage breast cancers. These tumours are treated with surgery and then with post-operative radiation therapy, chemotherapy, hormone therapy or a combination, depending on the biology of the tumour. Tumours of higher stages are considered advanced breast cancer. Advanced breast cancers in the absence of distant metastasis are

categorized as locally advanced breast cancer (LABC) (25). Guidelines for LABC suggest primary systemic therapy in order to downsize the primary tumour, eradicate micro-metastases and enable breast-conserving therapy, followed by surgery (26).

1.4 Magnetic Resonance Imaging

Magnetic resonance (MR) imaging is an imaging technique using non-ionizing radiation that can provide anatomical and functional images of the body both in health and disease.

1.4.1 Principles of Nuclear Magnetic Resonance

Nuclei with non-zero spin possess a magnetic moment when placed in a magnetic field (27). In an external magnetic field ($\boldsymbol{B_0}$) magnetic moments of a single nuclei precesses around an at a specific frequency, called Larmor frequency (ω_0), that is proportional to the strength of $\boldsymbol{B_0}$ (28).

$$\omega_0 = \gamma B_0 \quad \text{(Equation 1)}$$

where γ is the gyromagnetic ratio.

The hydrogen-1 (^{1}H) nucleus is the most common used nucleus in current MR applications. ^{1}H is a spin $\frac{1}{2}$ nucleus and has two possible spin states with different energy. In an external magnetic field ($\boldsymbol{B_0}$), slightly more ^{1}H nuclei are found at the lower energy state compared to the higher energy state, leading to a net magnetization vector, $\boldsymbol{M}$, that is aligned to $\boldsymbol{B_0}$ (29). Applying a radiofrequency (RF) pulse at the Larmor frequency of the nuclei under inspection will excite the nuclei to a higher energy state, a process called magnetic resonance, and tip $\boldsymbol{M}$ out of alignment with $\boldsymbol{B_0}$. The excited nuclei release the absorbed energy to the environment and $\boldsymbol{M}$ is re-established along $\boldsymbol{B_0}$ through a process called longitudinal relaxation, characterized by a time constant T_1(30).

$$M_z(t) = M_0(1 - e^{-t/T_1}) + M_z(0^+)e^{-t/T_1} \quad \text{(Equation 2)}$$

where M_z is the longitudinal component of $\boldsymbol{M}$, M_0 is the maximum value of M_z and t is the time. T_1 is when 63% of the longitudinal magnetization is re-established.

The process by which the transverse components of M decay to 0 and spins lose their phase coherence, is called transverse relaxation and is characterized by a time constant T_2 (30).

$$M_{xy}(t) = M_{xy}(0^+)e^{-t/T_2} \quad \text{(Equation 3)}$$

where M_{xy} is the transverse component of M and t is the time. T_2 is when 37% of the transversal magnetization is lost. Transverse relaxation, also called spin-spin relaxation, stems from interactions between neighbouring nuclei that lose their phase coherence. The transverse relaxation is affected by magnetic susceptibility differences and local B_0 inhomogeneity, resulting in a total transverse relaxation constant T_2^* (where $T_2^* \leq T_2$).

1.4.2 Image formation

Spatial localization methods use magnetic gradients to create a spatially varying magnetic field. When a magnetic field gradient G is added in a given direction, e.g. z-direction (G_z) (Figure 2), the protons experience a slightly different magnetic field that depends on their position, and their resonant frequency (ω) is modulated in the following manner:

$$\omega(z) = \gamma(B_0 + G_z z) \quad \text{(Equation 4)}$$

Therefore, the presence of a gradient gives rise to a variation in the resonant frequency in the gradient direction (31), with some spins precessing faster and others precessing slower compared to the ones in the isocentre ($z = 0$). Gradients can be applied in any orientation relative to B_0.
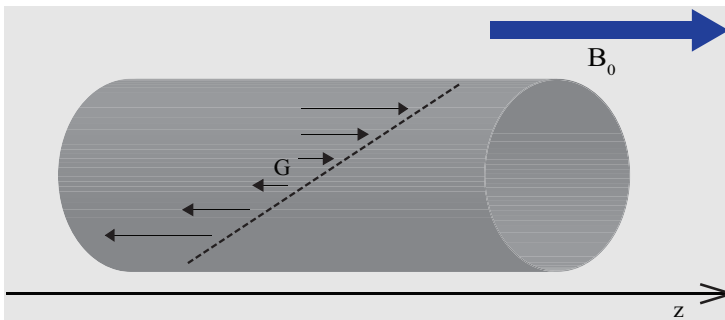
Figure 2. Linear gradient G in the direction of the magnetic field B_0.

<u>Slice selection</u>

In slice selection, an RF pulse is applied simultaneously with a slice-selective gradient, and the excitation is limited to a slice. Due to the slice-selective gradient, the nuclei's resonant frequencies depend on their position along the gradient. By choosing an RF pulse with a centre frequency corresponding to the resonance frequency at the wanted slice position, and a bandwidth corresponding to the wanted slice thickness, only the spins within the slice will be exited and the later detected MR signal will be from this specific slice (32).

$$Slice\ thickness = \frac{RF\ bandwidth}{\gamma G}\ \text{(Equation 5)}$$

<u>In-plane localization</u>

For spatial localization within the slice, two magnetic gradients in the other two directions are used; a phase-encoding gradient and a frequency-encoding (read-out) gradient. After the application of the phase encoding gradient, nuclei will precess at their Larmor frequency, but the phase is changed. The phase of the nuclei will depend on their position along the phase encoding gradient. The frequency-encoding gradient is applied simultaneously as the MR signal is acquired. At the presence of the frequency-encoding gradient, the frequency of the measured signal will depend on its position along the gradient (32).

Spatial localization results in protons in different locations having magnetic moments with a unique combination of frequency and phase. During acquisition, image data are stored in the spatial frequency domain, the k-space. Pulse sequences consisting of RF and gradient pulses acquire data that fill the k-space. Each line of the k-space is filled in the presence of the frequency encoding gradient. The lines are offset from each other by the phase encoding gradient. A pulse sequence diagram describes the process (Figure 3). At the end of a 2D acquisition, when the k-space is fully acquired, it is transformed using the inverse 2D Fourier transform, thus the final grayscale image is produced (33).
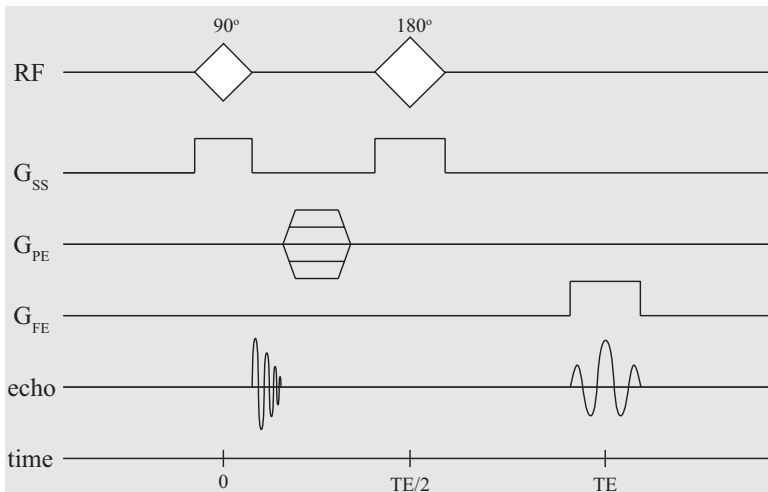
Figure 3. Diagram of a spin echo pulse sequence. The RF pulses are applied simultaneously with the slice-selective gradients (G_{SS}). The combination of RF pulses with G_{SS} results in the restriction of the excitation in a selected slice. In-plane spatial localization is performed using a phase-encoding gradient (G_{PE}) and a frequency-encoding gradient (G_{FE}) in the two orthogonal directions to (G_{SS}). The resulting MR signal is acquired in the presence of G_{FE}.

1.4.3 Image contrast

Tissues are complex, have structures and contain a variety of molecules. Different tissues have varying water and fat contents, and different relaxation properties. One way to produce contrast in MR images is to choose repetition time (TR) and echo time (TE) of the imaging sequence and exploit the difference in relaxation properties of different tissues. TR refers to the time from the application of an excitation pulse to the application of the next excitation pulse and controls the amount of T_1 relaxation. TE refers to the time from the application of an excitation pulse to the time when the MR signal is recorded and controls the amount of T_2 relaxation. A short TR and short TE, compared to the tissue T_1 and T_2 relaxation times respectively, will result in a T_1-weighted image, while using a long TR and a long TE will result in a T_2-weighted image (28). In T_1-weighted images, signals from fat is hyper-intense and signals from fluids are less intense. In T_2-weighted images, signals from both fluids and fat are hyper-intense.

Normal and pathological tissue will likely have different tissue properties and different T_1 and T_2 relaxation times. Appropriate choice of TR and TE, related to the tissue and the pathology

of interest, will create contrast between normal and pathological tissue, making it possible to visualize the pathology.

1.4.4 Short-tau inversion recovery

Short-tau inversion recovery (STIR) is a fat-suppressing MR imaging sequence where a magnetization preparation technique is followed by excitation (34), that exploits the difference in T1 relaxation time between fat and water containing tissue. Specifically, in STIR, a 180° inversion RF pulse flips the longitudinal component of $\boldsymbol{M}$ in the opposite direction. Due to T_1 relaxation, the longitudinal component of $\boldsymbol{M}$ will increase to return to its initial value, passing through null value. After a set amount of time, called inversion time (TI), the 90° RF pulse of the excitation is applied. The choice of TI is done, so that the longitudinal magnetization of fat, the tissue we want to suppress, at the time of the 90° RF pulse is null.

$$TI = T_{1,fat} \cdot \ln (2) \quad \text{(Equation 6)}$$

where $T_{1,fat}$ is the intrinsic longitudinal relaxation time of fat. At 1.5 Tesla, TI is approximately 140 ms. Fat will not emit any signal due to the absence of net magnetization at the time for the RF excitation pulse (Figure 4). Water has a longer T_1 relaxation time and therefore is not nulled and still produces signal. STIR does not provide a tissue specific suppression, but rather suppresses signal emitted from tissue with the short T_1 relaxation time as fat. STIR requires a long TR, resulting in a weighting similar to T_2-weighting (35).
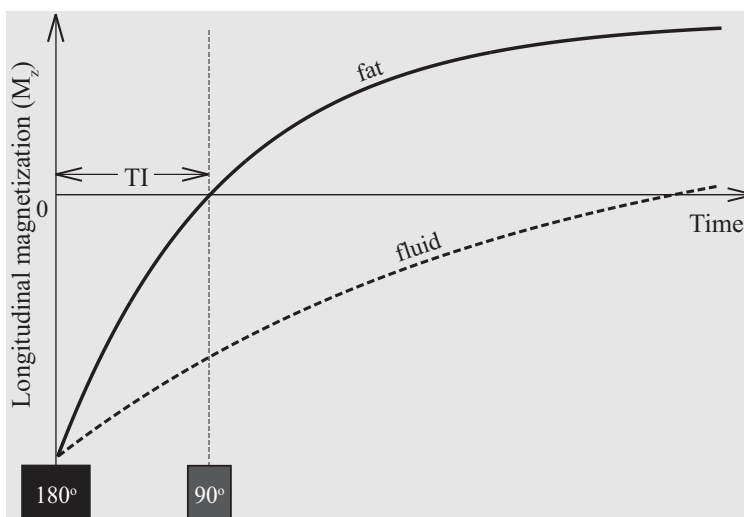
Figure 4. Longitudinal magnetization for fat and fluid when an inversion recovery sequence for fat suppression is applied.

The fat suppression achieved by STIR is not sensitive to heterogeneities in $\boldsymbol{B_0}$ and the effects of magnetic susceptibility, but should not be used after administration of T_1-shortening contrast-agents (36).

1.4.5 Contrast agents

In addition to TR and TE manipulation, exogenous contrast agents can be used to provide contrast in MR images. Each exogenous contrast agents is characterized by its relaxivity, which is its potency to reduce the T_1 and T_2 relaxation times in the tissue where it accumulates. This characteristic determines whether a contrast agent is suitable for contrast enhanced T_1-, T_2- or T_2^*-weighted imaging (37). Gadolinium-based contrast agents are routinely used in MR imaging of cancer (38). They exhibit high relaxivity and low toxicity and can specifically target tumour tissue. Tumour tissue usually has different physiology from normal tissue, including angiogenetic micro-vessels and increased vascular permeability. Clinically available gadolinium-based contrast agents accumulate in the extracellular-extravascular space in the tumour tissue through the leaky and chaotic tumour vessels and provide signal enhancement in T_1-weighted images. A commonly used sequence that makes use of contrast agents is dynamic contrast-enhanced (DCE) MR imaging and is described in detail in chapter 1.4.6.

1.4.6 Dynamic contrast-enhanced MR imaging

DCE MR imaging can provide knowledge about the tumour vasculature by utilizing the effect of a contrast agent in tissue relaxation times. In DCE MR imaging, one or more baseline images are acquired before a contrast agent injection and several contrast-enhanced images are acquired immediately after and at the time following the injection. The basic principle of DCE MR imaging is that the local concentration of the contrast agent in the tissue changes the signal intensity of the tissue in the MR images (39).

When using a conventional T_1-weighted spoiled gradient echo pulse sequence for the acquisition of DCE MR images, the measured signal intensity as a function of time $S(t)$ follows:

$$S(t) = M_0 \, e^{-TE/T_2^*(t)} \frac{sin\theta \, (1 - e^{-TR/T_1(t)})}{1 - cos\theta \, e^{-TR/T_1(t)}} \qquad \text{(Equation 7)}$$

where M_0 is the maximum magnetization value, T_2^* is the total transverse relaxation constant, T_1 is the longitudinal relaxation and θ is the flip angle.

Using the temporal changes in signal intensity, a curve of signal intensity in time is calculated. This curve is called signal enhancement (SE) kinetic curve and follows this form:

$$SE(t)[\%] = 100 \frac{S_{post-contrast}(t) - S_{baseline}}{S_{baseline}} \qquad \text{(Equation 8)}$$

where $SI_{baseline}$ is the signal intensity at baseline and $SI_{post-contrast}$ is the signal intensity at a post-contrast time-point.

In addition to the signal enhancement curve, one can estimate kinetic parameters that relate to physiological parameters of tissue by fitting DCE MR imaging data to a pharmacokinetic model (40). Pharmacokinetic modelling requires MR images with high temporal resolution.

1.5 MR imaging in psoriatic arthritis

MR imaging can be used to assist in the clinical management of psoriatic arthritis (41). MR imaging features of psoriatic arthritis are similar to those of rheumatoid arthritis and spondyloarthritis, including "synovitis, tendonitis, dactylitis, bone marrow oedema, bone erosions, spondylitis, sacroiliitis, soft tissue oedema and subclinical arthropathy" (42).

Sclerosis, fat deposition and ankylosis can be detected with T_1-weighted sequences, while T_2-weighted sequences and STIR might be better in the detection of erosions. In T_1-weighted fat-suppressed post gadolinium administration, active oedematous lesions, intra-articular changes (synovitis), capsulitis and enthesitis are visible. However, for active oedematous lesions, STIR is sufficient, with the exception of synovitis in the sacroiliac (SI) joints.

While its usefulness in differential diagnosis can be argued, MR imaging can give an additional insight and assist the clinician in making a diagnosis and monitoring the effect of intervention by detecting subclinical (asymptomatic) bone marrow oedema, that would otherwise go undiagnosed (41, 43). Fluid-sensitive fat-suppressed sequences, such as STIR, can be used to image oedematous lesions. In the STIR images, fat signal from normal bone marrow is suppressed, and active bone marrow oedema is hyper-intense compared to normal bone marrow. Thus, good lesion contrast is achieved. A pitfall of STIR in psoriatic arthritis is that blood vessels or haemangioma look similar to inflammation, limiting its usefulness in differential diagnosis.

1.5.1 Spondyloarthritis research consortium of Canada Index

There are many methods of radiographic scoring of disease activity and changes in axial psoriatic arthritis (44, 45). To assess bone marrow oedematous lesions in MR images of the spine and the SI joints, the spondyloarthritis research consortium of Canada (SPARCC) method has been proposed. In the SPARCC method, T_2-weighted STIR images of the spine and the SI joints are scored according to the SPARCC SI Joint and Spine Inflammation Indices (46, 47). The presence of oedema accounts for two thirds of the total SPARCC score, and the intensity and depth of the oedema, if present, account for one third.

SPARCC has been used in research settings, when assessing the effectiveness of new treatment methods. The scoring is performed by a trained reader, a radiologist or a rheumatologist, in a specifically designed environment (big screen and controlled settings). It is labour-intensive, taking 10 to 30 minutes per image set per patient depending on disease activity, but it has been shown as a reliable and sensitive to changes method for scoring bone marrow oedematous lesions in MR images of the spine and SI joints of psoriatic arthritis patients (48).

1.6 Dynamic contrast-enhanced MR imaging in breast cancer

According to guidelines from the European society of breast imaging, DCE MR imaging is a useful clinical tool in differential diagnosis, tumour detection, disease monitoring and treatment evaluation in breast cancer (49). DCE MR imaging can visualize new vessel formation in tumours and detect changes in tumour microvasculature due to anti-angiogenetic treatment (50, 51). Angiogenesis is a process where new vessels are formed from pre-existing vessels. While angiogenesis can be a normal physiological process, it is also part of several pathological conditions including cancer, where new vasculature fulfils the demand for nutrients and oxygen supply observed in fast growing tumours. Lastly, DCE MR imaging has potential as prognostic indicator for patient survival (52, 53).

Evaluation of the signal enhancement kinetic curve type has potential in differentiating benign from malignant tumours, achieving sensitivity and specificity of approximately 90% and 60% respectively (54, 55). The shape of the signal enhancement kinetic curve is evaluated by categorizing the washout pattern of the contrast agent. Three types of signal enhancement curves in breast cancer lesions can be recognized (Figure 5): type I (progressive enhancement), type II (rapid enhancement with plateau) and type III (rapid enhancement followed by rapid washout) (56). Progressive enhancement suggests a benign tumour, while rapid enhancement followed by rapid washout is linked with malignant tumours. Texture analysis of enhancement maps has also shown potential in differentiating benign and malignant breast cancer lesions (57). Regarding treatment response, DCE MR imaging can be useful in assessing, but also predicting response to neoadjuvant chemotherapy (NAC) treatment (58-64).
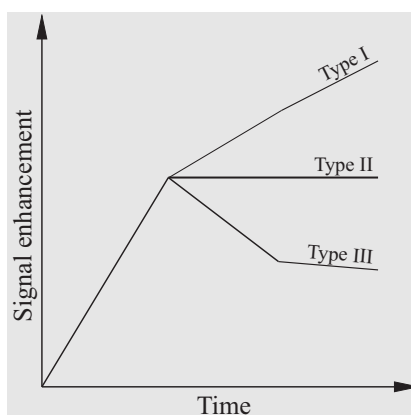
Figure 5. Types of signal enhancement curves in dynamic contrast-enhanced MR imaging: type I (progressive enhancement), type II (rapid enhancement with plateau) and type III (rapid enhancement followed by rapid washout).

1.7 MR image analysis

Radiomics is the extraction of quantitative image features from medical images to assist in establishing a diagnosis, estimating prognosis and predict response to therapy (65). Raw MR images contain noise and artefacts. Pre-processing is the initial step of quantitative MR image analysis that aims to minimize artefacts, reduce noise and generally improve image quality, which in turn can improve diagnostic accuracy. To obtain useful metrics from medical images, their technical validity and clinical usefulness must be established (66). In addition, the context of use, acquisition parameters and processing have to be known (66). Lastly, reliable metrics must be accurate (high repeatability) and precise (65). In this section, the main pre-processing techniques for quantitative analysis that are relevant in the context of this thesis are presented.

1.7.1 Intensity non-uniformity correction

Intensity non-uniformity or bias field is low frequency intensity variation in an image due to imperfect RF field uniformity, the acquisition sequence or patient anatomy (67). The presence of a bias field results in blurry images and altered grey-level distribution in the same tissue across an image. It is a common artefact in MR images that can hinder quantitative image analysis, where the grey-level values of the pixels are analysed. A bias-field correction algorithm is needed as a pre-processing step for images affected by bias field to improve image

quality. Several methods for bias-field corrections have been proposed (68). These either require an extra measurement of the bias field or are based on mathematical models that approximate the bias field (68). In the context of this thesis, a non-parametric non-uniform intensity normalization (N3) bias field correction algorithm (69) was used. N3 algorithms perform great and are insensitive to pathological data, independent of pulse sequence and do not require a tissue model to estimate the bias field and the true tissue intensities (70).

1.7.2 Normalization

Normalization is the process of changing the range of pixel intensity values. It can be a linear or non-linear process that aims at increasing the contrast in the image (histogram equalization) or render comparable images with the same pixel intensity ranges (histogram matching). In MR imaging, changes in imaging protocol parameters and the MR scanner used for different subjects or the same subject in longitudinal studies result in intensity variations, which in turn can affect further image analysis. Normalization is used to allow direct comparisons of images acquired with different settings.

Histogram equalization and matching

Histogram equalization is the process of adjusting the contrast using the intensity histogram of an image. The method modifies the pixel intensities in a way that each pixel has the same relative intensity and that each possible intensity value is represented in the same number of pixels. In other words, it adjusts the pixel intensities so that the image histogram is uniformly distributed. The mathematical description of the conventional histogram equalization has described extensively and more advanced methods have been developed (71). In histogram matching, the histogram of an image is transformed to match the histogram of a reference image (72).

1.7.3 De-noising and sharpening

Noise reduction and sharpening are common ways to improve overall image quality prior to post-processing, like segmentation or edge detection. Noise in MR images is typically Gaussian noise. Anisotropic diffusion, median filtering and wavelet transform are common ways to reduce noise in MR images while preserving the important parts of the image, such as edges, lines and detail. Anisotropic diffusion is a non-linear and space-variant noise reduction technique, that has been described by Perona and Malik (73). In this technique, a family of

parameterized images is produced. Each resulting image is the result of the convolution of the original image with a locally adaptive filter, a filter that depends on the local content of the original image. Thus, preservation of the edges and smoothing of the unimportant parts of the image can be achieved. Median filtering replaces the value of each pixel with the median of the values of its neighbouring pixels and can be used as bandpass filters in images where signal and noise overlap (74). The choice of method depends on the quality of the acquired image and the aim of the analysis. After noise reduction, sharpening increases the contrast between areas with low variation in pixel intensities. Sharpening can be achieved with the use of a high-pass filter that retains high frequency information and suppresses low frequency information in the image.

1.7.4 Segmentation

Image segmentation is the process where an image is labelled in regions containing relevant information (75). The remaining part of the image, that does not contain relevant information, is called background. Segmentation algorithms either detect discontinuities in pixel values (for example, edge detection) or are intensity-based (for example, thresholding and region growing). In the former, the underlying assumption is that pixels containing relevant information are in the same intensity range, which is different from the one background pixels occupy.

<u>Thresholding</u>

Thresholding is an image segmentation method, that results in a binary image (76). In its simplest implementation, it involves defining a range of pixel intensities as belonging to a category, while setting the remaining pixels as background. For example, if a pixel intensity is greater than some fixed constant T, it is replaced in the processed image ($I_{thresholded}$) with a white pixel. Otherwise, it is replaced by a black pixel (Figure 6).

$$I_{thresholded}(x,y) = \begin{cases} 1, if\ I(x,y) \geq T \\ 0, if\ I(x,y) < T \end{cases} \qquad \text{(Equation 9)}$$

A suitable threshold T can be selected manually or automatically by advanced thresholding algorithms, like Otsu's method (77) or histogram-based methods (78). The above implementation for thresholding is more suitable for images with a bimodal histogram, where a small percentage of pixels is misclassified (Figure 7). In other cases, it is possible to use

multiple thresholding, where more than one thresholds are set as cut-off values and region-growing, where the criteria for pixel classification include the proximity of the pixels as well as the similarity in their intensities.
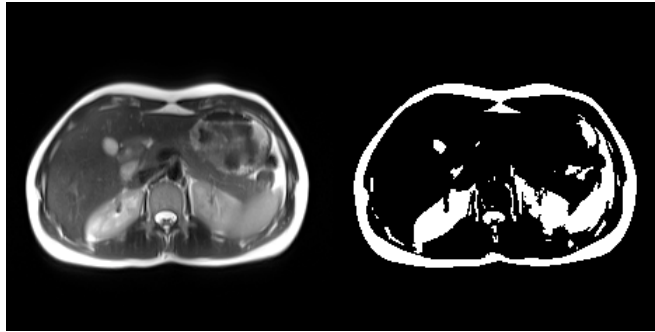
Figure 6. Example of thresholding in abdominal MR image.
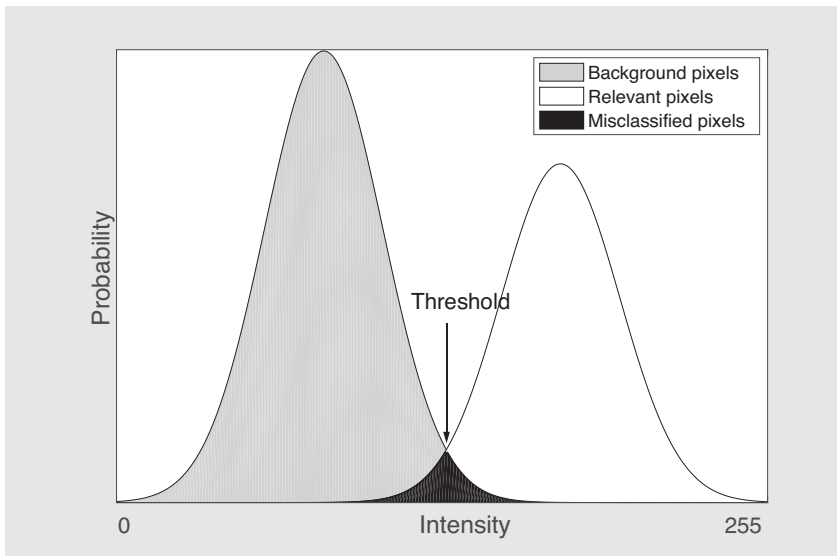
Figure 7. Bimodal histogram of digital image that is suitable for thresholding. A small percentage of pixels is misclassified.

Thresholding can be combined with non-linear operations that rely on the relative ordering of the pixels, making it more robust in noisy images or images with texture (79).

1.7.5 Image feature extraction

Feature extraction is a procedure where information is extracted from the image and is represented in the form of feature vectors. Image features can be shapes, gradients, textures, edges, points, objects etc.

Image gradients are directional changes in the grey-level intensity values in an image. Gradient-based descriptors have been used as features in visual object recognition (80). Texture is defined as the spatial variation of intensity within an image. A human observer can recognize texture (coarseness, contrast, etc.), but has limited ability to define and quantify it. Texture analysis is a method for assessing the spatial position of grey-level intensity of pixels in a digital image. Through texture analysis, a set of metrics that quantify the perceived texture of a digital image, called textural features, are computed from the distribution of pixels (81). Textural features identify the underlying patterns and textures in a digital image and increase the information that we can derive from it (82). In the context of this thesis, texture analysis is reserved for two-dimensional grey-level MR images. However, the principles described in the section can also be extended and applied to three-dimensional digital images as well as coloured images.

The method consists of two steps. In the first step, a matrix, called grey level co-occurrence matrix (GLCM), is created from the original image by counting how often a pair of pixel intensities, spaced apart at a distance d at a given direction φ, occur in an image. All possible pairs of pixels are considered and a matrix is created for each set of d and φ. By varying the distance, one can study different scales of texture within the image, whereas by varying the direction, one can study a particular orientation. Usually four GLCMs in horizontal ($\varphi = 0^o$), vertical ($\varphi = 90^o$) and oblique ($\varphi = 45^o$ and $\varphi = 90^o$) orientations are computed and then the mean GLCM is calculated. In the second step, textural features are derived from the GLCM matrix.

The most commonly used GLCM textural features were mathematically formulated by Haralick (83): angular second moment (f_1), contrast (f_2), correlation (f_3), variance (f_4), inverse difference moment (f_5), sum average (f_6), sum variance (f_7), sum entropy (f_8), entropy

(f_9), difference variance (f_{10}), difference entropy (f_{11}), information measure of correlation one (f_{12}), information measure of correlation two (f_{13}) and maximal correlation coefficient (f_{14}). Two additional features are described by Conners (84): cluster shade (f_{15}) and cluster prominence (f_{16}). The mathematical description of GLCM features is shown in Table 3. While textural features are defined mathematically, their visual representation is not always intuitive and therefore cannot be linked to specific image properties. In addition, several features correlate to each other as they reflect similar properties.

Table 3. Mathematical description of GLCM features.

		Description of feature	Equation
GLCM features	f_1	angular second moment	$\sum_i \sum_j p(i,j)^2$
	f_2	contrast	$\sum_{n=0}^{N-1} n^2\, p_{x-y}(n)$
	f_3	correlation	$\dfrac{\sum_i \sum_j (ij)p(i,j) - \mu_x^2}{\sigma_x^2}$
	f_4	variance	$\sum_i (i - \mu_x)^2 p_x(i)$
	f_5	homogeneity (inverse difference moment)	$\sum_i \sum_j \dfrac{1}{1 + (i-j)^2} p(i,j)$
	f_6	sum average	$\sum_{n=2}^{2N} n p_{x+y}(n)$
	f_7	sum variance	$\sum_{n=2}^{2N} (n - f_6)^2 p_{x+y}(n)$
	f_8	sum entropy	$-\sum_{n=2}^{2N} p_{x+y}(n)\log\left(p_{x+y}(n)\right)$
	f_9	entropy	$-\sum_i \sum_j p(i,j)\log\left(p(i,j)\right)$
	f_{10}	difference variance	$\sum_{n=0}^{N-1} (n - \mu_{x-y})^2 p_{x-y}(n)$
	f_{11}	difference entropy	$-\sum_{n=0}^{N-1} p_{x-y}(n)\log\left(p_{x-y}(n)\right)$
	f_{12}	information measure of correlation one	$\dfrac{f_9 + \sum_i \sum_j p(i,j)\log\left(p_x(i)p_x(j)\right)}{-\sum_i p_x(i)\log\left(p_x(i)\right)}$
	f_{13}	information measure of correlation two	$\sqrt{1 - e^{2(f_9 + \sum_i \sum_j p_x(i)p_x(j)\log\left(p_x(i)p_x(j)\right))}}$
	f_{14}	maximal correlation coefficient	$\sqrt{second\ largest\ eigenvalue\ of\ \sum_k \dfrac{p(i,k)p(j,k)}{p_x(i)p_x(k)}}$
	f_{15}	cluster shade	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i,j)$
	f_{16}	cluster prominence	$\sum_j (i + j - \mu_x - \mu_y)^4 p(i,j)$

N is the number of distinct gray levels in the histogram equalized image.

$p(i,j)$ is the (i,j)-th the entry in a normalized spatial grey level co-occurrence matrix (GLCM).

$p_x(i)$ is the ith entry in the marginal-probability matrix obtained by summing the rows of $p(i,j)$, $\sum_j p(i,j)$.

$p_{x+y}(n) = \sum_i \sum_j p(i,j)$, where $n = 2, 3, \ldots, 4N$ and $i + j = n$

$p_{x-y}(n) = \sum_i \sum_j p(i,j)$, where $n = 0, 1, \ldots, N - 1$ and $|i - j| = n$

μ_x and σ_x are the mean and standard deviation of p_x, respectively.

μ_{x-y} is the mean of p_{x-y}.

1.8 Classification and regression algorithms

Depending on the output, machine learning tasks can be categorized into classification, regression, clustering, density estimation and rule extraction (85). In the context of this thesis, classification and regression are relevant. In classification problems, the task is to predict a discrete number of categorical outputs, while in regression problems, the task is to predict a continuous quantity output (86). The inputs can be continuous, categorical or binary.

Machine learning algorithms can be grouped into four categories: supervised, unsupervised, semi-supervised and reinforcement learning (87). Supervised algorithms tune an adaptive model to a *training dataset*, where the inputs and corresponding outputs are known (86). The performance and predictive accuracy of the model is judged on a *validation dataset*. Then the model can make predictions on new data, where the output is unknown. A problem arises when the number of variables greatly exceeds the number of samples. In this case, because of the high number of variables, some correlate by chance, resulting in good class separation. The model produced is overly optimistic and would fit random datasets (over-fitting) (88). This issue should be considered and corrected for, when using machine learning algorithms.

In the following sub-sections, three algorithms are presented: partial least squares discriminant analysis (PLS-DA), principal component analysis (PCA) and linear mixed-effects models (LMMs).

1.8.1 Principal component analysis

PCA is an orthogonal linear transformation of data to a new co-ordinate system with lower dimensionality that retains as much of the variation that was present in the data as possible (89). The new set of uncorrelated variables are called principal components. PCA is mainly applied as a dimensionality reduction or feature selection method (90). Because of the sensitivity of the variance to outliers, PCA can also be used as an outlier detection method (90).

1.8.2 Partial least squares discriminant analysis

PLS-DA is a supervised classification algorithm that uses least squares methods to find linear discriminators separating the space into regions (Figure 8) (91). Each region contains a group/class. The separation is achieved by maximizing the covariance between two data matrices, one containing the independent variables *X* (raw data) and the second containing the

corresponding dependent variable Y (the group or class each sample belongs to) and finding a linear subspace of explanatory variables (92). Y can be predicted through a reduced number of factors, the PLS components or latent variables. The main PLS-DA equations are

$$X = LP + E \quad \text{(Equation 10)}$$

$$Y = Lq + f \quad \text{(Equation 11)}$$

where L is the matrix of PLS components (latent variables), E and f are residuals, P is the loadings matrix (91). The PLS-DA algorithm has two parts; first the model is built and then predictions can be made in new samples of unknown class or group, or test samples whose class or group is known.
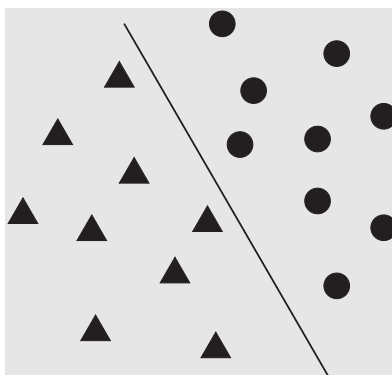
Figure 8. Possible discriminant function for two groups: one group is represented by triangles and the other group is represented by circles.

PLS-DA is useful in the analysis of complex datasets with high dimensionality. It is an appropriate method for analysing noisy data with many variables and few samples. PLS-DA provides a way to identify the most important variables in the classification and offers a visual way of illustrating the differences between the investigated groups. However, the method tends to over-fit the data when the number of included variables greatly exceeds the number of samples.

1.8.3 Linear mixed-effects models

LMMs are an extension of linear regression models. In linear regression, the output variable (dependent variable) is expressed as a linear combination of the input variables (independent/explanatory variables). In a mixed effects model, the independent variables are considered as either fixed effects, which are systematic and controlled, or random effects, which are unsystematic effects not accounted for by the fixed effects. An LMM can be represented as

$$y = X\beta + Zb + \varepsilon \qquad \text{(Equation 12)}$$

where y is the dependent variable vector, X is the matrix containing the fixed effects, β is the fixed effects coefficients vector, Z is the matrix containing the random effects, b is the random effects coefficients vector and ε is the observation error vector (93). The parameters of the model are estimated using the least squares method on a training dataset. It should be noted that if there is no linear relationship between the dependent variable and the independent variables, a logarithmic transformation can be used to conform to the linearity assumption of the model.

2 Thesis objectives

The main objective of this thesis was to explore radiomics analysis of MR images in order to obtain quantitative and objective MR image-based measures. More specifically, the research carried out in this thesis aimed to:

- Establish a framework for processing and analysis of data acquired longitudinally and with different scanners and protocols.
- Segment subtle bone marrow oedema in the spine and the SI joints of patients with psoriatic arthritis. (Paper 1)
- Evaluate the effect of HIIT in psoriatic arthritis patients by assessing bone marrow oedema in MR images of the spine at baseline and after intervention. (Paper 2)
- Predict seven-year overall in LABC patients based on textural features extracted from DCE MR images. (Paper 3)

3   Materials and methods

This thesis includes three papers in MR image analysis. Study A (paper I and II) focuses on the analysis of MR images of bone marrow for the detection and quantification of bone marrow oedema, and for the assessment of HIIT intervention effect in psoriatic arthritis patients. Study B (paper III) aims to predict survival through imaging biomarkers obtained prior to NAC in LABC patients. The materials and methods used in papers I-III are briefly outlined in Table 4.

Table 4. Outline of material and methods used in this thesis.

		Study A		Study B
		Paper I	Paper II	Paper III
Patient cohort		43 psoriatic arthritis patients	37 psoriatic arthritis patients	55 LABC patients
Data acquisition	MR imaging protocol	STIR of the spine and the SI joints	STIR of the spine	DCE MR imaging of the breast
Image pre-processing	Bias field correction	-	N4 bias field correction algorithm	-
	Image normalization and filtering	- Histogram matching	- Histogram matching - Anisotropic diffusion filter	- Histogram equalization
	ROI segmentation	Manually drawn	Manually drawn	Manually drawn and refined with thresholding and morphological operations
Image analysis	Extraction of quantitative measures	Thresholding	Texture and gradient analysis	Texture analysis
Statistical methods	Classification and regression	- LMMs	- LMMs	- LMMs - Orthogonalized PLS-DA
DCE: dynamic contrast-enhanced, LABC: locally advanced breast cancer, LMM: linear mixed-effect model, MR: magnetic resonance, PLS-DA: partial least squares discriminant analysis, ROI: region of interest, SI: sacroiliac, STIR: short-tau inversion recovery				

3.1 Patient cohorts

The studies have been approved by The Regional Committee for Medical and Health Research Ethics (Study A: REK number 2012/1646, Study B: REK number 2009/112). All patients signed informed consents. Two different patient populations were employed.

For study A, the cohort included patients diagnosed with psoriatic arthritis, participating in a randomized clinical trial with HIIT as intervention. Included patients fulfilled the CASPAR criteria for psoriatic arthritis, were between 18 and 65 years old and were able to exercise. Exclusion criteria were unstable psoriatic arthritis, unstable ischemic vascular disease, severe pulmonary disease, pregnancy, breastfeeding and drug or alcohol addictions. During the inclusion period 2013-2015, 43 patients were recruited and examined by MR imaging at Aleris Røntgen Trondheim. In paper II, a subset of data acquired for paper I was included.

For study B, the retrospective patient cohort consisted of LABC patients undergoing NAC prior to surgery. During the inclusion period 2008-2011, 55 patients were recruited and examined in St Olavs hospital in Trondheim, Norway.

3.2 Data acquisition

For study A, patients underwent MR imaging examinations on two 1.5T scanners (Scanner 1: Syngo MR B17 Model upgraded during the study to B19, Avanto, Siemens Healthcare, Germany, Scanner 2: Syngo MR D13, Avanto, Siemens Healthcare, Germany) at three time-points; before intervention, and three and nine months after the beginning of intervention. MR images were acquired using an inversion recovery-based sequence (STIR) in the spine in two stations (sagittal orientation; TR/TE/TI: 4250 ms/51 ms for the lower spine or 52 ms for the upper spine/145 ms; slice thickness: 4 mm for the lower spine or 3 mm for the upper spine) and the SI joints (semi-coronal orientation; TR/TE/TI: 3700/52/145 ms; slice thickness: 4 mm). T1 and T2 weighted sequences were acquired for anatomical reference (46, 47). In paper II, only MR images acquired at the first two time-points were included.

For study B, patients underwent an MR imaging examination of the breast before the start of NAC. MR images were acquired on a 3T Siemens Tim Trio (Erlangen, Germany) scanner using a dedicated four-channel breast coil. MR protocol included a T_1- weighted DCE MR imaging acquired with a three-dimensional RF-spoiled gradient-echo sequence without fat suppression (flip angle: 10º and TR/TE: 3.22/1.22 ms [flip angle: 6º and TR/TE: 3.5/1.2 ms for

the first 17 patients]). Images were reconstructed in the coronal view with in-plane resolution 1.1x1.1 mm^2, matrix size: 320x320; and slice thickness: 1.1-1.5 mm. After the acquisition of one baseline image, a bolus injection of 0.1 mmol/kg of body weight of gadolinium-based contrast agent (Magnevist [Bayer Healthcare Pharmaceuticals, Montville, NJ, US] for the first nine examinations; or Omniscan [GE Healthcare, Oslo, Norway] for the rest) was given automatically at a rate of 2 ml/s, followed by a 20 ml saline flush. After contrast agent administration, seven post-contrast images were acquired with a temporal resolution of one minute.

3.3 Reference standard

In study A, a trained rheumatologist scored the STIR images of the spine and SI joints for the presence, intensity and depth of oedema using the SPARCC scoring methods (46, 47). The presence of oedema accounts for two thirds of the maximum score, while intensity and depth account for the rest one third. The total maximum score for the spine and the SI joints is 108 and 72, respectively. Figure 9 illustrates the methodology for SPARCC scoring in the SI joints. In paper II, a radiologist who was blinded with respect to the intervention evaluated the images and provided psoriatic arthritis lesion segmentations. Images were also categorized according to the change from the first to the second scan as stable, increased or reduced bone marrow oedema.

In study B, patient status seven years post diagnosis was used as classification output. Patients with relapse or residual tumour that were alive seven years post-diagnosis, were considered overall survivors.
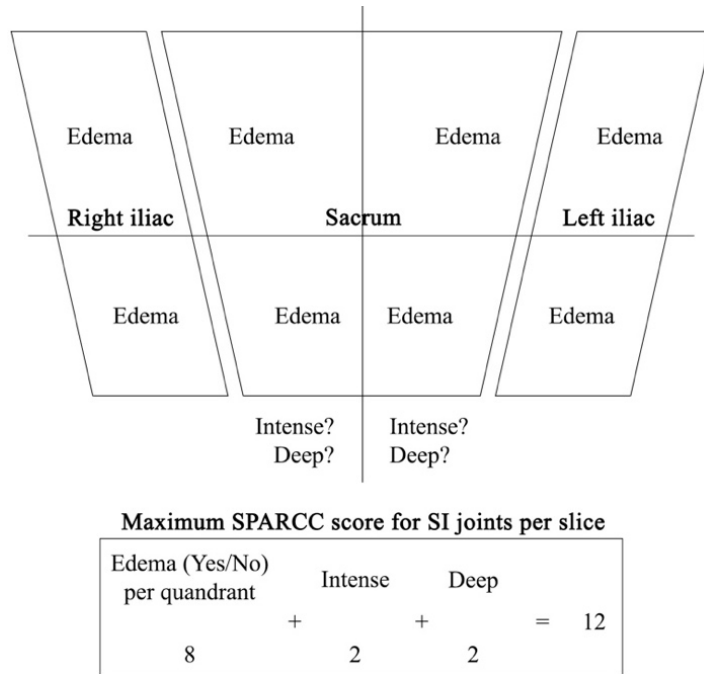
Figure 9. Illustration of SPARCC scoring methodology in the SI joints.

3.4 Image quality assurance protocol

In study A, a quality assurance protocol using the American College of Radiology (ACR) MR imaging accreditation phantom test was used (94). The test evaluates MR image quality through the assessment of geometric accuracy, high-contrast resolution, slice thickness accuracy, slice position accuracy, image intensity uniformity, percent signal ghosting, low-contrast object detectability, signal-to-noise ratio and central frequency. Phantom images were obtained using the sequences described in the ACR MR imaging accreditation protocol, as well as in the MR protocol of study A. Images were there analysed following the ACR large phantom test guidance (95).

3.5 Image pre-processing

In paper I, all spinal MR images were histogram-matched to one reference spinal image and all MR images from SI joints were histogram-matched to one reference SI joint image.

Both in papers I and II, bone marrow of the sacrum and the iliac bones in the SI joints and vertebral bone marrow in the spine, excluding vascular and neural structures, were manually outlined using 3D Slicer (MIT Artificial Intelligence Lab, USA).

In paper II, spinal images were pre-processed using a customized intensity adjustment procedure based on the nonparametric nonuniform intensity normalization bias field correction algorithm (69) which was applied to the region of the lumbar spinal column. Image histogram normalization was performed to standardize the intensity values by matching the histogram (extracted from the spinal column) to the histogram of a randomly selected spinal image. An anisotropic (Perona-Malik) diffusion filter was then applied in MATLAB R2017b (MathWorks, Natick, MA, USA) (96).

Image pre-processing in study B was performed by a previous PhD student at the MR Cancer group. In brief, images were motion-corrected by a non-linear image registration using the FNIRT tool in the FSL package (Oxford FMRIB Centre, University of Oxford, Oxford, UK) optimized for breast images. Segmentation of breast cancer lesions was performed on post-contrast subtracted images at two minutes post-injection by manual segmenting contrast enhancing region, thresholding regions with a relative enhancement ration higher than 0.2 and applying an opening morphological operation. The segmenting procedure should exclude pixels of necrotic, healthy, and fatty tissue, and vessels. Histogram equalization to 32 levels was applied only to the segmented regions.

3.6 Image analysis

Table 5 summarizes the methods that were used for the extraction of quantitative measures.

Table 5. Methods for extraction of quantitative measures.

	Study A		Study B
	Paper I	**Paper II**	**Paper III**
	Thresholding	Feature extraction	
Type of implementation	Pixel-wise approach		Region-based approach
Information	Grey-level intensity	Grey-level intensity, gradient and textural features	Textural features

In paper I, quantitative measures of bone marrow oedematous lesions were extracted from images using thresholding. The threshold was defined as the

$$Threshold = \mu + a \cdot \sigma \qquad \text{(Equation 13)}$$

Where μ is the mean signal intensity in the reference normal bone marrow region of interest (ROI), a is the optimal percentage and σ is the standard deviation of the signal intensities in that ROI. The reference normal bone marrow ROI was chosen as a circular ROI (≥ 200 pixels) at a healthy vertebra in one slice of the spinal image series and at the centre of the first sacral vertebra in one slice of the SI joint image series. The optimal threshold for the spine and the SI joints was explored by a receiver operatic characteristic curve analysis. For the spine, the optimal threshold was found to be $\mu + 4.15 \cdot \sigma$, while for the SI joints, the optimal threshold was $\mu + 2.64 \cdot \sigma$.

Thresholding provided a mask of pixels with higher signal intensity that the applied threshold, which was consistent with inflammation (97). Connected components of less than 10 pixels were considered artefacts and were removed. All hyper-intense pixels were normalized to the mean signal intensity of normal bone marrow. Thresholding provided three quantitative metrics: the volume covered by STIR hyper-intense pixels, the relative hyper-intensity of STIR hyper-intense pixels ($S_{RelHyper}$) to normal bone marrowand the number of lesions per image

set. For each image set the total volume and the mean, median, 75-percentile and 90-percentile of $S_{RelHyper}$ were calculated for all image sets.

In paper II, texture and gradient feature maps were calculated from bone marrow ROIs in STIR spinal images of psoriatic arthritis patients. The feature set that was calculated included seven intensity features; grey-level intensity value of the central pixel in the ROI, mean, median, standard deviation, minimum, maximum and semi-interquartile range of the grey-level intensity values in the ROI. In addition to the intensity features, ten gradient features were part of the feature set. For the extraction of the gradient features, 2-dimensional directional gradients for x-axis (G_x) and y-axis (G_y) were calculated using a Sobel gradient operator in the *imgradientxy* function in MATLAB R2017b. The $L1_{norm}$ and $L2_{norm}$ of the directional gradients were calculated based on equations 14 and 15, respectively.

$$L1_{norm} = |G_x| + |G_y| \qquad \text{(Equation 14)}$$

$$L2_{norm} = \sqrt{G_x{}^2 + G_y{}^2} \qquad \text{(Equation 15)}$$

The ten gradient features were the sum, mean, standard deviation, median, minimum, maximum and semi-interquartile range of $L1_{norm}$ inside the ROI and the sum, mean and standard deviation of $L2_{norm}$ inside the ROI. Lastly, four GLCM features were extracted using the *graycomatrix* and *graycoprops* functions in MATLAB R2017b; energy ($\sqrt{f_1}$), f_2, f_3 and f_5. An overview of the features used in Paper II is shown in Table 6.

Table 6. Feature set description (Paper II).

		Description of feature
Intensity features	i_1	Grey-level intensity value of the central pixel
	i_2	Mean of grey-level intensity values
	i_3	Median of grey-level intensity values
	i_4	Standard deviation of grey-level intensity values
	i_5	Minimum of grey-level intensity values
	i_6	Maximum of grey-level intensity values
	i_7	Semi-interquartile range of the grey-level intensity values
Gradient features	g_1	Sum of $L1_{norm}$
	g_2	Sum of $L2_{norm}$
	g_3	Mean of $L1_{norm}$
	g_4	Mean of $L2_{norm}$
	g_5	Standard deviation of $L1_{norm}$
	g_6	Standard deviation of $L2_{norm}$
	g_7	Median of $L1_{norm}$
	g_8	Minimum of $L1_{norm}$
	g_9	Maximum of $L1_{norm}$
	g_{10}	Semi-interquartile range of $L1_{norm}$
GLCM features*	$\sqrt{f_1}$	energy
	f_2	contrast
	f_3	correlation
	f_5	homogeneity (inverse difference moment)

GLCM: grey level co-occurrence matrix

$L1_{norm} = |G_x| + |G_y|$, G_x and G_y **are the 2-dimensional directional gradients for x-axis and y-axis, respectively.**

$$L2_{norm} = \sqrt{G_x^2 + G_y^2}$$

* Mathematical descriptions of the GLCM features are shown in Table 3.

Feature maps were calculated using a sliding window implementation. A sliding window is the operation, where an orthogonal kernel of a certain size slides from top left of the ROI row-by-row down to bottom right. At every position of the kernel, texture analysis is performed using

the pixels inside the kernel as input. The resulting features corresponded to the pixel in the centre of the kernel. In this case, the ROI was the bone marrow of each vertebra in the spine. The analysis was performed using a three-by-three pixels kernel. Figure 10 illustrates a sliding window implementation.
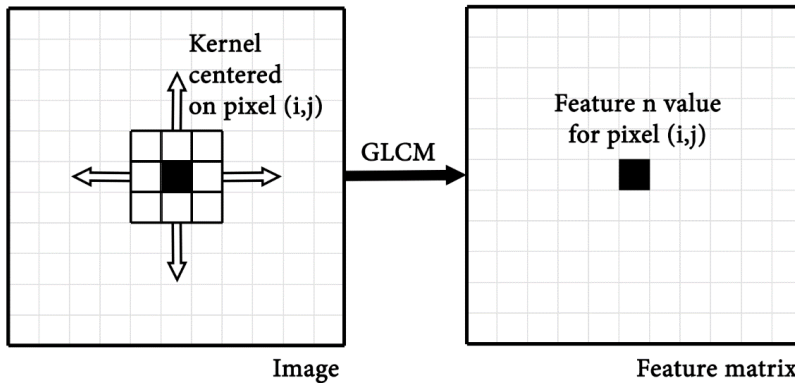
Figure 10. Illustration of a sliding window operation with a three-by-three pixels kernel.

Texture analysis in paper III was performed by a previous PhD student at the MR Cancer group. In brief, texture analysis was performed using two-dimensional GLCMs. From these matrices, 16 features were computed. 14 out of the 16 texture features were described by Haralick (83) and the additional two features were described by Conners (84). A mathematical description of GLCM features is shown in the introduction. While textural features are defined mathematically, their visual representation is not always intuitive and therefore cannot be linked to specific image properties. In addition, several features correlate to each other as they reflect similar properties.

Texture analysis was performed using in-house MATLAB scripts developed at Centre for Magnetic Resonance Investigations, Hull York Medical School at University of Hull, Hull, UK.

3.7 Statistical methods

The work carried out in this thesis required advanced image processing and analysis. Non-commercial in-house software developed at NTNU using MATLAB R2017b, in addition to the

commercial software R (R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistics (IBM SPSS Statistics for Macintosh, Version 22.0) were used.

In all studies, LMMs were employed to assess the effect of time-related changes. In paper I, the effect of the MR scanner and intervention were also considered in LMMs. LMMs (98) were built in R 3.1.1 using the function lme from the 'nlme' package (99) employing the method of restricted maximum likelihood. In addition to the work included in Paper II that evaluated the effect of HIIT in patients with psoriasis arthritis, an orthogonalized PLS-DA classifier was used to segment axial psoriatic arthritis lesions using intensity, gradient and textural features. In paper III, orthogonalized PLS-DA was used for the classification of patients based on their seven-year overall survival status (PLS toolbox version 8.2.1, Eigenvector Research, Inc., Washington, USA in MATLAB MathWorks, Natick, MA, USA). Random subsets with 10 data splits and 20 iterations were used for cross-validation.

The specific methods used in each paper are described in detail at the *Material and Methods* section of each paper.

4 Summary of papers

4.1 Paper I

Quantifying bone marrow inflammatory oedema in the spine and sacroiliac joints with thresholding

Ioanna Chronaiou, Ruth S. Thomsen, Else M. Huuse, Leslie R. Euceda, Susanne J. Pedersen, Mari Hoff, Beathe Sitter

The main purpose of this paper was to assess the effectiveness of thresholding in the quantification of bone marrow oedema in the spine and SI joints of psoriatic arthritis patients and compare it with the SPARCC scoring system.

STIR images of the spine (N=85) and the SI joints (N=95) of patients with psoriatic arthritis (N=41) were acquired on two 1.5T Siemens Avanto scanners at three time-points. After normalization, bone marrow was manually segmented on MR images and a threshold was applied to visible bone marrow. The hyper-intense, compared to normal bone marrow, areas were masked and were considered as oedematous lesions. The volume and relative signal intensity of the hyper-intense areas were calculated and were compared to SPARCC spine and SI joint indices, through LMMs. The effect of the MR scanner, intervention and time of acquisition were assessed.

For both the spine and the SI joints, LMMs showed a significant positive correlation between the volumes and relative signal intensities of the hyper-intense areas, and the SPARCC indices. In the spine, thresholding provided a sensitivity and specificity of 90% and 84%, respectively, in the spine and 49% and 68% respectively, in the SI joints. Thresholding performed better in the spine that the SI joints, possibly due to signal intensity in homogeneities within the SI joints images and signal intensity differences between the SI joints images that were not corrected by normalization. Such differences were not present in the spinal images. The MR scanner, intervention and time of acquisition did not have a significant effect on the volumes and relative signal intensities.

In conclusion, thresholding allows quantification of bone marrow oedematous lesions in psoriatic arthritis patients and has potential in becoming an automated alternative to SPARCC scoring system in the spine.

4.2 Paper II

Evaluating the impact of high intensity interval training on axial psoriatic arthritis based on MR images

Ioanna Chronaiou, Guro Fanneløb Giskeødegård, Ales Neubert, Tamara V. Hoffmann-Skjøstad, Ruth S. Thomsen, Mari Hoff, Tone F. Bathen, Beathe Sitter

The purpose of the study was to evaluate the effect of HIIT in psoriatic arthritis patients by investigating spinal MR images at baseline and after intervention. Patients were randomized into an intervention group that performed HIIT three times per week for 11 weeks, and a control group with no change in pre-study physical exercise habits.

STIR and T1-weighted MR images of the spine of patients with psoriatic arthritis (N=37) were acquired on 1.5T Siemens Avanto scanners at baseline and after intervention. All images were evaluated by a radiologist for bone marrow oedema and scored by a trained rheumatologist using the SPARCC scoring system. Vertebral bone marrow was manually segmented from STIR MR images. Segments were further bias field corrected, noise reducing image smoothing algorithms were applied in pre-processing and the MR signal intensities were normalized to an atlas using histogram matching. For each image pixel, 21 features were extracted using local image intensity statistics with a three-by-three pixels neighbourhood kernel.

MR imaging established that the disease burden in terms of bone marrow edema showed minor variations in both groups. We observed no differences between the intervention and control group. Disease activity score (DAS) in 44 joints and four textural features of psoriatic arthritis lesions decreased for both groups, but the textural features changes were not significant after Bonferroni correction.

In conclusion, MR imaging showed that bone marrow edema in psoriatic arthritis did not change significantly after HIIT, supporting that HIIT is safe for psoriatic arthritis patients.

Texture analysis of spinal MR images could be more sensitive to changes in BME than radiological evaluation and SPARCC scoring.

4.3 Paper III

Feasibility of contrast-enhanced MRI derived textural features to predict overall survival in locally advanced breast cancer

Ioanna Chronaiou, Guro F. Giskeødegård, Pål E. Goa, Jose Teruel, Roja Hedayati, Steinar Lundgren, Else M. Huuse, Martin Pickles, Peter Gibbs, Beathe Sitter and Tone F. Bathen

This paper aimed to evaluate the ability of GLCM textural features extracted from post-contrast DCE MR images obtained before NAC to predict seven-year overall survival in LABC patients for better treatment stratification.

One baseline and seven post-contrast pre-treatment DCE MR images from 55 LABC patients were acquired at a 3T Siemens Trio scanner. 16 GLCM textural features were extracted from segmented tumours at each time-point. The ability of GLCM textural features from all time-points to distinguish patients based on their seven-year overall survival status was assessed through LMMs. The time-point with the highest classification accuracy between seven-year overall survivors and non-survivors was established through orthogonalized PLS-DA. The predictive value of GLCM textural features from this time-point, alone or in combination with traditional prognostic factors, was evaluated by PLS-DA and using Kaplan-Meier analysis with log-rank test.

LMMs revealed significant differences in five GLCM features (f_1, f_2, f_5, f_{10}, f_{11}) in survivors and non-survivors. In the PLS-DA model, the same features, as well as f_3 and f_9, from two minutes post-contrast images were the most different and resulted in a classification accuracy of 73.1% (p<0.001), compared to 67.2% (p=0.005) and 77.8% (p<0.001) achieved by traditional prognostic factors and a combination of these with textural features, respectively. The median value of f_1, f_2, f_{10} and f_{11} from two minutes post-contrast provided the most significantly different survival curves in Kaplan-Meier analysis.

In conclusion, GLCM textural features from pre-treatment post-contrast DCE images were found to be associated with seven-year overall survival. Prognostic value of textural features should be investigated further.

4.4 Additional analysis

In addition to the work included in Paper II that evaluated the effect of HIIT in patients with psoriasis arthritis, it was investigated whether machine learning can segment axial psoriatic arthritis lesions using intensity, gradient and textural features. The classification model was established using an orthogonalized PLS-DA classifier built on extracted image features from 90% of the patients, using a leave one patient out (LOPO) cross-validation. The number of latent variables was optimized using an internal LOPO cross-validation. Data from the remaining 10% patients were used to test the model. The 21 features that were calculated for each pixel were used as predictors. Healthy pixels in the training dataset were under-sampled at a ratio of 1:1 (healthy to pathological) to achieve a more balanced dataset and avoid a model bias towards the over-presented class in the dataset. Model evaluation was performed using sensitivity and specificity metrics on the test dataset. PLS-DA classification achieved a sensitivity and specificity of 97% and 96%, respectively, for the presence of bone marrow edema.

To assess the effect of neighbourhood kernel size in feature values, features were extracted using both a three-by-three and a five-by-five kernel size. All features, but one (f_3 calculated with a 5-by-5 pixel kernel), were significantly different between healthy and pathological bone marrow ($p<0.05$).

5 Discussion

The expanding use of MR imaging as an assisting tool in the diagnosis, staging, monitoring progression and evaluating treatment response in a variety of diseases makes it an interesting target for clinically relevant research (100-102). MR imaging has become a cornerstone in clinical evaluations performed at a hospital, resulting in more examinations per patient. Additionally, new methods emerge and multi-parametric MR imaging is common. As a result, MR imaging produces an increasing amount of data that must be analysed and interpreted. Qualitative MR analysis is based on visual inspection of the images, while quantitative MR imaging uses the MR scanner as scientific measuring instrument and allows us to measure the burden of disease, assess its progression and evaluate treatment responses (103). Most clinicians rely on qualitative image analysis, which requires long training, is time-consuming and is subject to inter-reader variance. With an increasing amount of information available to clinicians, radiomics analysis can assist in the time-consuming interpretation of complex datasets with automation and the extraction of useful information for decision making (104).

The main challenge in the field of radiomics is the lack of standardization in the methodologies used for image acquisition and pre-processing, segmentation of targeted regions and feature extraction (105-108), which hinders the comparison of results from different studies. Several features depend on acquisition and image reconstruction parameters (106, 107) and there is an increasing amount of research assessing the reproducibility and repeatability of radiomics features (109-111). The choice of segmentation method will also affect the repeatability of the features (106). It is important that segmentations should be consistent and exhibit low inter-observer variability (107, 108). In addition, different implementations for the calculation of the same feature value impact the interpretation of the study results (106, 107). It is therefore necessary to standardize the way that features are calculated (107). Radiomics also suffers from the curse of dimensionality (112), because the number of calculated features is often much larger than the sample size (106, 107). One solution to this problem is to use feature selection or dimensionality reduction methods, which can help reduce the false discovery rate and the risk of overfitting (106-108). Another way to avoid finding correlations by chance is to correct for multiple hypothesis testing (107). Lastly, an external validation dataset is recommended for the confirmation of study results (107).

The papers included in this thesis aim to facilitate the development of quantitative analysis methods of MR images, with applications demonstrated in breast cancer and psoriatic arthritis.

In the context of this thesis, two methods for segmentation of bone marrow oedema in STIR images were explored: thresholding and classification using intensity, gradient and textural features as predictors. Baseline and after intervention values of these features were used to assess the effect of HIIT in psoriatic arthritis patients. Textural features were also extracted from pre-treatment DCE MR images of breast cancer patients for prediction of long-term outcome. A routine including quality control, normalization, and statistical analysis was implemented to pool data acquired longitudinally and with different scanners and protocols. The main findings of the research presented in this thesis are discussed in the following sections.

5.1 Processing MR data acquired at multiple sites or with different protocols

In a clinical setting, it is common that the imaging parameters vary slightly between the patients. This could be due to protocol optimization for each patient (e.g. a change of field of view or acquisition matrix may result in automatic optimization of TE) or due to the examinations being performed at different scanners with different possibilities and limitations. In daily clinical practice, a radiologist relies on the visual inspection of the images to decide on a diagnosis (113). In this qualitative approach, the contrast between tissues is important, but the actual values of the pixel intensities are not. In quantitative approaches, images are treated as data and are automatically analysed by algorithms (104). For quantitative comparisons, variability due to the use of different scanners, or variability occurring from a longitudinal setting, should be controlled (114). To some extent, minimizing such differences is possible by implementing a standardized protocol, relevant to the study, at all scanners for all patients. However, protocol standardization across hardware setups, field strengths and vendors can be challenging (115). Differences in the hardware (e.g. different coils for different scanners) may result in different signal-to-noise ratio and image quality, while different vendors may allow slightly different implementations and acquisition parameters. Additionally, scanner upgrades may be impossible to delay in longitudinal clinical studies and can result in differences in the protocols, reconstruction algorithms, correction algorithms or hardware. In general, scanner upgrades improve the performance of the scanner and therefore may improve the clinical results. Metrics that are robust against protocol and scanner changes are preferred to ensure that research results can be translated in clinical settings, where data acquisition is affected by multiple parameters including the hardware, the software and the radiographer (116). A

specific framework for the interpretation of data acquired and the evaluation of the calculated metrics in different settings should be designed.

In study A, quality control and system performance testing procedure were performed. The standardized image quality assurance protocol and phantom developed by the ACR MR imaging accreditation program were used in accordance with the ACR large phantom test guidance (95). Advantages of the ACR phantom test include the use of a single phantom appropriate for nearly all clinical MR scanners, the efficient testing procedure (total examination time was approximately one hour per scanner) and the assessment of both qualitative and quantitative performance of the scanner (117). The chosen procedure assessed the overall quality of the image (signal-to-noise ratio, geometrical accuracy and distortions, image intensity uniformity, slice thickness and position accuracy, object detectability and ghosting). The phantom was scanned once at each of the scanners, after all patient examinations were performed, to control that the scanners were properly working and provide comparable image quality. Ideally, phantom testing should be performed longitudinally in longitudinal studies to ensure the performance stability of the scanners. The image quality assurance step is necessary in longitudinal and/or multi-site studies (118).

The use of multiple scanners or protocols may induce intensity variations due to scanner-related and image weighting differences, thus hindering further quantitative analysis or post-processing. This is particularly evident in retrospective studies, where there was no intention to pool the data. The studies included in this thesis were designed so that these variations were minimal. For study A, a standardized protocol was implemented in the two scanners and for study B, only one scanner was used for data acquisition. Depending on the study design, there are several methods that correct for intensity non-uniformity in MR imaging (119). In all three papers included in this thesis, histogram equalization or matching was used to normalize the images and correct for intensity variations between them. This ensured that images acquired at multiple time-points or with multiple protocols could be compared and analysed together.

In study A, MR imaging examinations were performed in two scanners at three time-points. Histogram matching was used as a normalization method to create similar intensity distributions between the images. In this study, the MR protocol for the lower spine was changed due to human error for ten patients, examined in the same scanner, at one time-point. For these patients, TR was 3500 ms and TE was 31 ms (in contrast with the rest of the patients,

for which TR was 4250 ms and TE was 51 ms TR was 4250 ms), resulting in different image weighting. Histogram matching was applied to these images but was not sufficient to obtain the same contrast between tissues that was necessary for quantitative analysis, and a decision was made to exclude these scans. In study B, the imaging protocol was not consistent throughout the study due to a platform upgrade and a successive sequence update after the first 17 patients had been examined. This resulted in different imaging parameters (TR, TE and flip angle) in these patients, which in turn had an effect on the maximum enhancement. Histogram equalization was performed prior to textural feature extraction to minimize the effect of the change in the imaging parameters.

As a last step, the interaction between the scanner and the MR image-derived measurements of interest (signal intensities and textural features) had to be accounted for by statistical methods. Especially for texture analysis, there is a need to assess the sensitivity of textural parameters to the acquisition protocol changes and the normalization method (65, 120, 121). LMMs can account for the influence of between-subject and within-subject variations to MR image-derived measurements due to the MR scanner and acquisition parameter changes (122). In the studies included in this thesis, the scanner or protocol was included as a fixed effect in LMMs, to examine and account for the possibility that any observed effects to MR image-derived measurements are confounded by the changes in the MR scanner or protocol. In study A, the effect of the MR scanner on the measurements by the thresholding method was not significant. In study B, several textural features were significantly affected by the change in acquisition protocol. Most of these features were not significantly different between patient groups of interest. When data after the scanner upgrade and the protocol change was analysed separately, the same features were found important in the classification between patient groups and the difference was more significant, suggesting that the protocol change could be masking group differences. A decision was made to include all the data, prior and after the scanner upgrade, as these circumstances relate more to a clinical setting, where standardization may not be possible.

Based on the methods applied in the papers included in this thesis, the suggested steps to proper interpretation of these data is outlined in Table 7 and discussed in this section. A more general outline of evaluation criteria is discussed in (65).

Table 7. Steps for appropriate quantitative interpretation of MR imaging data acquired by multiple scanners and/or protocols.

Image acquisition phase	Run an image quality protocol using phantom or human control.
Image pre-processing phase	Apply intensity non-uniformity correction (ex. histogram matching) on acquired images.
Statistical analysis phase	Use statistical models that include confounding variables and can account for the change in the scanner and/or the protocol (e.g. LMMs).

5.2 Classification problems: Pixel-wise or region-based approach

Texture analysis was performed in papers II and III. In paper II, features were calculated in a pixel-wise approach where the features were extracted in the immediate neighbourhood of each pixel using a three-by-three kernel, resulting in a feature map for each feature calculated. In paper III, a region-based approach was used, where the calculated features refer to the whole ROI. In general, the choice of the appropriate approach depends on the research question. Both pixel-wise and region-based approaches have advantages and limitations.

The pixel-wise approach gives more detailed information about the part of the image that is analysed, provides input for pixel-wise classification and potentially could be used for lesion segmentation. It is particularly suitable when a small part of the image that has specific characteristics has to be identified or in images with high to very high spatial resolution (123). In paper II, several psoriatic arthritis patients exhibited small oedematous lesions, compared to the size of their vertebra (Figure 11). In this case, a region-based approach, that would calculate mean features from the whole vertebra, would be less appropriate, because the number of pixels that contained healthy tissue was much higher than the number of pixels containing oedema. The features then reflect the physiology of the healthy tissue, unless the inflammatory burden was larger, resulting in undetected lesions. The main limitation of the pixel-wise approach is that pixel-wise feature extraction is computationally intensive, compared to region-based approaches. Feature values are calculated by some operation, for example averaging, that takes

into account the values of the neighboring voxels. This could be a limitation in some applications, such as edge detection.
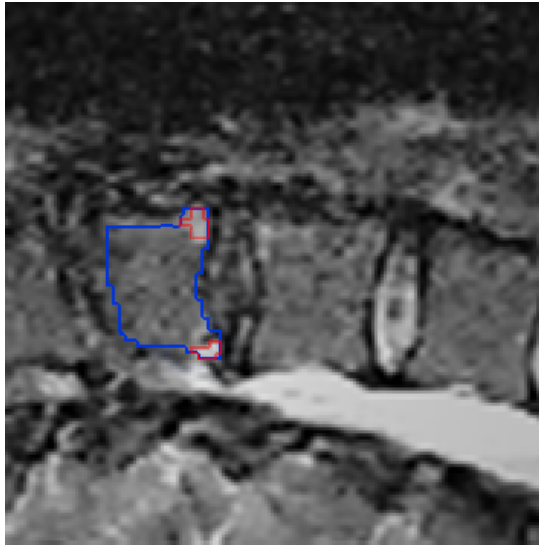
Figure 11. Pre-processed STIR MR image of a spinal vertebra of a psoriatic arthritis patient. Bone marrow oedema (delineated in red) is much smaller compared to total vertebral bone marrow (region of interest, delineated in blue).

The region-based approach is suitable for cases where the part of the image that is analysed contains one type of tissue relevant in the analysis. For example, in case of large oedematous lesions, the region-based approach can be used to classify slices into healthy or containing oedematous lesions, as seen in (124, 125). In paper III, the analysis aimed at classifying tumour lesions to seven-year overall survivors. The regions that were analysed contained areas with visible contrast enhancement, consistent with tumour and a morphological operation was applied to ensure that no enhanced vessels or spuriously enhanced voxels were included in the ROI (Figure 12). Therefore, the features that were calculated refer to the overall image characteristics of tumours. Using a pixel-wise approach in this case would have provided more, possibly excessive, information and reduce the computational efficiency.
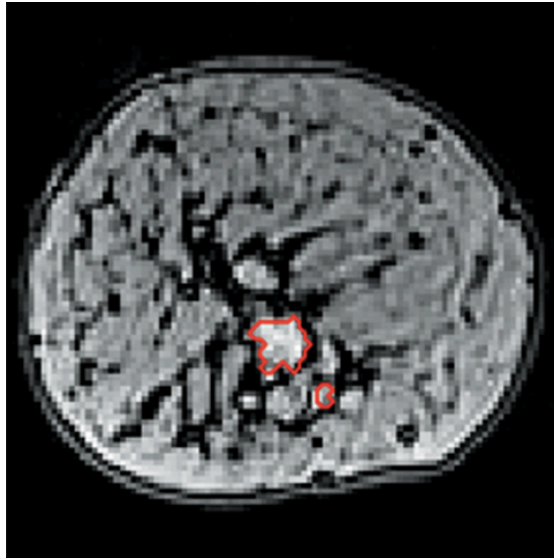
Figure 12. Two minutes post-contrast DCE MR image of locally advanced breast cancer. Tumour (region of interest) is outlined in red.

It has been reported that the size of the ROI may have an impact in the calculated feature values (126, 127). In paper II, the kernel size has an impact on the feature value for most of the GLCM and gradient features, but smaller impact on the intensity features. The intensity features are dependent on the local contrast and brightness. It is expected that the variation between the samples increases as the kernel size is reduced (126). In paper III, the largest tumour diameter was significantly higher in non-survivors, as larger tumour size is associated with a poorer prognosis in breast cancer. However, no significant correlation between textural features and tumour size was found in the cohort analysed in paper III.

5.3 Texture analysis in MR images: Relating textural features to physiology

Texture translates information from biomedical images that may not be perceptible by the human eye to quantitative biologically relevant measurements (128). Texture analysis of MR images has received increased attention during the past two decades (82, 129, 130). In both studies included in this thesis, GLCM textural features were used as an input to segmentation and/or classification algorithms.

In the pathologies of the bones, textural features are considered sensitive in detecting alteration of the bone architecture, segmenting and characterizing lesions and evaluating treatment

response (124, 125, 131-136). Texture features have been further used in assessing tissue heterogeneity and disease progression in osteoarthritis (124, 125, 131-134). Intensity (standard deviation, smoothness and third moment) (125), gradient (variance and mean of absolute gradient) and textural features (correlation and sum entropy) (124) extracted from segmented bone marrow could effectively classify between healthy slices and slices that were affected by bone marrow oedema of the knee. Lastly, MR image-derived textural features could evaluate the degree of post-radiation bone marrow oedema in patients with skeletal metastases (136).

In cancer, textural features appear useful in differentiating cancerous from non-cancerous tissue, staging of the disease, assessing response to treatment and prognosis (137-142), while relatively few studies assess prognosis through textural features from pre-treatment DCE MR images (141, 143-146). In study B, several GLCM textural features, extracted from segmented tumours in pre-treatment DCE 3T MR images, were associated with seven-year overall survival outcome. Five out of 16 GLCM textural features [f_1, f_2, f_5, f_{10} and f_{11}] were significantly different between seven-year overall survivors and non-survivors using LMM analysis. Most of the same features from 2-minutes post-contrast images resulted in significantly different survival curves in Kaplan-Meier survival analysis, while all of them were the most important variables at distinguishing between overall survivors and non-survivors by orthogonalized PLS-DA. Lastly, GLCM textural features added value to the clinical prognostic factors in the prediction of overall survival. Since studies have different acquisition parameters, pre-processing and follow-up time, direct quantitative comparison of results is challenging. The results shown in paper III were in agreement with most of the published literature (141, 143-146) and with previously published results of the same cohort regarding NAC treatment response (138). Patients with less heterogeneous tumours on contrast enhanced T1-weighted subtraction images were related to poorer outcomes (145). The opposite was true for T2-weighted images, where high heterogeneity was related with poorer outcomes (145). A previous study (141) that attempted to predict breast cancer survival based on parameters calculated by pre-treatment DCE MR images found GLCM textural features related to heterogeneity [f_7 and f_8] and symmetry [f_{15} and f_{16}] consistently higher in patients with shorter survival. These features were not significantly different between the patient groups in study B but were all affected by changes in the MR protocol.

GLCM textural features are derived from the grey-value spatial distribution in the GLCM. Although mathematically defined, their visual interpretation and their relationship to

physiology is not always intuitive. As several features describe similar image properties, several features are correlated with each other, as seen on the heat-map of paper III (Figure 13). Specifically, f_2, which measures the intensity variation between a reference pixel and its neighbour and relates to contrast, is positively correlated with f_{10} and f_{11}, which are associated with heterogeneity. f_1, a measure of local uniformity of intensities in an image, and f_5, which measures "how close the distribution of elements in the GLCM matrix is to the diagonal of GLCM" (147), are associated with homogeneity and are also positively correlated. In contrast, f_2 and f_5 are negatively correlated. f_7 and f_8, which relate to heterogeneity, are positively correlated with each other and negatively correlated with f_1. The use of statistical analysis to recognize the most important features to a specific research question could be useful before further analysis. Direct comparison of textural features calculated from different studies can be challenging. Therefore, assessing the textural features in more generic terms can be useful when comparing with existing literature.
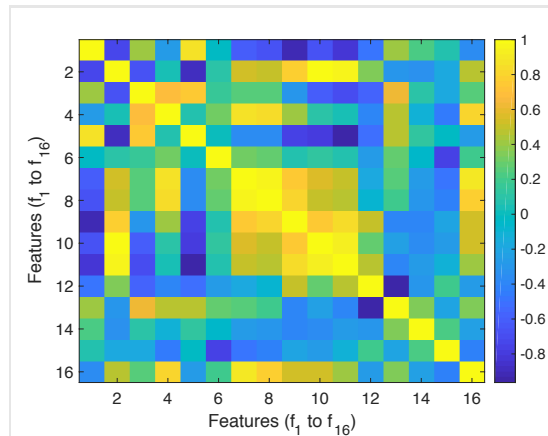
Figure 13. Heat-map based on Pearson correlations between the textural features illustrates correlations, ranging from strong negative (-1) to strong positive (+1), between GLCM features f_1 to f_{16} obtained from two minutes post-contrast images of LABC patients.

5.4 Quantifying bone marrow oedema in patients with psoriatic arthritis

In study A, quantitative MR image-based measures for bone marrow oedema in patients with psoriatic arthritis were extracted and compared to the reference standard MR image scoring, SPARCC. Two pixel-wise methods for quantification of bone marrow oedema in patients with psoriatic arthritis were implemented. Both methods resulted in the segmentation of the

oedematous voxels. Paper I evaluated thresholding as a method for quantification of bone marrow oedema in the spine and SI joints of psoriatic arthritis patients. The additional analysis of Paper II assessed the use of intensity, textural and gradient features as predictors for the segmentation of oedematous voxels in the spine of the same patients.

Lesion detectability by either method is affected by poor image quality and signal intensity inhomogeneities. Appropriate pre-processing, that includes bias-field correction and normalization is crucial. It should also be noted that both methods primarily aim to quantify the volume of oedematous lesions, while SPARCC primarily scores for the presence of oedema. Patients with the same SPARCC score do not necessarily experience the same inflammatory load in terms of volumes of oedematous lesions, which is why SPARCC scores and calculated volumes do not always agree. SPARCC and calculated volumes provide complementary information. The number of lesions was also assessed as a measure of disease burden in paper I. A combination of number and volume of oedematous lesions would provide similar information to SPARCC score. In addition, both methods perform best in patients with high inflammatory load. Many patients in this study had either low or no inflammation in their spine and/or SI joints, with the mean SPARCC score for the spine being 7.4, ranging from 0 to 51 out of a maximum of 108 and for the SI joints being 1.6, ranging from 0 to 17 out of a maximum possible score of 72.

Both SPARCC score and MR image-based calculated volumes provide additional information regarding psoriatic arthritis disease burden, compared to clinical evaluation scores (BASDAI, DAS 44, high sensitivity CRP, pain scores). A lack of correlation between clinical evaluation scores and MR image bone oedema scores has been reported for psoriatic arthritis patients, suggesting that these clinical measures may not accurately reflect the burden of inflammation in these patients (148). It can be that inflammation seen in MR images is subclinical and does not necessarily cause any clinical symptoms. This subclinical disease activity may have prognostic value, as it has been hypothesized for rheumatoid arthritis patients (149).

Thresholding is the simplest histogram-based method for classification and segmentation. The grey-level intensity value of each voxel determines the classification result. It performs best in images with bi-modal grey-level intensity histogram, where the objects and the background have different, but uniform brightness (150). Bone marrow oedematous lesions are hyper-intense compared to normal bone marrow in STIR images, resulting in bi-modal intensity

histogram in the ROI, the bone marrow (97). Hence, thresholding is an appropriate method for quantification of bone marrow in psoriatic arthritis patients and other types of spondyloarthritis, that is fast and easily implemented. On the other hand, more advanced methods allow for a set of predictors to determine the result of the classification. In paper II, the chosen predictors were local measures of grey-level intensity, as well as their spatial variations as described by textural and gradient features. This approach performs best in lesions that do not necessarily have the same overall brightness but exhibit similar spatial variations in grey-level intensities. Spatial information of the signal intensities in an oedematous lesion could be useful in the identification and quantification. In study B, thresholding based on relative enhancement was performed in the pre-processing of the images after manual lesion segmentation to exclude areas of necrosis, healthy or fatty tissue from tumour segmentation (138).

In the additional analysis of Paper II, the PLS-DA classifier achieved a pixel-wise sensitivity and specificity of 97% and 96%, respectively, compared to manual lesion segmentations. Intensity features i_1 i_2, i_3 and i_6 were the most important features. PLS-DA classification failed in several image sets. Overall, PLS-DA identifies the same lesions as manual lesion segmentation, but the area of the lesion identified by PLS-DA classification is generally smaller than the area by the manual lesion segmentation. Manual segmentations were performed before application of the diffusion filter, which can contribute to the discrepancy of lesion sizes. The diffusion filter is a shape adapting smoothing that reduces noise while preserving edges and structures in the images.

A common limitation of papers I and II is that bone marrow segmentations were drawn manually. There are multiple methods for automatic segmentation of the spine, but they were out of scope of these papers (151-153). However, with the addition of an algorithm that performs automatic bone segmentation for the spine and SI joints, the methods presented in these papers can provide fully automated quantification of the volume of oedematous lesions in the spine and SI joints of psoriatic arthritis patients, as well as patients suffering from other types of spondyloarthritis who experience inflammation in the same areas. Lastly, it is possible that the lack of external validation set resulted in overfitting in Paper I. This means that the model presented in paper I would behave more poorly in segmenting oedematic lesions in new images of psoriatic arthritis patients. There are many ways to prevent overfitting. An example

would be the feature selection that was performed using LOPO cross validation when developing the PLS-DA classifier for the additional analysis of Paper II.

5.5 Predicting survival with texture analysis of dynamic contrast-enhanced MR images of breast cancer

NAC is the standard initial treatment for inoperable LABC patients, before assessing operability (24). Response to NAC and long-term outcomes are variable. The treatment course following NAC is decided on an individual basis depending on patient's response to treatment, disease stage and lesion characteristics. Clinical factors, such as ER status and tumour grade, may allow identification of LABC patients who would benefit from low-intensity chemotherapy or even no chemotherapy (24). Predicting patient long-term survival outcome prior to NAC can affect treatment stratification offering more aggressive treatment to patients with lower expected survival and recognizing patients who need close post treatment follow-up. Even though clinical prognostic factors have shown promising results in assessing prognosis of breast cancer patients, MR imaging can provide additional biomarkers with potential for long-term survival prediction that provide added value to clinical prognostic factors (154).

Biomarkers are measurable indicators of biological processes. Prognostic biomarkers provide information about the disease outcome and affect the management of the disease by stratifying the patients (155). Clinical prognostic indicators for breast cancer include age at diagnosis, largest tumour dimension, TNM stage, ER, PgR and HER2 status and histological grade and type (141, 156, 157). Intrinsic breast cancer subtypes may have diverse biological mechanisms, resulting in different response to treatment and prognostic implications (158), with luminal-A tumours having favourable prognosis (159). Intrinsic subtypes have been shown to add significant predictive and prognostic value compared to other clinical prognostic indicators (160). Within the HER2+ patient subgroup, HER2-enriched patients responded better to trastuzumab-based therapy compared to non-HER2-enriched patients, indicating that using knowledge from intrinsic breast cancer subtyping can assist in personalizing breast cancer care (161).

In study B, the use of GLCM textural features from a single time-point of pre-treatment post-contrast DCE MR images as prognostic indicators of overall survival in breast cancer patients

was assessed. Textural features showed added value to clinical prognostic factors in predicting overall survival. The combination of clinical prognostic factors with GLCM features resulted in significantly higher classification accuracy for predicting overall survival. In a previous study, combination of clinical prognostic indicator with MR-derived metrics, not including textural features, but shape, resulted in significantly more accurate prediction of five-year survival of LABC patients (141). These results indicate that textural information from pre-treatment DCE MR images can potentially assist in increasing personalized treatment for breast cancer patients.

A limitation of study B is that the number of features (N=16) is comparable to the number of patients (N=55). As a rule of thumb, 10-15 patients are needed for each feature added in the model to reduce false discovery rate (107). PLS-DA is handles this problem, the curse of dimentionality, very well, requiring at least twice as many samples as features (162). Due to the retrospective nature of this study, it was impossible to include more patients. To this end, the reported p-values were corrected using Bonferroni multiple-comparison correction.

6 Conclusions and future perspectives

The work described in this thesis signifies the importance of furthering the knowledge of radiomics to obtain diagnostic and prognostic quantitative MR image-based measures.

Firstly, in the work presented in this thesis, we demonstrated the use of a framework that included a quality control procedure, image pre-processing steps and appropriate statistical analysis, for analysis of data acquired longitudinally, and with different scanners and protocols. Secondly, we evaluated the potential of STIR images to quantify the inflammatory burden in the spine and the SI joints of patients with psoriatic arthritis and to assess the effect of HIIT in psoriatic arthritis patients at baseline and after intervention. Lastly, we found an association between seven-year overall survival and textural features of pre-treatment breast cancer DCE-MR images and demonstrated the added value of textural features to the clinical prognostic factors in the prediction of overall survival.

Following the results of this research thesis, several aspects of this work are considered for further study. Reliable measures obtained from images must be technically and clinically validated, accurate and precise. The use of acquisition protocols and pre-processing steps optimized for the extraction of textural features should be explored, for both psoriatic arthritis and breast cancer imaging. Different methods for features extraction, such as 3D texture and run-length texture feature extraction, are worth looking into. Furthermore, considering that ROI size and segmentation method influence textural features, different approaches for ROI segmentation should be explored. Since textural features may depend on the acquisition parameters and can be affected by the acquisition settings, multi-centre studies are necessary to ensure the reproducibility of the results. Additionally, the usefulness of textural image features in segmenting bone marrow oedema, assessing the effect of intervention in psoriatic arthritis and predicting survival of LABC patients have to be validated in larger cohorts, as well as in patient subgroups, such as patients with low versus high disease burden. The correlation of textural features with disease activity should also be investigated in larger cohorts of patients with axial psoriatic arthitis. Lastly, as results of this research thesis and published literature indicated that texture relates to underlying physiology, investigating the relationship between measured obtained from radiomics analysis to physiology is another possible line of research. A better understanding of what physiological phenomena the textural parameters describe would contribute to more open and interpretable models when such

features are used for e.g. prognostic purposes. In the long term, this could contribute to easier translation of such methodology to clinical use.

References

1. Wright V, Moll J. Psoriatic arthritis, seronegative poliarthritis, p169-235, Amsterdam, North-Hollander, Pub. CO; 1976.

2. Madland TM, Apalset EM, Johannessen AE, Rossebö B, Brun JG. Prevalence, disease manifestations, and treatment of psoriatic arthritis in Western Norway. The Journal of Rheumatology. 2005;32(10):1918-22

3. Hoff M, Gulati AM, Romundstad PR, Kavanaugh A, Haugeberg G. Prevalence and incidence rates of psoriatic arthritis in central Norway: data from the Nord-Trøndelag Health Study (HUNT). Ann Rheum Dis. 2013

4. Gottlieb A, Korman NJ, Gordon KB, Feldman SR, Lebwohl M, Koo JYM, et al. Guidelines of care for the management of psoriasis and psoriatic arthritis: Section 2. Psoriatic arthritis: Overview and guidelines of care for treatment with an emphasis on the biologics. J Am Acad Dermatol. 2008;58(5):851-64 DOI: https://doi.org/10.1016/j.jaad.2008.02.040.

5. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H. Classification criteria for psoriatic arthritis: Development of new criteria from a large international study. Arthritis Rheum. 2006;54(8):2665-73 DOI: 10.1002/art.21972.

6. Haroon M, Gallagher P, FitzGerald O. Diagnostic delay of more than 6 months contributes to poor radiographic and functional outcome in psoriatic arthritis. Ann Rheum Dis. 2014

7. El Miedany Y, El Gaafary M, Youssef S, Ahmed I, Nasr A. Tailored approach to early psoriatic arthritis patients: clinical and ultrasonographic predictors for structural joint damage. Clin Rheumatol. 2015;34(2):307-13 DOI: 10.1007/s10067-014-2630-2.

8. Gladman D. Mortality in psoriatic arthritis. Clin Exp Rheumatol. 2008;26(5):62

9. Thomsen RS, Nilsen TIL, Haugeberg G, Bye A, Kavanaugh A, Hoff M. Effect of high-intensity interval training on cardiovascular disease risk factors and body composition in psoriatic arthritis: a randomised controlled trial. RMD Open. 2018;4(2) DOI: 10.1136/rmdopen-2018-000729.

10. Thomsen RS, Nilsen TIL, Haugeberg G, Bye A, Kavanaugh A, Hoff M. The impact of high intensity interval training on disease activity and patient disease perception in patients with psoriatic arthritis: a randomized controlled trial. 2018

11. Karlsen T, Aamot I-L, Haykowsky M, Rognmo Ø. High Intensity Interval Training for Maximizing Health Outcomes. Prog Cardiovasc Dis. 2017;60(1):67-77 DOI: https://doi.org/10.1016/j.pcad.2017.03.006.

12. Baraliakos X, Coates LC, Braun J. The involvement of the spine in psoriatic arthritis. Clin Exp Rheumatol. 2015;33(5 Suppl 93):S31-5

13. Gladman DD. Axial disease in psoriatic arthritis. Curr Rheumatol Rep. 2007;9(6):455-60 DOI: 10.1007/s11926-007-0074-2.

14. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. The Journal of rheumatology. 1994;21(12):2286-91

15. Smolen JS, Schöls M, Braun J, Dougados M, FitzGerald O, Gladman DD, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis. 2017

16. Cancer Registry of Norway. Cancer in Norway 2019 - Cancer incidence, mortality, survival and prevalence in Norway. Oslo, Norway: Cancer Registry of Norway; 2020.

17. Cowell CF, Weigelt B, Sakr RA, Ng CKY, Hicks J, King TA, et al. Progression from ductal carcinoma in situ to invasive breast cancer: Revisited. Mol Oncol. 2013;7(5):859-69 DOI: 10.1016/j.molonc.2013.07.005.

18. Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. JAMA. 2003;289(11):1421-4 DOI: 10.1001/jama.289.11.1421.

19. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74 DOI: 10.1073/pnas.191367098.

20. Cuccurullo V, Mansi L. AJCC Cancer Staging Handbook: from the AJCC Cancer Staging Manual (7th edition). Eur J Nucl Med Mol Imaging. 2011;38(2):408 DOI: 10.1007/s00259-010-1693-9.

21. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res. 2010;12(4):207 DOI: 10.1186/bcr2607.

22. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, et al. Breast Cancer—Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67(4):290-303 DOI: 10.3322/caac.21393.

23. Fleeman N, Martin Saborido C, Payne K, Boland A, Dickson R, Dundar Y, et al. The clinical effectiveness and costeffectiveness of genotyping for CYP2D6 for the management of women with breast cancer treated with tamoxifen: A systematic review. Health technology assessment (Winchester, England). 2011;15:1-102 DOI: 10.3310/hta15330.

24. Helsedirectoratet. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft. IS-2669 ed. Oslo, Norway. 2017

25. Garg PK, Prakash G. Current definition of locally advanced breast cancer. Current Oncology. 2015;22(5):409-10 DOI: 10.3747/co.22.2697.

26. Cardoso F, Costa A, Senkus E, Aapro M, Andre F, Barrios CH, et al. 3rd ESO-ESMO international consensus guidelines for Advanced Breast Cancer (ABC 3). Breast (Edinburgh, Scotland). 2017;31:244-59 DOI: 10.1016/j.breast.2016.10.001.

27. Brown RW, Cheng Y-CN, Haacke EM, Thompson MR, Venkatesan R. Classical Response of a Single Nucleus to a Magnetic Field. Magnetic Resonance Imaging: John Wiley & Sons, Inc.; 2014. p. 19-36.

28. Ansorge R, Graves M. The Physics and Mathematics of MRI: Morgan & Claypool Publishers; 2016. Available from: http://dx.doi.org/10.1088/978-1-6817-4068-3.

29. Brown RW, Cheng Y-CN, Haacke EM, Thompson MR, Venkatesan R. The Quantum Mechanical Basis of Precession and Excitation. Magnetic Resonance Imaging: John Wiley & Sons, Inc.; 2014. p. 67-83.

30. Bloch F. Nuclear Induction. Physical Review. 1946;70(7-8):460-74

31. Rajan SS. Magnetic Field Gradient Pulses and Spatial Encoding of MR Signal. MRI: A Conceptual Overview. New York, NY: Springer New York; 1998. p. 12-24.

32. McRobbie DW, Moore EA, Graves MJ. MRI from Picture to Proton: Cambridge university press; 2017.

33. Gallagher TA, Nemeth AJ, Hacein-Bey L. An introduction to the Fourier Transform: relationship to MRI. American Journal of Roentgenology. 2008;190(5):1396-405

34. Bernstein MA, King KF, Zhou XJ. Chapter 14 - Basic pulse sequences. Handbook of MRI Pulse Sequences. Burlington: Academic Press; 2004. p. 579-647.

35. Krinsky G, Rofsky NM, Weinreb JC. Nonspecificity of short inversion time inversion recovery (STIR) as a technique of fat suppression: pitfalls in image interpretation. AJR Am J Roentgenol. 1996;166(3):523-6 DOI: 10.2214/ajr.166.3.8623620.

36. Delfaut EM, Beltran J, Johnson G, Rousseau J, Marchandise X, Cotten A. Fat Suppression in MR Imaging: Techniques and Pitfalls. RadioGraphics. 1999;19(2):373-82 DOI: 10.1148/radiographics.19.2.g99mr03373.

37. Strijkers GJ, M Mulder WJ, F van Tilborg GA, Nicolay K. MRI contrast agents: current status and future perspectives. Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents). 2007;7(3):291-305

38. Zhou Z, Lu Z-R. Gadolinium-Based Contrast Agents for MR Cancer Imaging. Wiley interdisciplinary reviews Nanomedicine and nanobiotechnology. 2013;5(1):1-18 DOI: 10.1002/wnan.1198.

39. Gordon Y, Partovi S, Müller-Eschner M, Amarteifio E, Bäuerle T, Weber M-A, et al. Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. Cardiovascular Diagnosis and Therapy. 2014;4(2):147-64 DOI: 10.3978/j.issn.2223-3652.2014.03.01.

40. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusable tracer: Standardized quantities and symbols. Journal of Magnetic Resonance Imaging. 1999;10(3):223-32 DOI: 10.1002/(SICI)1522-2586(199909)10:3<223::AID-JMRI2>3.0.CO;2-S.

41. Baraliakos X, Conaghan PG, D'Agostino M-A, Maksymowych W, Naredo E, Ostergaard M, et al. Imaging in rheumatoid arthritis, psoriatic arthritis, axial spondyloarthritis, and osteoarthritis: An international viewpoint on the current knowledge and future research priorities. Eur J Rheumatol. 2019;6(1):38-47 DOI: 10.5152/eurjrheum.2018.18121.

42. McQueen F, Lassere M, Ostergaard M. Magnetic resonance imaging in psoriatic arthritis: a review of the literature. Arthritis Research and Therapy. 2006;8(2):207

43. Mandl P, Aletaha D. The role of ultrasound and magnetic resonance imaging for treat to target in rheumatoid arthritis and psoriatic arthritis. Rheumatology. 2019;58(12):2091-8 DOI: 10.1093/rheumatology/kez397.

44. Biagioni BJ, Gladman DD, Cook RJ, Eder L, Wakhlu A, Shen H, et al. Reliability of Radiographic Scoring Methods in Axial Psoriatic Arthritis. Arthritis Care Res (Hoboken). 2014;66(9):1417-22 DOI: 10.1002/acr.22308.

45. Ibrahim A, Gladman DD, Thavaneswaran A, Eder L, Helliwell P, Cook RJ, et al. Sensitivity and Specificity of Radiographic Scoring Instruments for Detecting Change in Axial Psoriatic Arthritis. Arthritis Care Res (Hoboken). 2017;69(11):1700-5 DOI: 10.1002/acr.23189.

46. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Krishnananthan R, Stone M, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Care Res. 2005;53(4):502-9

47. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Williams M, Stone M, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Care Res. 2005;53(5):703-9

48. LandewÉ RB, Hermann K-GA, van der Heijde DM, Baraliakos X, Jurik A-G, Lambert RG, et al. Scoring sacroiliac joints by magnetic resonance imaging. A multiple-reader reliability experiment. J Rheumatol. 2005;32(10):2050-5

49. Mann RM, Kuhl CK, Kinkel K, Boetes C. Breast MRI: guidelines from the European Society of Breast Imaging. European Radiology. 2008;18(7):1307-18 DOI: 10.1007/s00330-008-0863-7.

50. Wilmes LJ, Pallavicini MG, Fleming LM, Gibbs J, Wang D, Li K-L, et al. AG-013736, a novel inhibitor of VEGF receptor tyrosine kinases, inhibits breast cancer growth and decreases vascular permeability as detected by dynamic contrast-enhanced magnetic resonance imaging. Magnetic Resonance Imaging. 2007;25(3):319-27 DOI: https://doi.org/10.1016/j.mri.2006.09.041.

51. Li L, Wang K, Sun X, Wang K, Sun Y, Zhang G, et al. Parameters of dynamic contrast-enhanced MRI as imaging markers for angiogenesis and proliferation in human breast cancer. Med Sci Monit. 2015;21:376-82 DOI: 10.12659/MSM.892534.

52. Johansen R, Jensen LR, Rydland J, Goa PE, Kvistad KA, Bathen TF, et al. Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI. Journal of Magnetic Resonance Imaging. 2009;29(6):1300-7 DOI: 10.1002/jmri.21778.

53. Tuncbilek N, Tokatli F, Altaner S, Sezer A, Türe M, Omurlu IK, et al. Prognostic value DCE-MRI parameters in predicting factor disease free survival and overall survival for breast cancer patients. European Journal of Radiology. 2012;81(5):863-7 DOI: https://doi.org/10.1016/j.ejrad.2011.02.021.

54. El Khouli RH, Macura KJ, Jacobs MA, Khalil TH, Kamel IR, Dwyer A, et al. Dynamic Contrast-Enhanced MRI of the Breast: Quantitative Method for Kinetic Curve Type Assessment. American Journal of Roentgenology. 2009;193(4):W295-W300 DOI: 10.2214/AJR.09.2483.

55. Kul S, Cansu A, Alhan E, Dinc H, Gunes G, Reis A. Contribution of Diffusion-Weighted Imaging to Dynamic Contrast-Enhanced MRI in the Characterization of Breast Tumors. American Journal of Roentgenology. 2011;196(1):210-7 DOI: 10.2214/AJR.10.4258.

56. Kuhl CK, Mielcareck P, Klaschik S, Leutner C, Wardelmann E, Gieseke J, et al. Dynamic Breast MR Imaging: Are Signal Intensity Time Course Data Useful for Differential Diagnosis of Enhancing Lesions? Radiology. 1999;211(1):101-10 DOI: 10.1148/radiology.211.1.r99ap38101.

57. Karahaliou A, Vassiou K, Arikidis NS, Skiadopoulos S, Kanavou T, Costaridou L. Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. The British Journal of Radiology. 2010;83(988):296-309 DOI: 10.1259/bjr/50743919.

58. Chen JH, Feig B, Agrawal G, Yu H, Carpenter PM, Mehta RS, et al. MRI evaluation of pathologically complete response and residual tumors in breast cancer after neoadjuvant chemotherapy. Cancer. 2008;112(1):17-26 DOI: 10.1002/cncr.23130.

59. Loo CE, Teertstra HJ, Rodenhuis S, van de Vijver MJ, Hannemann J, Muller SH, et al. Dynamic Contrast-Enhanced MRI for Prediction of Breast Cancer Response to Neoadjuvant Chemotherapy: Initial Results. American Journal of Roentgenology. 2008;191(5):1331-8 DOI: 10.2214/AJR.07.3567.

60. Wang T, Wu X, Cui Y, Chu C, Ren G, Li W. Role of apparent diffusion coefficients with diffusion-weighted magnetic resonance imaging in differentiating between benign and malignant bone tumors. World J Surg Oncol. 2014;12:365 DOI: 10.1186/1477-7819-12-365.

61. Ah-See M-LW, Makris A, Taylor NJ, Harrison M, Richman PI, Burcombe RJ, et al. Early Changes in Functional Dynamic Magnetic Resonance Imaging Predict for Pathologic Response to Neoadjuvant Chemotherapy in Primary Breast Cancer. Clinical Cancer Research. 2008;14(20):6580-9 DOI: 10.1158/1078-0432.ccr-07-4310.

62. Craciunescu OI, Blackwell KL, Jones EL, Macfall JR, Yu D, Vujaskovic Z, et al. DCE-MRI parameters have potential to predict response of locally advanced breast cancer patients to neoadjuvant chemotherapy and hyperthermia: A pilot study. International Journal of Hyperthermia. 2009;25(6):405-15 DOI: 10.1080/02656730903022700.

63. Abramson RG, Li X, Hoyt TL, Su P-F, Arlinghaus LR, Wilson KJ, et al. Early assessment of breast cancer response to neoadjuvant chemotherapy by semi-quantitative analysis of high-temporal resolution DCE-MRI: Preliminary results. Magnetic Resonance Imaging. 2013;31(9):1457-64 DOI: https://doi.org/10.1016/j.mri.2013.07.002.

64. Atuegwu NC, Arlinghaus LR, Li X, Chakravarthy AB, Abramson VG, Sanders ME, et al. Parameterizing the Logistic Model of Tumor Growth by DW-MRI and DCE-MRI Data to Predict Treatment Response and Changes in Breast Cancer Cellularity during Neoadjuvant

Chemotherapy. Translational Oncology. 2013;6(3):256-64 DOI: https://doi.org/10.1593/tlo.13130.

65. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-62 DOI: 10.1038/nrclinonc.2017.141.

66. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015;277(3):813-25 DOI: 10.1148/radiol.2015142202.

67. Sing JK, Adhikari SK, Kahali S, editors. On estimation of bias field in MRI images. 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS); 2015 2-3 Nov. 2015.

68. Song S, Zheng Y, He Y. A review of Methods for Bias Correction in Medical Images. Biomedical Engineering Review. 2017;1(1) DOI: 10.18103/bme.v3i1.1550.

69. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 Bias Correction. IEEE Trans Med Imaging. 2010;29(6):1310-20 DOI: 10.1109/TMI.2010.2046908.

70. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998;17(1):87-97 DOI: 10.1109/42.668698.

71. Kaur M, Kaur J, Kaur J. Survey of contrast enhancement techniques based on histogram equalization. International Journal of Advanced Computer Science and Applications. 2011;2(7):137-41

72. Morovic J, Shaw J, Sun P-L. A fast, non-iterative and exact histogram matching algorithm. Pattern Recognition Letters. 2002;23(1):127-35

73. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1990;12(7):629-39 DOI: 10.1109/34.56205.

74. Justusson BI. Median Filtering: Statistical Properties. Two-Dimensional Digital Signal Prcessing II: Transforms and Median Filters. Berlin, Heidelberg: Springer Berlin Heidelberg; 1981. p. 161-96.

75. Rogowska J. Chapter 5 - Overview and Fundamentals of Medical Image Segmentation A2 - BANKMAN, ISAAC N. Handbook of Medical Image Processing and Analysis (Second Edition). Burlington: Academic Press; 2009. p. 73-90.

76. Sahoo PK, Soltani S, Wong AKC. A survey of thresholding techniques. Computer Vision, Graphics, and Image Processing. 1988;41(2):233-60 DOI: https://doi.org/10.1016/0734-189X(88)90022-9.

77. Otsu N. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics. 1979;9(1):62-6

78. Glasbey CA. An analysis of histogram-based thresholding algorithms. CVGIP: Graphical models and image processing. 1993;55(6):532-7

79. Haralick RM, Sternberg SR, Zhuang X. Image Analysis Using Mathematical Morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1987;PAMI-9(4):532-50 DOI: 10.1109/TPAMI.1987.4767941.

80. Dalal N, Triggs B, editors. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005 25-25 June 2005.

81. Shapiro LG, Stockman GC. Computer vision. Upper Saddle River, NJ: Prentice Hall; 2001. 580 p.

82. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. Clin Radiol. 2004;59(12):1061-9 DOI: https://doi.org/10.1016/j.crad.2004.07.008.

83. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. Ieee Transactions on Systems Man and Cybernetics. 1973;Smc3(6):610-21 DOI: 10.1109/Tsmc.1973.4309314.

84. Conners RW, Harlow CA. A Theoretical Comparison of Texture Algorithms. Ieee Transactions on Pattern Analysis and Machine Intelligence. 1980;2(3):204-22

85. Thomas T, Vijayaraghavan AP, Emmanuel S. Introduction to Machine Learning. Machine Learning Approaches in Cyber Security Analytics: Springer; 2020. p. 17-36.

86. Alpaydin E. Introduction to machine learning: MIT press; 2014.

87. Yu FR, He Y. Introduction to Machine Learning. Deep Reinforcement Learning for Wireless Networks. Cham: Springer International Publishing; 2019. p. 1-13.

88. Christopher MB. Pattern recognition and machine learning: Springer-Verlag New York; 2016.

89. Jollife I. Principal component analysis. 1986. Springer, New York.

90. Sawant P, Billor N, Shin H. Functional outlier detection with robust functional principal component analysis. Computational Statistics. 2012;27(1):83-102 DOI: 10.1007/s00180-011-0239-3.

91. Brereton RG, Lloyd GR. Partial least squares discriminant analysis: taking the magic away. Journal of Chemometrics. 2014;28(4):213-25 DOI: 10.1002/cem.2609.

92. Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, et al. A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding. Anal Chim Acta. 2015;879:10-23 DOI: 10.1016/j.aca.2015.02.012.

93. Gałecki A, Burzykowski T. Linear Mixed-Effects Model. Linear Mixed-Effects Models Using R: A Step-by-Step Approach. New York, NY: Springer New York; 2013. p. 245-73.

94. Chen C-C, Wan Y-L, Wai Y-Y, Liu H-L. Quality Assurance of Clinical MRI Scanners Using ACR MRI Phantom: Preliminary Results. J Digit Imaging. 2004;17(4):279-84 DOI: 10.1007/s10278-004-1023-5.

95. Phantom test guidance for the ACR MRI accreditation program. American College of Radiology. Available from: https://www.acraccreditation.org/-/media/ACRAccreditation/Documents/MRI/LargePhantomGuidance.pdf.

96. Perona P, Malik J. Scale-Space and Edge-Detection Using Anisotropic Diffusion. Ieee Transactions on Pattern Analysis and Machine Intelligence. 1990;12(7):629-39 DOI: Doi 10.1109/34.56205.

97. Fritz J, Henes JC, Thomas C, Clasen S, Fenchel M, Claussen CD, et al. Diagnostic and interventional MRI of the sacroiliac joints using a 1.5-T open-bore magnet: a one-stop-shopping approach. AJR Am J Roentgenol. 2008;191(6):1717-24 DOI: 10.2214/ajr.08.1075.

98. Pinheiro J, Bates D. Linear Mixed-Effects Models: Basic Concepts and Examples. Mixed-Effects Models in S and S-PLUS. New York, NY, USA. New York: Springer; 2000.

99. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117. 2014.

100. Gareth ED, Nisha K, Yit L, Soujanye G, Emma H, Massat NJ, et al. MRI breast screening in high-risk women: cancer detection and survival analysis. Breast Cancer Res Treat. 2014;145(3):663-72 DOI: 10.1007/s10549-014-2931-9.

101. Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol. 2015;26(suppl_5):v8-v30 DOI: 10.1093/annonc/mdv298.

102. Sudoł-Szopińska I, Jurik AG, Eshed I, Lennart J, Grainger A, Østergaard M, et al. Recommendations of the ESSR Arthritis Subcommittee for the Use of Magnetic Resonance

Imaging in Musculoskeletal Rheumatic Diseases. Semin Musculoskelet Radiol. 2015;19(04):396-411 DOI: 10.1055/s-0035-1564696.

103. Tofts PS. Concepts: Measurement and MR. Quantitative MRI of the Brain: John Wiley & Sons, Ltd; 2004. p. 1-15.

104. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-77 DOI: 10.1148/radiol.2015151169.

105. Sanduleanu S, Woodruff HC, de Jong EEC, van Timmeren JE, Jochems A, Dubois L, et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. Radiother Oncol. 2018;127(3):349-60 DOI: 10.1016/j.radonc.2018.03.033.

106. Larue RTHM, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. The British Journal of Radiology. 2016;90(1070) DOI: 10.1259/bjr.20160665.

107. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol. 2016;61(13):R150-R66 DOI: 10.1088/0031-9155/61/13/R150.

108. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. Int J Radiat Oncol Biol Phys. 2018;102(4):1143-58 DOI: https://doi.org/10.1016/j.ijrobp.2018.05.053.

109. Fiset S, Welch ML, Weiss J, Pintilie M, Conway JL, Milosevic M, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. Radiother Oncol. 2019;135:107-14

110. Baeßler B, Weiss K, dos Santos DP. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. Invest Radiol. 2019;54(4):221-8

111. Moradmand H, Aghamir SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. J Appl Clin Med Phys. 2019

112. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5(1):13087 DOI: 10.1038/srep13087.

113. Yankeelov TE, Pickens DR, Price RR. Quantitative MRI in cancer: Taylor & Francis; 2011.

114. Zhou X, Sakaie KE, Debbins JP, Kirsch JE, Tatsuoka C, Fox RJ, et al. Quantitative quality assurance in a multicenter HARDI clinical trial at 3T. Magn Reson Imaging. 2017;35:81-90 DOI: 10.1016/j.mri.2016.08.022.

115. Kruggel F, Turner J, Muftuler LT, The Alzheimer's Disease Neuroimaging I. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage. 2010;49(3):2123-33 DOI: 10.1016/j.neuroimage.2009.11.006.

116. Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage. 2009;46(1):177-92 DOI: 10.1016/j.neuroimage.2009.02.010.

117. Chen C-C, Wan Y-L, Wai Y-Y, Liu H-L. Quality Assurance of Clinical MRI Scanners Using ACR MRI Phantom: Preliminary Results. J Digit Imaging. 2004;17(4):279-84 DOI: 10.1007/s10278-004-1023-5.

118. Davids M, Zöllner FG, Ruttorf M, Nees F, Flor H, Schumann G, et al. Fully-automated quality assurance in multi-center studies using MRI phantom measurements. Magn Reson Imaging.32(6):771-80 DOI: 10.1016/j.mri.2014.01.017.

119. Belaroussi B, Milles J, Carme S, Zhu YM, Benoit-Cattin H. Intensity non-uniformity correction in MRI: Existing methods and their validation. Med Image Anal.10(2):234-46 DOI: 10.1016/j.media.2005.09.004.

120. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. Br J Radiol. 2017;90(1070) DOI: 10.1259/bjr.20160642.

121. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging. 2004;22(1):81-91 DOI: https://doi.org/10.1016/j.mri.2003.09.001.

122. Chua AS, Egorova S, Anderson MC, Polgar-Turcsanyi M, Chitnis T, Weiner HL, et al. Handling changes in MRI acquisition parameters in modeling whole brain lesion volume and atrophy data in multiple sclerosis subjects: Comparison of linear mixed-effect models. NeuroImage : Clinical. 2015;8:606-10 DOI: 10.1016/j.nicl.2015.06.009.

123. Puissant A, Hirsch J, Weber C. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. International Journal of Remote Sensing. 2005;26(4):733-45 DOI: 10.1080/01431160512331316838.

124. Chuah TK, Van Reeth E, Sheah K, Poh CL. Texture analysis of bone marrow in knee MRI for classification of subjects with bone marrow lesion - data from the Osteoarthritis Initiative. Magn Reson Imaging. 2013;31(6):930-8 DOI: 10.1016/j.mri.2013.01.014.

125. Chuah TK, Poh CL, Sheah K. Quantitative texture analysis of MRI images for detection of cartilage-related bone marrow edema. Conf Proc IEEE Eng Med Biol Soc. 2011;2011:5112-5 DOI: 10.1109/IEMBS.2011.6091266.

126. Sikio M, Holli KK, Harrison LCV, Ruottinen H, Rossi M, Helminen MT, et al. Parkinson's Disease: Interhemispheric Textural Differences in MR Images. Acad Radiol. 2011;18(10):1217-24 DOI: 10.1016/j.acra.2011.06.007.

127. Robins M, Solomon J, Hoye J, Abadi E, Marin D, Samei E, editors. How reliable are texture measurements? SPIE Medical Imaging; 2018: SPIE.

128. Depeursinge A, Al-Kadi OS, Mitchell JR. Biomedical Texture Analysis: Fundamentals, Tools and Challenges: Academic Press; 2017.

129. Lerski RA, Straughan K, Schad LR, Boyce D, Blüml S, Zuna I. VIII. MR image texture analysis—An approach to tissue characterization. Magn Reson Imaging. 1993;11(6):873-87 DOI: https://doi.org/10.1016/0730-725X(93)90205-R.

130. Kassner A, Thornhill RE. Texture Analysis: A Review of Neurologic MR Imaging Applications. American Journal of Neuroradiology. 2010;31(5):809-16 DOI: 10.3174/ajnr.A2061.

131. Joseph GB, Baum T, Alizai H, Carballido-Gamio J, Nardo L, Virayavanich W, et al. Baseline mean and heterogeneity of MR cartilage T2 are associated with morphologic degeneration of cartilage, meniscus, and bone marrow over 3 years--data from the Osteoarthritis Initiative. Osteoarthritis Cartilage. 2012;20(7):727-35 DOI: 10.1016/j.joca.2012.04.003.

132. Blumenkrantz G, Stahl R, Carballido-Gamio J, Zhao S, Lu Y, Munoz T, et al. The feasibility of characterizing the spatial distribution of cartilage T-2 using texture analysis. Osteoarthritis Cartilage. 2008;16(5):584-90 DOI: 10.1016/j.joca.2007.10.019.

133. MacKay JW, Murray PJ, Kasmai B, Johnson G, Donell ST, Toms AP. MRI texture analysis of subchondral bone at the tibial plateau. Eur Radiol. 2016;26(9):3034-45 DOI: 10.1007/s00330-015-4142-0.

134. MacKay JW, Murray PJ, Low SBL, Kasmai B, Johnson G, Donell ST, et al. Quantitative analysis of tibial subchondral bone: Texture analysis outperforms conventional trabecular microarchitecture analysis. J Magn Reson Imaging. 2016;43(5):1159-70 DOI: doi:10.1002/jmri.25088.

135. Fripp J, Bourgeat P, Crozier S, Ourselin S. Segmentation of the bones in MRIs of the knee using phase, magnitude, and shape information. Acad Radiol. 2007;14(10):1201-8 DOI: 10.1016/j.acra.2007.06.021.

136. Romanos O, Solomou E, Georgiadis P, Kardamakis D, Siablis D. Magnetic resonance imaging and image analysis of post - radiation changes of bone marrow in patients with skeletal metastases. J BUON. 2013;18(3):788-94

137. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol. 2015;25(10):2840-50 DOI: 10.1007/s00330-015-3701-8.

138. Teruel JR, Heldahl MG, Goa PE, Pickles M, Lundgren S, Bathen TF, et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. NMR Biomed. 2014;27(8):887-96 DOI: 10.1002/nbm.3132.

139. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. Eur J Radiol. 2016;85(4):824-9 DOI: 10.1016/j.ejrad.2016.01.013.

140. Fox MJ, Gibbs P, Pickles MD. Minkowski functionals: An MRI texture analysis tool for determination of the aggressiveness of breast cancer. J Magn Reson Imaging. 2016;43(4):903-10 DOI: 10.1002/jmri.25057.

141. Pickles MD, Lowry M, Gibbs P. Pretreatment Prognostic Value of Dynamic Contrast-Enhanced Magnetic Resonance Imaging Vascular, Texture, Shape, and Size Parameters Compared With Traditional Survival Indicators Obtained From Locally Advanced Breast Cancer Patients. Invest Radiol. 2016;51(3):177-85 DOI: 10.1097/Rli.0000000000000222.

142. Lisson CS, Lisson CG, Flosdorf K, Mayer-Steinacker R, Schultheiss M, von Baer A, et al. Diagnostic value of MRI-based 3D texture analysis for tissue characterisation and discrimination of low-grade chondrosarcoma from enchondroma: a pilot study. Eur Radiol. 2017 DOI: 10.1007/s00330-017-5014-6.

143. Tuncbilek N, Tokatli F, Altaner S, Sezer A, Türe M, Omurlu IK, et al. Prognostic value DCE-MRI parameters in predicting factor disease free survival and overall survival for breast cancer patients. Eur J Radiol. 2009;81(5):863-7 DOI: 10.1016/j.ejrad.2011.02.021.

144. Heldahl MG, Bathen TF, Rydland J, Kvistad KA, Lundgren S, Gribbestad IS, et al. Prognostic value of pretreatment dynamic contrast-enhanced MR imaging in breast cancer

patients receiving neoadjuvant chemotherapy: Overall survival predicted from combined time course and volume analysis. Acta Radiol. 2010;51(6):604-12 DOI: 10.3109/02841851003782059.

145. Kim JH, Ko ES, Lim Y, Lee KS, Han BK, Ko EY, et al. Breast Cancer Heterogeneity: MR Imaging Texture Analysis and Survival Outcomes. Radiology. 2017;282(3):665-75 DOI: 10.1148/radiol.2016160261.

146. Dogan B, Yuan Q, Bassett R, Guvenc I, Jackson EF, Cristofanilli M, et al. Comparing the performances of Magnetic Resonance Imaging size versus pharmacokinetic parameters to predict response to neoadjuvant chemotherapy and survival in breast cancer patients. Curr Probl Diagn Radiol. 2018 DOI: https://doi.org/10.1067/j.cpradiol.2018.03.003.

147. Zayed N, Elnemr HA. Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities. International Journal of Biomedical Imaging. 2015;2015:7 DOI: 10.1155/2015/267807.

148. Tan YM, Østergaard M, Doyle A, Dalbeth N, Lobo M, Reeves Q, et al. MRI bone oedema scores are higher in the arthritis mutilans form of psoriatic arthritis and correlate with high radiographic scores for joint damage. Arthritis Res Ther. 2009;11(1) DOI: 10.1186/ar2586.

149. Krabben A, Stomp W, van Nies JAB, Huizinga TWJ, van der Heijde D, Bloem JL, et al. MRI-detected subclinical joint inflammation is associated with radiographic progression. Ann Rheum Dis. 2014;73(11):2034

150. Sharma N, Aggarwal LM. Automated medical image segmentation techniques. Journal of Medical Physics / Association of Medical Physicists of India. 2010;35(1):3-14 DOI: 10.4103/0971-6203.58777.

151. Chu C, Belavý DL, Armbrecht G, Bansmann M, Felsenberg D, Zheng G. Fully automatic localization and segmentation of 3d vertebral bodies from ct/mr images via a learning-based method. PLoS One. 2015;10(11)

152. Korez R, Likar B, Pernuš F, Vrtovec T, editors. Model-Based Segmentation of Vertebral Bodies from MR Images with 3D CNNs2016; Cham: Springer International Publishing.

153. Suzani A, Rasoulian A, Seitel A, Fels S, Rohling RN, Abolmaesumi P, editors. Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling; 2015: International Society for Optics and Photonics.

154. O'Connor JPB, Jackson A, Parker GJM, Jayson GC. DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. Br J Cancer. 2007;96:189 DOI: 10.1038/sj.bjc.6603515.

155. Bay B-H, Yip GW-C. Breast Cancer Prognostic Biomarkers. In: Schwab M, editor. Encyclopedia of Cancer. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 533-7.

156. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1989;63(1):181-7 DOI: doi:10.1002/1097-0142(19890101)63:1<181::AID-CNCR2820630129>3.0.CO;2-H.

157. Connolly R, Stearns V. Current approaches for neoadjuvant chemotherapy in breast cancer. Eur J Pharmacol. 2013;717(0):58-66 DOI: 10.1016/j.ejphar.2013.02.057.

158. Wilson TR, Yu J, Lu X, Spoerke JM, Xiao Y, O'Brien C, et al. The molecular landscape of high-risk early breast cancer: comprehensive biomarker analysis of a phase III adjuvant population. Npj Breast Cancer. 2016;2:16022 DOI: 10.1038/npjbcancer.2016.22 https://www.nature.com/articles/npjbcancer201622#supplementary-information.

159. Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol. 2010;28(10):1684-91

160. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol. 2009;27(8):1160-7 DOI: 10.1200/JCO.2008.18.1370.

161. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast (Edinburgh, Scotland). 2015;24 Suppl 2:S26-35 DOI: 10.1016/j.breast.2015.07.008.

162. Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? BMC Bioinformatics. 2020;21(1):1-10

Paper I

BMC Musculoskeletal Disorders

Open Access

Quantifying bone marrow inflammatory edema in the spine and sacroiliac joints with thresholding

Ioanna Chronaiou[1*], Ruth S. Thomsen[2], Else M. Huuse[3], Leslie R. Euceda[1], Susanne J. Pedersen[4], Mari Hoff[2,5] and Beathe Sitter[1]

Abstract

Background: Psoriatic Arthritis (PsA) is a chronic inflammatory arthritis that develops in patients with psoriasis. Inflammatory edema in the spine may reflect subclinical disease activity and be a predictor of radiographic progression. A semi-quantitative method established by the spondyloarthritis research consortium of Canada (SPARCC) is commonly used to assess the disease activity in MR images of the spine. This study aims to evaluate thresholding for quantification of subtle bone marrow inflammation in the spine and the sacroiliac (SI) joints of patients with PsA and compare it with the SPARCC scoring system.

Methods: Short tau inversion recovery (STIR) MR images of the spine ($N = 85$) and the SI joints ($N = 95$) of patients with PsA ($N = 41$) were analyzed. A threshold was applied to visible bone marrow in order to mask areas with higher signal intensity, which are consistent with inflammation. These areas were considered as inflammatory lesions. The volume and relative signal intensity of the lesions were calculated. Results from thresholding were compared to SPARCC scores using linear mixed-effects models. The specificity and sensitivity of thresholding were also calculated.

Results: A significant positive correlation between the volumes and mean relative signal intensities, which were calculated by thresholding analysis, and the SPARCC scores was detected for both spine ($p < 0.001$) and SI joints ($p < 0.001$). For the spine, thresholding had sensitivity and specificity of 83% and 76% respectively, while for the SI joints the values were 51% and 88% respectively.

Conclusions: Thresholding allows quantification of subtle bone marrow inflammatory edema in patients with psoriatic arthritis, and could support SPARCC scoring of the spine. Improved image processing and inclusion of automatic segmentation are required for thresholding of STIR images to become a rapid and reliable method for quantitative measures of inflammation.

Trial registration: NCT02995460 (December 14, 2016) – Retrospectively registered.

Keywords: Psoriatic arthritis, MRI, Image processing, SPARCC, Bone marrow inflammatory edema

---

* Correspondence: ioanna.chronaiou@ntnu.no
[1]Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, 7491 Trondheim, Norway, Trondheim, Norway
Full list of author information is available at the end of the article

Chronaiou *et al. BMC Musculoskeletal Disorders* (2017) 18:497

Page 2 of 8

Background

Psoriatic Arthritis (PsA) is a chronic inflammatory joint disease associated with psoriasis [1] that manifests with inflammation in peripheral joints, axial skeleton, enthesitis and dactylitis [2]. Magnetic resonance imaging (MRI) allows visualization of inflammation and damage in all structures involved in PsA [3] and has been found to be more sensitive to inflammatory changes than clinical examination [4]. The prevalence of PsA ranges from 20 to 420 per 100,000 population in all countries except Japan, where prevalence is lower [5].

The prevalence of axial PsA varies from 25% to 75% of PsA patients depending on the criteria used [6, 7]. In a subgroup of patients with axial PsA, there is subclinical inflammation in the absence of clinical symptoms. Detecting radiographic involvement of the spine and the sacroiliac (SI) joints in these patients is important for diagnosis and classification. Accurate quantification of small inflammatory lesions in the spine and SI joints is important as it may reflect subclinical disease activity [3] and be a predictor of radiographic progression [8, 9]. Additionally, an accurate method that can detect minor changes will be able to assess the effect of treatment or intervention. A semi-quantitative method established by the spondyloarthritis research consortium of Canada (SPARCC) can be used in order to assess the disease activity in MR images of the spine and SI joints. This scoring method is reliable and sensitive to changes [10], but it requires a trained reader and is labor-intensive. A computer-aided and potentially automatic method for the quantification of bone marrow inflammation is thus a possible time-efficient alternative.

Manual methods for image analysis rely on human vision, which is very sensitive, but are reader-dependent and prone to subjective errors and variation. Automatic methods offer advantages over manual methods of analysis. They are standardized and reproducible and have a consistent accuracy. Moreover, automatic methods follow a systematic approach, thus are highly repeatable. Once established, the procedure can easily be consistently applied in a large number of images, is objective and less time-consuming.

Thresholding has been used in a previous study to quantify inflammation in the SI joints of patients with chronic lower back pain originating in the SI joints [11]. Application of this approach is based on the fact that inflammatory lesions have higher signal intensity than normal bone marrow in short tau inversion recovery (STIR) images [11], which are typically used for imaging of bone marrow inflammation. The proposed method is potentially faster, easier and more robust than SPARCC and more importantly, eliminates the need of a trained reader. Another advantage of thresholding compared to SPARCC is that the former uses all images in the image set, while in the latter only a selection of slices is scored. Altogether, thresholding could be an alternative to SPARCC for quantification of subtle bone marrow inflammation in the spine and SI joints of patients with psoriatic arthritis. However, the validity of thresholding in this setting has not yet been tested.

This study aims to validate thresholding as a method suitable for accurate quantification of subtle bone marrow inflammation in patients with PsA and compare it with the SPARCC scoring system.

Methods

Patients

Patients diagnosed with PsA ($N = 43$) were recruited to the study, all being under optimal treatment at the time. Eligible patients were participating in a randomized clinical trial with high intensity interval training as intervention. Trial participants fulfilled the CASPAR-criteria for PsA, were between 18 and 65 years old and were able to exercise. Exclusion criteria were unstable PsA, unstable ischemic vascular disease, severe pulmonary disease, pregnancy, breastfeeding and drug or alcohol addictions. Two patients were excluded due to conditions that could influence the MR image analysis, one due to incidental findings (lymphoma) and one due to anomaly in the SI joints. Thus, 13 men with a mean age of 48 years (range: 30–64 years) and 28 women with a mean age of 48 years (range: 23–65 years) were included ($N = 41$). All patients have signed informed consent and the Norwegian Regional Committee for Medical and Health Research Ethics has approved the study. Patients were randomized into a control and an intervention group as part of a separate study. Effects of intervention are out of the scope of this study. Clinical evaluation at baseline, patient global assessment [mean ± standard deviation (SD): 42 ± 23 mm], disease activity score of 28 joints (mean ± SD: 2.9 ± 1.1), Bath ankylosing spondylitis disease activity index (mean ± SD: 3.4 ± 1.8), quality of life questionnaire, and high-sensitivity C-reactive protein (*hs-CRP*, median: 4.2 mg/L, range: 0.1 to 28.7 mg/L) provided patient health status.

MRI

All patients underwent MRI examinations of the spine and the SI joints based on standardized protocols [12, 13]. Examinations were performed on two 1.5 T scanners (Scanner 1: Syngo MR B17 upgraded during the study to B19, Scanner 2: Syngo MR D13, Avanto, Siemens Healthcare, Germany). An inversion recovery based sequence (STIR) was used for the examination of the SI joints and the spine in two stations (Table 1). The protocol also included T1 and T2 weighted sequences for anatomical reference. American College of Radiology phantom tests were performed on both scanners as image

Chronaiou *et al. BMC Musculoskeletal Disorders* (2017) 18:497

Page 3 of 8

Table 1 Acquisition parameters

	Spine	SI joints
Orientation	Sagittal	Semi-coronal
TR (msec)	4250	3700
TE (msec)	51 (for lower spine) 52 (for upper spine)	52
TI (msec)	145	145
Slice thickness (mm)	4 (for lower spine) 3 (for upper spine)	4
Gap	10%	10%
Number of slices	Minimum of 16	15

TR time to recovery, *TE* time to echo, *TI* time to inversion

quality control [14]. The effect of using different MR scanners was assessed with statistical analysis.

Clinical evaluation and MRI of the spine and SI joints were performed at one ($N = 4$), two ($N = 20$) and three time-points ($N = 17$). A total of 95 scans of the spine and the SI joints were acquired. Ten image sets of the spine were excluded from the analysis due to human error during the acquisition that resulted in altered protocol and different image weighting. A total of 85 scans of the spine and 95 scans of the SI joints were thus included in image analyses.

Acquisition parameters of short-tau inversion recovery (STIR) sequence used for the examination of the spine and the SI joints. Orientation, time to recovery (TR), time to echo (TE), time to inversion (TI), slice thickness, gap and number of slices are presented in the table. Images of the spine we acquired in two stations (lower spine and upper spine).

SPARCC scoring

A rheumatologist (RST) trained for the SPARCC scoring methods, blindly scored the STIR images of the spine and the SI joints according to the SPARCC SI Joint and Spine Inflammation Indices [12, 13]. In short, for the spine, the six most abnormal disco-vertebral levels on the STIR sequence are selected. Three consecutive sagittal slices, that represent the most abnormal slices for each level, are chosen for scoring at that level. The total maximum SPARCC score is 108 for all six levels of the spine. In the SI joints, the six consecutive slices covering the cartilaginous part of the joints, which is the most relevant part of the SI joints when looking for inflammation, are scored. The total maximum SPARCC score is 72 for all six slices of SI joints. Cases with positive SPARCC scores were considered positive for the presence of bone marrow inflammatory edema, whereas cases with SPARCC score of 0 were considered negative.

For the spine, only a total SPARCC score per image set ($N = 85$) was provided, while for the SI joints both a

total SPARCC score per image set ($N = 95$) and a SPARCC score for each chosen slice ($N = 570$) were available.

Thresholding
Image pre-processing
Histogram-matching [15] is a histogram-based intensity normalization method that transforms the histogram of an image so that it is a match to the histogram of a reference image. Histogram-matching was performed to ensure that all image sets had the same overall brightness. All spinal MR images were histogram-matched to one reference spinal image and all MR images from SI joints were histogram-matched to one reference SI joint image. The function imhistmatch in MATLAB (MathWorks, Natick, MA, USA) was used.
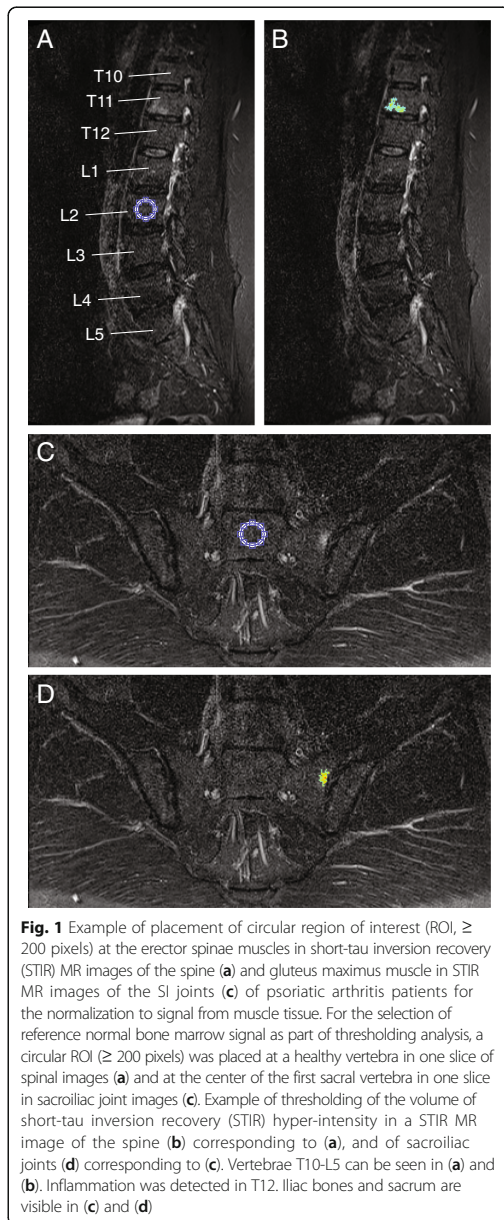
Segmentation of bone marrow
Bone marrow of the sacrum and the iliac bones in the SI joints and vertebral bone marrow in the spine, excluding vascular and neural structures, were manually outlined using 3D Slicer (MIT Artificial Intelligence Lab, USA).

Volume of STIR hyper-intensity
All data processing was performed in Matlab R2016b (The MathWorks Inc., Natick, MA, 2000) using in-house scripts.

A signal intensity threshold consistent with inflammation was calculated from a circular ROI (≥ 200 pixels) at a healthy vertebra in one slice of the spinal image series and at the center of the first sacral vertebra in one slice of the SI joint image series (Fig. 1a, c). The criterion for choosing the ROI placement was the absence of bone marrow inflammatory edema. The mean signal intensity in this ROI was used as reference normal bone marrow signal intensity. A threshold was defined as the sum of the mean signal intensity in the reference normal bone marrow ROI and a percentage of the SD of signal intensity in that ROI. A receiver operating characteristic (ROC) curve was used to define the optimal threshold for the spine (area under curve [AUC] = 0.81) and the SI joints (AUC = 0.70) (Fig. 2). For the spine, the optimal threshold was defined as the sum of the mean signal intensity in the reference normal bone marrow ROI and 4.15 times the SD of signal intensity in that ROI. For the SI joints, the optimal threshold was defined as the sum of the mean signal intensity in the reference normal bone marrow ROI and 2.64 times the SD of signal intensity in that ROI.

All pixels with higher signal intensity than the threshold, consistent with inflammation [11], were selected and further used for the calculation of the volume of STIR hyper-intensity ($volume_{hyper}$) in the vertebral bodies. All connected components (objects) in the resulting

Chronaiou *et al. BMC Musculoskeletal Disorders* (2017) 18:497

Page 4 of 8

Fig. 1 Example of placement of circular region of interest (ROI, ≥ 200 pixels) at the erector spinae muscles in short-tau inversion recovery (STIR) MR images of the spine (**a**) and gluteus maximus muscle in STIR MR images of the SI joints (**c**) of psoriatic arthritis patients for the normalization to signal from muscle tissue. For the selection of reference normal bone marrow signal as part of thresholding analysis, a circular ROI (≥ 200 pixels) was placed at a healthy vertebra in one slice of spinal images (**a**) and at the center of the first sacral vertebra in one slice in sacroiliac joint images (**c**). Example of thresholding of the volume of short-tau inversion recovery (STIR) hyper-intensity in a STIR MR image of the spine (**b**) corresponding to (**a**), and of sacroiliac joints (**d**) corresponding to (**c**). Vertebrae T10-L5 can be seen in (**a**) and (**b**). Inflammation was detected in T12. Iliac bones and sacrum are visible in (**c**) and (**d**)

volumes that have fewer than 10 pixels were removed, as they were considered artefacts. $Volume_{hyper}$ was acquired by adding the volumes of all hyper-intense pixels. The number of objects per image set, which represent different lesions, was calculated.

Fig. 2 A receiver operating characteristic curve for the spine (continuous line) and the sacroiliac joints (dashed line) was plotted in order to define the optimal thresholds (shown in circle)
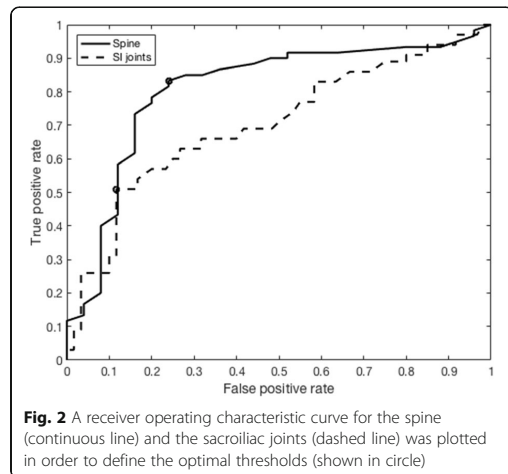
Relative signal intensities of STIR hyper-intense pixels

All hyper-intense pixels were normalized to the mean signal intensity of normal bone marrow [11]. The relative signal intensities of STIR hyper-intense pixels ($S_{RelHyper}$) were calculated according to Eq. 1.

$$S_{RelHyper} = \left(S_{hyper} - \frac{\sum_{i=1}^{n} S_{bone,i}}{n} \right) / \frac{\sum_{i=1}^{n} S_{bone,i}}{n} \qquad (1)$$

where S_{hyper} is the signal intensity value of the respective hyper-intense pixel, S_{bone} is the signal intensity of the pixel included in the reference normal bone marrow ROI and n represents the number of pixels in the reference normal bone marrow ROI [11]. The mean ($S_{RelHyper,\ mean}$), the median ($S_{RelHyper,\ median}$), 75-percentile ($S_{RelHyper,\ 75perc}$) and 90-percentile ($S_{RelHyper,\ 90perc}$) of $S_{RelHyper}$ were calculated for all image sets.

Statistical analysis

Spearman's rank-order correlation between *SPARCC scores* and *volume_{hyper}, $S_{RelHyper,\ mean}$,* number of objects per image set and *hs-CRP* and was calculated in IBM SPSS Statistics (IBM SPSS Statistics for Macintosh, Version 22.0).

Linear mixed-effects models (LMM) [16] were built in R 3.1.1 using the function *lme* from the 'nlme' package [17] employing the method of restricted maximum likelihood. LMM incorporate two types of effects: fixed, which are systematic and controlled, and random, which encompass unsystematic differences not accounted for by the fixed effects, e.g. variation between patients. The fixed effects are essentially different explanatory variables or classification factors whose relationship with the response variable is evaluated simultaneously. LMM

Chronaiou et al. BMC Musculoskeletal Disorders (2017) 18:497

Page 5 of 8

models were built for data from both spine and SI joints separately, including the categorical fixed effects of *intervention group* (intervention or control), *time of scan* (time-point 1, 2 or 3), and *MR scanner* (machine 1 or 2) (without interaction terms). The continuous fixed effect of *SPARCC score* was also included, while the random effect was the *patient number* and the response variables were $volume_{hyper}$, $S_{RelHyper, \ mean}$, $S_{RelHyper, \ median}$, $S_{RelHyper, \ 75perc}$, $S_{RelHyper, \ 90perc}$ or the number of objects per image set from thresholding. The latter were log10 transformed to comply with normality assumptions, confirmed by visual inspection of residual q-q plots and histograms.

We calculated sensitivity and specificity of thresholding compared to SPARCC from the proportion of patients identified with inflammatory lesions. Both for the spine ($N = 85$) and the SI joints ($N = 95$), the calculations were performed per image set including all the slices in each image set. In addition, for the SI joints, the calculations were performed per image set including only the six slices that were chosen for the SPARCC scoring method ($N = 95$) and per slice for the slices that were chosen for the SPARCC scoring method ($N = 570$).

Results

SPARCC
For the 85 image sets covering the spine, 60 were positive for inflammation using the SPARCC scoring method. For the 95 image sets covering the SI joints, 35 had a positive SPARCC score. Overall, 84 out of 570 slices of the SI joints were given a positive SPARCC score.

For the image sets with positive SPARCC scores, the mean score was 10.5 for the spine and 4.3 for the SI joints. Including all image sets, with positive or zero SPARCC scores, the mean SPARCC score for the spine was 7.4 ranging from 0 to 51 out of maximum possible score 108. The mean SPARCC score for the SI joints is 1.6, ranging from 0 to 17, out of a maximum possible score of 72.

Thresholding
Thresholding revealed inflammatory lesions in 56 out of 85 image sets of the spine and 25 out of 95 image sets of the SI joints. In the analysis of SI joints, when including only the six slices that were chosen with the SPARCC method, 25 out of 95 image sets were found positive for the presence of inflammatory lesions. In total, 92 out of 570 slices of the SI joints showed inflammation when analyzed using thresholding.

For the image sets that had inflammatory lesions, mean $volume_{hyper}$ was 2.92 cm^3 and 2.77 cm^3 for the spine and the SI joints, respectively. Including all image

sets, with or without inflammatory lesions, mean $volume_{hyper}$ was 1.92 cm^3, ranging from 0 to 17.86 cm^3, in the spine and 0.73 cm^3, ranging from 0 to 19.04 cm^3, in the SI joints.

The mean and the range of $volume_{hyper}$, $S_{RelHyper, \ mean}$, $S_{RelHyper, \ median}$, $S_{RelHyper, \ 75perc}$ and $S_{RelHyper, \ 90perc}$ for the spine and the SI joints using all the slices are presented in Table 2. Examples of thresholding of the volume of STIR hyper-intensity in the SI joints and the spine are presented in Fig. 1.

Volume of short-tau inversion recovery (STIR) hyper-intense pixels ($volume_{hyper}$) and measures of lesion relative signal intensities; mean ($S_{RelHyper, \ mean}$), median ($S_{RelHyper, \ median}$), 75-percentile ($S_{RelHyper, \ 75perc}$) and 90-percentile ($S_{RelHyper, \ 90perc}$) of the relative signal intensities of STIR hyper-intense pixels for the spine and the SI joints calculated by thresholding. All values are given with standard deviations and parameter range in brackets.

Statistics
Spearman's rank-order correlation analysis revealed a significant positive correlation between SPARCC score and $volume_{hyper}$ both for the spine (correlation coefficient: 0.74, $p < 0.001$) and the SI joints (correlation coefficient: 0.52, $p < 0.001$). SPARCC score did not correlate significantly with *hs-CRP*. Correlation coefficients calculated by Spearman's rank-order correlation analysis are presented in Table 3.

Results from multilevel LMMs to simultaneously assess the relationship between $volume_{hyper}$, $S_{RelHyper, \ mean}$, $S_{RelHyper, \ median}$, $S_{RelHyper, \ 75perc}$, $S_{RelHyper, \ 90perc}$ or the number of objects per image set and the fixed effects of SPARCC score, intervention group, time of scan and MR scanner are summarized in Table 4. A significant positive correlation between $volume_{hyper}$ and SPARCC score was detected for spine (coefficient ± standard error: 0.11 ± 0.02, $p < 0.001$,) and SI joints (coefficient ± standard error: 0.31 ± 0.05, $p < 0.001$). The *intervention group*, *time of scan* (not shown) and the *MR scanner* were determined to not have a significant effect on the measurements by the thresholding method.

Table 2 Volume of short-tau inversion recovery hyper-intense pixels and measures of lesion relative signal intensities

	Spine ($N = 58$)	SI joints ($N = 36$)
$volume_{hyper}$ (cm^3)	2.92 ± 3.86 (0.04–17.86)	2.77 ± 4.18 (0.03–19.04)
$S_{RelHyper, \ mean}$	1.69 ± 0.12 (1.41–2.01)	0.66 ± 0.13 (0.45–0.92)
$S_{RelHyper, \ median}$	1.72 ± 0.23 (1.31–2.26)	0.63 ± 0.23 (0.38–1.07)
$S_{RelHyper, \ 75perc}$	2.09 ± 0.16 (1.79–2.26)	0.83 ± 0.23 (0.47–1.07)
$S_{RelHyper, \ 90perc}$	2.24 ± 0.04 (2.05–2.26)	1.00 ± 0.14 (0.57–1.07)

STIR short-tau inversion recovery, *SI* sacroiliac, $volume_{hyper}$ volume of STIR hyper-intensity, $S_{RelHyper}$ relative signal intensities of STIR hyper-intense pixels, *75perc* 75-percentile, *90perc* 95-percentile

Chronaiou *et al. BMC Musculoskeletal Disorders* (2017) 18:497

Page 6 of 8

Table 3 Spearman's rank-order correlation

| | SPARCC score | | | |
| | Spine | | SI joints | |
	Coefficient	p-value	Coefficient	p-value
$volume_{hyper}$	0.74	< 0.001	0.52	< 0.001
$S_{RelHyper,\ mean}$	0.67	< 0.001	0.47	< 0.001
Number of lesions	0.72	< 0.001	0.52	< 0.001
hs-CRP	−0.14	0.215	0.091	0.380

SI sacroiliac, *SPARCC* spondyloarthritis research consortium of Canada, *STIR* short-tau inversion recovery, $volume_{hyper}$ volume of STIR hyper-intensity, $S_{RelHyper}$ relative signal intensity of STIR hyper-intense pixels, *hs-CRP* high-sensitivity C-reactive protein

The two methods, SPARCC and thresholding, agreed on the absence of inflammatory activity in 19 out of 85 image sets of the spine, resulting in a sensitivity of 83% and a specificity of 76%. For the SI joints, the agreement was for 53 out of 95 image sets, resulting in a sensitivity of 51% and a specificity of 88%. When comparing the scores of each slice from the whole image set of SI joints, the two methods agreed on 434 slices out of 570 showing no inflammation, resulting in a sensitivity of 48% and a specificity of 89%.

Spearman's rank-order correlation coefficients and p-values for the relationship of thresholding-derived metrics (volume, number of lesions and high-sensitivity C-reactive protein to spondyloarthritis research consortium of Canada.

Linear mixed-effects model (LMM) coefficients and p-values for the relationship of thresholding-derived metrics and number of lesions to spondyloarthritis research consortium of Canada (SPARCC) scores and MR scanner (scanner 1 or 2). The coefficients indicate how much $volume_{hyper}$, $S_{RelHyper,\ mean}$, $S_{RelHyper,\ median}$, $S_{RelHyper,\ 75perc}$ and number of lesions increase (positive coefficient) or decrease (negative coefficient) for every unit increase in the SPARCC score.

Discussion

This study evaluates thresholding as a computer-aided method for quantification of subtle bone marrow

inflammation in the spine and SI joints of PsA patients. Thresholding-derived metrics ($volume_{hyper}$, $S_{RelHyper,\ mean}$, $S_{RelHyper,\ median}$, $S_{RelHyper,\ 75perc}$, $S_{RelHyper,\ 90perc}$ and number of objects per image set) correlate significantly with SPARCC scores both for the spine and the SI joints. However, the agreement on absence or presence of inflammation between the two methods was higher for the spine than for the SI joints, indicating that the proposed method of analysis performs better in the former. All metrics (mean, median, 75th-percentile and 90th-percentile) for the relative signal intensity of the hyper--intense lesions correlate with the same level of significance with the SPARCC scores. We therefore suggest that the $S_{RelHyper,\ mean}$ can be used as a standard metric for relative signal hyper-intensity of inflammatory lesions.

To validate the use of the proposed method, we compared thresholding data to SPARCC scores for 85 image sets of the spine and 95 image sets of the SI joints from 41 PsA patients. In addition, for the 570 slices from SI joints, a slice-by-slice comparison was performed on results from the two methods. There was some disagreement between the two methods. The lesions that thresholding failed to detect in the spine ($N = 10$) had a mean SPARCC score of 3.8, while correctly identified lesions ($N = 50$) had a mean SPARCC score of 11.8. The disagreement was bigger for the SI joints, where lesions that thresholding failed to detect ($N = 17$) had a mean SPARCC score of 2.5, while correctly identified lesions ($N = 18$) had a mean SPARCC score of 5.9. Sensitivity and specificity measures show that thresholding analysis is more accurate in the spine. Spearman's rank-order correlation analysis confirms higher correlation for the spine than the SI joints. Patients included in this study had little to no inflammation, especially in the SI joints, which may suggest that the method performs better in areas with higher inflammatory activity. Additionally, the examined anatomical structures in the spine are in the homogeneous image center of all slices, whereas the examined anatomical structures of the SI joints are more

Table 4 Results from linear mixed-effects model

| | Spine | | | SI joints | | |
| | SPARCC score | | MR scanner | SPARCC score | | MR scanner |
	Coefficient	p-value	p-value	Coefficient	p-value	p-value
$volume_{hyper}$	0.11	< 0.001	0.467	0.31	< 0.001	0.804
$S_{RelHyper,\ mean}$	0.09	0.001	0.347	0.25	< 0.001	0.597
$S_{RelHyper,\ median}$	0.09	0.001	0.348	0.25	< 0.001	0.596
$S_{RelHyper,\ 75perc}$	0.09	0.001	0.338	0.26	< 0.001	0.560
$S_{RelHyper,\ 90perc}$	0.09	0.001	0.348	0.26	< 0.001	0.590
Number of lesions	0.13	< 0.001	0.424	0.37	< 0.001	0.672

SI sacroiliac, *SPARCC* spondyloarthritis research consortium of Canada, *STIR* short-tau inversion recovery, $volume_{hyper}$ volume of STIR hyper-intensity, $S_{RelHyper}$ relative signal intensities of STIR hyper-intense pixels, *75perc* 75-percentile, *90perc* 95-percentile

distant from the homogeneous image center, and also in varying distance through slices. This may affect the homogeneity of the acquired image. Areas that are closer to the coil appear more hyper-intense, resulting in slightly different signal intensities through an image. This issue could have been resolved using appropriate pre-processing. Anatomical differences may also contribute to lower lesion detectability in the SI joints. Additional pre-processing of the SI joint images could be used to correct for the inhomogeneities and signal intensity differences and improve the performance of thresholding in the SI joints.

Examinations were acquired using two MR scanners with different software platforms over the course of a year. During that time, one of the scanners underwent software upgrade. This should have no effect in the results of this study, and LMM also showed that different MR scanners used for imaging did not affect the measurements by the thresholding method.

The thresholding method presented here was first introduced in a previous study [11], where it was used to measure inflammatory changes in the SI joints of patients with lower back pain. However, in that study, the method was not compared to any clinical evaluation score and its validity was not tested. Additionally, the method was not tested for different threshold values to justify for the specific choice of threshold. In our study, the method is also applied in the spine.

One limitation of the thresholding method is that the ROIs of the bone marrow in the spine and the SI joints of the patients were drawn manually, in order to accurately exclude neural structures and blood vessels, but include possible inflammatory lesions. This presupposes a basic knowledge of the anatomy of SI joints. Fully automated methods for the selection of the sacrum and iliac bone ROIs should be explored. A fully automated method for the localization and segmentation of the vertebral units has been used in a previous study as part of a semi-automated framework for comparative visualization of inflammatory bone marrow lesions in MR images of the spine [18]. Combining fully automated segmentation of the spine and thresholding in such a setting could potentially assist in assessing radiological progression of patients with inflammatory lesions in the spine. Time required for SPARCC scoring depends on the experience of the reader, but also on how many lesions a patient has. A trained reader will need approximately 10 min for a patient without lesions and 30–40 min for a patient with many lesions. Time required for manual segmentation of bone marrow of a single image set is approximately 10 min. However, a fully automated segmentation of inflammation will reduce the reading time significantly and make thresholding a quantitative method feasible in the clinic.

A disadvantage of intensity-based methods for image analysis, such as thresholding, is that these methods are not able to differentiate between different pathologies that lead to increased signal intensities in the images, which is something a trained human can do easily. However, SPARCC scoring is used in patients who already have a diagnosis with a pre-investigative probability of having inflammation due to the primary diagnosis (psoriatic arthritis, spondyloarthritis, ankylosing spondylitis). Other approaches, including textural analysis, may be more beneficial in this instance. Another limitation of this study is the absence of a control group.

Overall, automatic thresholding is a novel method which performs relatively well at detecting inflammatory lesions in the spine of PsA patients, but more poorly in the SI joints. In addition to the presence or absence of inflammation, it provides volumetric information and allows localization of the lesions. The implementation of the method is generic enough to allow for application in the quantification of bone marrow inflammation in other types of spondyloarthritis. Fully automated implementation of the thresholding method should be explored.

Conclusion

Thresholding allows quantification of subtle bone marrow inflammation in PsA patients with low SPARCC scores for inflammatory activity. The significant correlation for low inflammatory scores suggests that this method can provide reliable and sensitive quantitative measures for the presence of subtle inflammation in bone marrow. With further studies, automatic segmentation and technique optimization, it is possible that automatic thresholding may eventually be an alternative or supplement to SPARCC scoring.

Chronaiou *et al. BMC Musculoskeletal Disorders* (2017) 18:497

Page 8 of 8

Authors' contributions

Ethics approval and consent to participate

Consent for publication

Competing interests

Publisher's Note

Author details

[1]Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, 7491 Trondheim, Norway, Trondheim, Norway. [2]Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, Trondheim, Norway. [3]Department of Radiology and Nuclear Medicine, St Olav's Hospital, University Hospital in Trondheim, Trondheim, Norway. [4]Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet - Glostrup, Copenhagen, Denmark. [5]Department of Rheumatology, St Olav's Hospital, University Hospital in Trondheim, Trondheim, Norway.

References

1. Kalkan G, Karadag AS. Psoriatic arthritis epidemiology. East J Med. 2014;19(1):1–7.
2. Østergaard M, Glinatsi D, Pedersen SJ, Sørensen IJ. Utility in clinical trials of magnetic resonance imaging for psoriatic arthritis: a report from the GRAPPA 2014 annual meeting. J Rheumatol. 2015;42(6):1044–7.
3. Poggenborg R, Sørensen I, Pedersen S, Østergaard M. Magnetic resonance imaging for diagnosing, monitoring and prognostication in psoriatic arthritis. Clin Exp Rheumatol. 2015;33(5 Suppl 93):66.
4. Wiell C, Szkudlarek M, Hasselquist M, Møller JM, Vestergaard A, Nørregaard J, Terslev L, Østergaard M. Ultrasonography, magnetic resonance imaging, radiography, and clinical assessment of inflammatory and destructive changes in fingers and toes of patients with psoriatic arthritis. Arthritis Res Ther. 2007;9(6):R119.
5. Dhir V, Aggarwal A. Psoriatic arthritis: a critical review. Clin Rev Allerg Immu. 2013;44(2):141–8.
6. Baraliakos X, Coates LC, Braun J. The involvement of the spine in psoriatic arthritis. Clin Exp Rheumatol. 2015;33(5 Suppl 93):S31–5.
7. Gladman DD. Axial disease in psoriatic arthritis. Curr Rheumatol Rep. 2007;9(6):455–60.
8. Savnik A, Malmskov H, Thomsen HS, Graff LB, Nielsen H, Danneskiold-Samsøe B, Boesen J, Bliddal H. MRI of the wrist and finger joints in inflammatory joint diseases at 1-year interval: MRI features to predict bone erosions. Eur Radiol. 2002;12(5):1203–10.
9. Bond SJ, Farewell VT, Schentag CT, Gladman DD. Predictors for radiological damage in psoriatic arthritis: results from a single centre. Ann Rheum Dis. 2007;66(3):370–6.
10. LandewÉ RB, Hermann K-GA, van der Heijde DM, Baraliakos X, Jurik A-G, Lambert RG, Østergaard M, Rudwaleit M, Salonen DC, Braun J. Scoring sacroiliac joints by magnetic resonance imaging. A multiple-reader reliability experiment. J Rheumatol. 2005;32(10):2050–5.
11. Fritz J, Henes JC, Thomas C, Clasen S, Fenchel M, Claussen CD, Lewin JS, Pereira PL. Diagnostic and interventional MRI of the sacroiliac joints using a 1.5-T open-bore magnet: a one-stop-shopping approach. AJR Am J Roentgenol. 2008;191(6):1717–24.
12. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Krishnananthan R, Stone M, Conner-Spady B, Palsat J, Lambert RG. Spondyloarthritis research consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Care Res. 2005;53(4):502–9.
13. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Williams M, Stone M, Conner-spady B, Palsat J, Lambert RG. Spondyloarthritis research consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Care Res. 2005;53(5):703–9.
14. Chen C-C, Wan Y-L, Wai Y-Y, Liu H-L. Quality Assurance of Clinical MRI scanners using ACR MRI phantom: preliminary results. J Digit Imaging. 2004;17(4):279–84.
15. Manjón JV. MRI preprocessing. In: Martí-Bonmatí L, Alberich-Bayarri A, editors. Imaging biomarkers: development and clinical integration. Switzerland: Springer International Publishing; 2017. p. 53–63.
16. Pinheiro J, Bates D. Linear mixed-effects models: basic concepts and examples. Mixed-effects models in S and S-PLUS. New York, NY, USA. New York: Springer; 2000.
17. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team: Nlme: linear and nonlinear mixed effects models. R package version 3.1–117. 2014.
18. Griffith JF, Wang D, Shi L, Yeung DK, Lee R, Shan TL. Computer-aided assessment of spinal inflammation on magnetic resonance images in patients with Spondyloarthritis. Arthritis Rheumatol. 2015;67(7):1789–97.

Paper II

Evaluating the impact of high intensity interval training on axial psoriatic arthritis based on MR images

Ioanna Chronaiou[1], Guro F. Giskeødegård[2], Ales Neubert[3], Tamara V. Hoffmann-Skjøstad[4], Ruth S. Thomsen[5,6], Mari Hoff [5,6], Tone F. Bathen[1], Beathe Sitter[1*]

[1] Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
[2] Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
[3] The Australian E-Health Research Centre, CSIRO Health & Biosecurity, Brisbane, Australia
[4] Department of Radiology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[5] Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
[6] Department of Rheumatology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway

*Corresponding author:

E-mail: beathe.sitter@ntnu.no

Abstract

Introduction

High intensity interval training (HIIT) has shown to benefit patients with psoriatic arthritis (PsA). However, MR imaging has uncovered bone marrow edema (BME) in healthy volunteers after vigorous exercise. The purpose of this study was to investigate MR images of the spine of PsA patients for changes in BME after HIIT.

Materials and Methods

PsA patients went through 11 weeks of HIIT (N=19) or no change in physical exercise habits (N=20). We acquired scores for joint affection and pain and STIR and T1-weighted MR images of the spine at both timepoints. MR images were evaluated for BME by a trained radiologist, by SpondyloArthritis Research Consortium of Canada (SPARCC) scoring, and by extraction of textural features.

Results

No significant changes of BME were detected in MR images of the spine after HIIT. This was consistent for all three image evaluations. Disease activity score in 44 joints (DAS44) and four textural features of PsA lesions decreased for both groups, but the textural features changes were not significant after Bonferroni correction.

Conclusion

BME in spine was not changed after HIIT, supporting that HIIT is safe for PsA patients. Texture analysis of MR images could be more sensitive to changes in BME than radiological evaluation and SPARCC scoring.

Introduction

Psoriatic Arthritis (PsA) is a chronic inflammatory joint disease associated with skin psoriasis that can manifest in the axial skeleton and peripheral joints, and can include dactylitis and enthesitis [1]. The prevalence of PsA range from 20-670 per 100,000 population [2, 3]. Between 30 -50% of PsA patients will develop axial PsA, involving the spine or the sacroiliac joints [4].

Physical exercise gives beneficial effects on inflammation, joint damage and symptoms [5], and is recommended as supplementary treatment to patients with arthritis [6]. It has been suggested that physical exercise, and in particular high intensity interval training (HIIT), can have anti-inflammatory effect [7, 8]. It is however also possible that vigorous exercise can increase the disease burden, as mechanical strain drives both entheseal inflammation and new bone formation and may contribute to further development of spondyloarthritis [9, 10]. To our knowledge, there are no reported studies of PsA patients under HIIT with imaging of the axial skeleton. It has thus not been explored if vigorous exercise decreases or increases the disease burden in terms of BME in PsA patients.

Short and long-term beneficial effects of HIIT on disease activity, patient disease perception and the risk of cardiovascular disease were recently reported in PsA patients [11, 12]. Exercise led to reduced fatigue and cardiovascular risk factor in terms of truncal fat and maximal oxygen uptake, whereas scores for joint affection and pain were compatible with the control group. This finding is important, as it shows that HIIT have no negative impact on the disease burden and can be recommended for PsA patients [11, 12]. A sub-group of these patients were also examined by MR imaging before and after intervention. MR imaging can portray inflammation in the structures involved [13] and evaluate inflammatory changes with higher sensitivity than

68　clinical examination [14]. Short tau inversion recovery (STIR) is the recommended MR

69　imaging sequence for axial PsA [15, 16], presenting edematous lesions with hyperintense

70　signals [17]. Radiological assessment of edematous lesions is based on hyperintensity in STIR

71　MR images, located in two or more sites and/or two or more slices [18]. Lesions can be

72　confirmed as hypointense signals in T1-weighted MR images. A semi-quantitative scoring

73　system of disease activity can be highly sensitive to changes in the spine, like the

74　Spondyloarthritis Research Consortium of Canada (SPARCC) scoring system [16].

75

76　It is further possible to characterize disease through quantitative image features, also diseases

77　that are difficult to identify in the MR image by human vision alone. Spatial variations of grey-

78　level intensity in an image are perceived by a human observer as texture [19]. Texture analysis

79　allows the quantification of these spatial variations by computing a set of metrics, called

80　textural features, from the distribution of pixel grey-level intensities in an image [20], and has

81　been used in studies of BME [21-26]. Image gradient measures the magnitude of directional

82　grey-level intensity variations in an image [27]. Spatial position, variation and directionality of

83　the signal intensities inside an image could provide more detailed image characteristics than

84　observer-based methods for MR image evaluation, and thus be more sensitive to subtle changes

85　[23].

86

87　The aim of this study was to assess whether HIIT in PsA patients led to detectable changes in

88　the axial skeleton, by investigating MR images of the spine for BME. Additionally, we have

89　explored the potential of textural features to detect BME changes.

Materials and Methods

Patient cohort

The presented study is part of a randomized controlled trial with HIIT as intervention, conducted at St. Olavs hospital and NTNU–Norwegian University of Science and Technology, Trondheim, Norway from 2013 to 2015 [11, 12]. Participants fulfilled the ClASsification for Psoriatic ARthritis (CASPAR) criteria and were between 18 and 65 years. The intervention group (N=19) performed HIIT three times per week for 11 weeks, whereas the control group (N=20) did no change in pre-study physical exercise habits [12]. All patients signed informed consent and the Norwegian Regional Committee for Medical and Health Research Ethics approved the study (Trial registration: NCT02995460).

Disease activity scores

Scores for joint affection and pain were assessed at baseline and after 11 weeks as previously described [12], and included patient global assessment (PGA), high sensitivity C-reactive protein (hs-CRP), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and Disease Activity Score in 44 joints (DAS44).

MR image acquisition

The MR imaging was performed of the spine in two stations using an inversion recovery-based sequence (STIR) and T1-weighted turbo spin-echo (TSE) sequence based on a standardized protocol [16] as previously described [28]. For two patients, both in the HIIT group, the second MR imaging was not performed, and both patients were excluded from evaluation by radiologist and SPARCC scoring, leaving MR images from 37 patients for the further analyses.

113 For 10 of the participants, the MR imaging protocol deviated with respect to spatial resolution

114 in the first (N=1) or second MR imaging (N=9), and these images were excluded from analysis

115 by textural features.

116

Image analysis

Radiological evaluation

119 The MR images were evaluated by a radiologist for BME at both time-points. The radiologist

120 was blinded with respect to intervention. BME was identified by hyperintense signals in STIR

121 MR images, supported by hypointense signals in T1-weighted MR images. To be considered

122 positive for BME, the hyperintense signal should be located in two or more sites and/or two or

123 more slices [18, 29]. Images were also assessed with respect to change from first to second MR

124 imaging; categorized to stable, increased, or reduced BME.

125

SPARCC scoring

127 The STIR MR images at both time-points were scored by a trained rheumatologist as

128 previously described [16]. In brief, the six most abnormal disco-vertebral levels on the STIR

129 sequence are selected. Three consecutive sagittal slices, that represent the most abnormal slices

130 for each level, are chosen for scoring at that level. The total maximum SPARCC score is 108

131 for all six levels of the spine. The SPARCC scores were also categorized with respect to change

132 from first to second MR imaging; stable, increased, or reduced SPARCC score.

133

Image pre-processing and textural feature extraction

135 Vertebral bone marrow, excluding vascular and neural structures, were manually segmented

136 using 3D Slicer (MIT Artificial Intelligence Lab, USA) in MR images from all patients (N=37),

137 comprising images from both time-points (N=27), only first (N=9) or only last (N=1) time-

138 point. Images of the spinal column were pre-processed using a customized intensity adjustment

139 procedure based on the nonparametric nonuniform intensity normalization (N4) bias field

140 correction algorithm [30]. Pixel intensity values were normalized by matching the histogram

141 extracted from the spinal column to the histogram of a randomly selected atlas image. Image

142 noise was reduced using an in-house implementation of anisotropic (Perona-Malik) diffusion

143 smoothing filter (iterations = 15, an integration constant = 1/7, time step = 0.01, conductance

144 = 1.0) [31]. BME in all image sets was manually segmented using 3D Slicer. The segments

145 were verified by a trained radiologist.

146

147 Three pixel-wise types of image textural features were calculated; seven intensity features, 10

148 gradient features and four grey level co-occurrence matrix (GLCM) textural features, all

149 described in Supplementary Table 1. Intensity features were the grey-level intensity value of

150 the central pixel, and the mean, the median, the standard deviation, the minimum, the maximum

151 and the semi-interquartile range of the grey-level intensity values. For the extraction of gradient

152 features, 2-dimensional directional gradients for x-axis (G_x) and y-axis (G_y) were calculated

153 using a Sobel gradient operator in the *imgradientxy* function in MATLAB (MathWorks,

154 Natick, MA, USA). GLCM features were extracted using the *graycomatrix* and *graycoprops*

155 functions in MATLAB (MathWorks, Natick, MA, USA) at four orientations (0º, 45º, 90º and

156 135º) with a distance of 1 pixel. The resulting GLCM feature was the mean of the GLCM

157 feature values in these orientations. Feature maps were created for each feature using a sliding

158 window implementation. In this approach, an orthogonal 3-by-3 box/kernel "slides" in the

159 region of interest, in this case the segmented bone marrow. The features were calculated in

160 each orthogonal kernel position and correspond to the central pixel of the box.

161

Statistical analysis

Changes in patient characteristics (PGA, hs-CRP, BASDAI, DAS44, and SPARCC score) before and after intervention were analyzed using Wilcoxon signed-rank test. Changes in BME status and SPARCC score were categorized, and differences between HIIT and control group were analyzed by Fisher's exact probability test. Analyses were performed in SPSS (IBM SPSS Statistics v 26), and p-values < 0.05 were considered statistically significant.

Differences in feature values from voxels in bone marrow edema and healthy voxels were assessed by linear mixed models. We used patient number and scan as random effects, whether the voxel is healthy or pathological as fixed effect, and textural feature as response variable. Further, linear mixed models were used to assess changes between first and second time point and if these changes were different between the HIIT and control groups. For this analysis, average feature values per individual per time point were used as response variables, as it was not possible to match lesions between the two time points. Fixed effects were time (whether the scan was a baseline or 11 weeks scan), intervention (whether the patient was in the HIIT or the control group) and the interaction term between time and intervention (time*intervention), and patient number was random effect. The time variable was reference coded to the baseline measurement, and the intervention variable was sum coded.

Bonferroni correction was used to correct p-values for multiple comparisons from all three linear mixed-models. The statistical level of significance was set to p<0.05. Statistical analyses were performed in in RStudio: Integrated Development for R (RStudio, PBC, Boston, MA) and Matlab R2019A.

Results

Patient cohort

Demographic characteristics of study participants in control and HIIT groups, and their scores for joint affection and pain at baseline and after 11 weeks are shown in Table 1. BASDAI decreased for the HIIT group and DAS44 decreased for both groups after 11 weeks.

Table 1. Description of study participants.

	HIIT			Controls		
	Baseline	11 weeks	p	Baseline	11 weeks	p
Number, n	19		-	20		-
Men, n (%)	4 (21)		-	8 (40)		-
Women, n (%)	15 (79)		-	12 (60)		-
Years, median (range)	52 (39 – 64)		-	45 (23 – 64)		-
PGA[a], median (range)	50.0 (1.0–95.0)	43.0 (3.0– 81.0)	0.465	46.0 (6-85)	35.5 (0-89)	0.227
hs-CRP[b] mg/L, median (range)	1.8 (0.4-24.0)	2.1 (0.5-10.2)	0.096	2.2 (0.1-28.7)	2.2 (0.3-22.0)	0.571
BASDAI[c], median (range)	4.0 (0.4-8.3)	3.2 (0.5-6.6)	0.049	3.7 (0.3-6.7)	2.6 (0.2-7.7)	0.133
DAS44[d], median (range)	2.3 (0.8-3.3)	1.9 (0.5-2.4)	0.001	2.3 (0.6-3.1)	1.7 (0.6-3.0)	0.007

Demographic characteristics of study participants in high intensity interval training (HIIT) and control groups, and their scores for joint affection and pain at baseline and after 11 weeks [11]. Mean values of scores for joint affection and pain within groups at baseline and 11 weeks were compared by Wilcoxon signed-rank test.

[a] PGA: Patient Global Assessment, range from 0 to 100. [b] hs-CRP: high-sensitive C-reactive protein. [c] BASDAI: Bath Ankylosing Spondylitis Disease Activity Index, range from 0 to 10. [d] DAS44: disease activity score in 44 joints, range from 0.2 to 9.9.

Image analysis

Radiological evaluation

Examples of acquired MR images are shown in Figure 1, demonstrating the PsA lesions of two patients with low and high disease burden. MR images of the spine from 21 of 37 patients were found negative with respect to BME in the radiological evaluation at both timepoints. Sixteen patients (43%) were identified with BME, consistent with axial manifestation of PsA. The

206   radiologically manifested axial PsA was considered mild to moderate for both groups, and

207   disease burden in terms of BME was stable. The findings are summarized in Table 2. The

208   number of patients with changes of BME after 11 weeks was not significantly different between

209   the HIIT and control group (p-value: 0.50).

210

211   **Figure 1. MR image examples.** Sagittal STIR MR images of upper spine in two patients.

212   Image A (baseline) and B (11 weeks) are of a patient with low SPARCC scores (4 and 2,

213   respectively). The images show the spine from the second cervical vertebrae (C2) to thoracic

214   vertebrae 12 (TH12). Segmented region in red in B show minor BME in corner of TH10. Image

215   C (baseline) and D (11 weeks) are of patient with high SPARCC scores (39 and 50,

216   respectively). The images show the spine from the second cervical vertebrae (C2) to thoracic

217   vertebrae 11 (TH11), and segmented regions in red in D show BME in C5 to C7, and in TH5

218   to TH9.

219

220   **Table 2. Bone marrow edema by radiological evaluation of MR images.**

	HIIT		Controls	
	Baseline	11 weeks	Baseline	11 weeks
Number, n	17		20	
BME detected, n (%)	9 (53)	9 (53)	5 (25)	5 (25)
No change, n (%)	17 (100)		17 (85)	
Increased BME, n (%)	0 (0)		1 (5)	
Reduced BME, n (%)	0 (0)		2 (10)	

225   Results from radiological evaluation for BME in STIR and T1-weighted MR images of the

226   participants in high intensity interval training (HIIT) and control groups at baseline and after

227   11 weeks. Number of participants with detectable changes after 11 weeks were not significantly

228   different for the two groups (Fisher's Exact Probability Test p-value: 0.50).

229

SPARCC scoring

231 MR images for 11 of the 37 patients were found negative with respect to BME by the SPARCC

232 scoring at both time-points. 23 of the patients (62%) had a positive SPARCC score at both

233 timepoints. Participants in this study had a median baseline SPARCC score of 4.0 (Table 3).

234 The number of patients with changes in SPARCC scores after 11 weeks was not significantly

235 different between the HIIT and control group (p-value: 1).

236

237 The radiological evaluation and SPARCC scoring were consistent for 49 of the MR images at

238 baseline and 11 weeks. Out of these, 23 were found negative and 26 were found positive for

239 BME by both methods. 23 MR images with a positive SPARCC score were judged to be BME

240 negative by the radiological evaluation. Two MR images with a SPARCC score of 0 were

241 identified as BME positive in the radiological evaluation. SPARCC scoring (Table 3) identified

242 changes from baseline to 11 weeks in more patients than the radiological evaluation for BME

243 (Table 2). The changes in BME identified by radiological evaluation for three patients were in

244 agreement with changes in SPARCC scores for these patients.

245

246 **Table 3. Bone marrow edema by SPARCC scoring of MR images.**

	HIIT		Controls	
	Baseline	11 weeks	Baseline	11 weeks
Number, n	17		20	
SPARCC score > 0, n (%)	13 (72)	13 (72)	13 (65)	10 (50)
SPARCC score, median (max value)	4.0 (39)	5.0 (50)	4.0 (36)	0.5 (20)
No change in SPARCC score, n (%)	5 (29)		9 (45)	
Increased SPARCC score, n (%)	5 (29)		2 (10)	
Reduced SPARCC score, n (%)	7 (41)		9 (45)	
No change by MIC, n (%)	14 (82)		16 (80)	
Increased SPARCC by MIC, n (%)	1 (6)		1 (5)	
Reduced SPARCC by MIC, n (%)	2 (12)		3 (15)	

247 Results from evaluation of STIR MR images of the participants in high intensity interval

248 training (HIIT) and controls groups at baseline and after 11 weeks by SPARCC scoring [16].

249   Changes in SPARCC scores are given for absolute values, and by using a threshold of SPARCC

250   score of five according to minimally important changes (MIC) [32]. Number of participants

251   with detectable changes after 11 weeks were not significantly different for the two groups

252   (Fisher's Exact Probability Test p-value: 1).

253   [a] SPARCC: SpondyloArthritis Research Consortium of Canada, score range from 0 to 108.

254

255

256   ## Textural features

257   Mean and standard deviation of features extracted from MR images of pathological (BME

258   lesions) and healthy voxels are presented in Table 4. The mean values for all but one (g_7)

259   extracted features were significantly different in pathological compared to healthy voxels. With

260   the exception for one of the GLCM features ($f1$), all mean values for textural features were

261   higher in pathological than healthy voxels. Mean values for four textural features ($g6, g10, f3$

262   and $f4$) of pathological voxels were reduced when observing the whole cohort after 11 weeks

263   (Table 5). These changes were not significant after Bonferroni-correction for multiple

264   hypothesis testing. We observed no significant differences in the changes from baseline to

265   week 11 in textural features of PsA lesions between the HIIT group and control group.

266

267

268 **Table 4. Textural features of voxels in bone marrow edema and healthy voxels.**

	Voxels with BME (N = 3289)		Healthy voxels (N = 3289)			
	Mean	*SD*	Mean	*SD*	p-value	q-value
i_1	159.8	*21.0*	117.1	*21.7*	<0.001	**<0.001**
i_2	150.1	*18.0*	116.6	*16.9*	<0.001	**<0.001**
i_3	153.2	*19.0*	117.5	*16.5*	<0.001	**<0.001**
i_4	20.6	*10.2*	13.4	*7.3*	<0.001	**<0.001**
i_5	104.3	*34.2*	87.8	*29.8*	<0.001	**<0.001**
i_6	178.2	*18.4*	139.9	*16.7*	<0.001	**<0.001**
i_7	14.1	*8.2*	8.6	*5.2*	<0.001	**<0.001**
g_1	3219.6	*1571.4*	2231.4	*1188.8*	<0.001	**<0.001**
g_2	2557.3	*1250.6*	1777.9	*948.2*	<0.001	**<0.001**
g_3	128.8	*62.9*	89.3	*47.6*	<0.001	**<0.001**
g_4	102.3	*50.0*	71.1	*38.0*	<0.001	**<0.001**
g_5	78.1	*43.3*	51.4	*31.5*	<0.001	**<0.001**
g_6	61.5	*34.0*	40.9	*25.1*	<0.001	**<0.001**
g_7	115.5	*59.9*	80.4	*43.9*	<0.001	**<0.001**
g_8	20.3	*17.8*	15.9	*12.8*	0.062	1
g_9	292.2	*150.5*	201.0	*115.1*	<0.001	**<0.001**
g_{10}	56.3	*34.4*	36.2	*23.6*	<0.001	**<0.001**
f_1	4.52	*1.7*	5.31	*1.91*	<0.001	**<0.001**
f_2	0.52	*0.23*	0.32	*0.22*	<0.001	**<0.001**
f_3	0.13	*0.09*	0.09	*0.03*	<0.001	**<0.001**
f_4	0.56	*0.08*	0.51	*0.06*	<0.001	**<0.001**

269 Mean values and standard deviation (SD) of features, extracted from MR images of voxels in

270 bone marrow edema (BME) and healthy voxels. Differences in feature values for pathological

271 and healthy voxels were examined using linear mixed-effects models. p-values and Bonferroni

272 corrected p-values (q-values) are reported.

Table 5. Textural features of voxels in bone marrow edema at baseline and 11 weeks.

	Baseline		11 weeks		Changes from baseline to week 11		Differences between HIIT and control
	Mean	*SD*	Mean	*SD*	p-value	q-value	p-value
i1	160.5	*22.6*	159.2	*19.7*	0.660	1	0,912
i2	149.7	*19.4*	150.4	*16.8*	0.592	1	0,986
i3	153.6	*20.9*	152.9	*17.4*	0.666	1	0,873
i4	21.6	*10.9*	19.8	*9.5*	0.531	1	0,526
i5	99.2	*37.9*	108.2	*30.4*	0.100	1	0,984
i6	177.7	*17.6*	178.6	*18.9*	0.488	1	0,908
i7	14.5	*8.9*	13.8	*7.6*	0.798	1	0,858
g1	3387.5	*1708.0*	3089.4	*1443.7*	0.633	1	0,672
g2	2679.1	*1344.4*	2462.7	*1164.3*	0.592	1	0,620
g3	135.5	*68.3*	123.6	*57.7*	0.633	1	0,672
g4	107.2	*53.8*	98.5	*46.6*	0.592	1	0,620
g5	85.3	*47.4*	72.4	*39.0*	0.073	1	0,348
g6	66.7	*36.6*	57.4	*31.2*	**0.049**	1	0,270
g7	119.6	*66.2*	112.3	*54.4*	0.868	1	0,8070
g8	20.2	*19.5*	20.3	*16.3*	0.426	1	0,845
g9	316.3	*164.4*	273.6	*135.9*	0.243	1	0,421
g10	60.5	*37.8*	53.0	*31.1*	**0.047**	0.980	0,820
f1	4.6	*1.8*	4.4	*1.6*	0.945	1	0,784
f2	0.48	*0.24*	0.54	*0.22*	0.088	1	0,573
f3	0.14	*0.11*	0.11	*0.06*	**0.026**	0.555	0,282
f4	0.57	*0.09*	0.55	*0.07*	**0.030**	0.620	0,429

Mean values and standard deviation (SD) of pixel-wise textural features of BME in STIR MR images of psoriatic arthritis patients. Changes in feature values from the baseline to the 11 weeks scan for the full cohort, and differences in changes between the HIIT and control groups, were investigated using a linear mixed effect model. p-values and Bonferroni corrected p-values (q-values) are reported.

Discussion

No significant changes were observed in MR images of the spine after HIIT training for 11 weeks. This finding was consistent for radiological evaluation, SPARCC scoring and textural features of MR images. Values for 20 out of 21 textural features were significantly different in voxels of BME compared to voxels of healthy bone marrow. No textural features of PsA lesions

284   were significantly different when comparing changes in values after 11 weeks between the

285   HIIT and control groups.

286

287   43% and 62% of study participants were found positive for BME by radiological evaluation

288   and SPARCC scoring, respectively (Table 2). The fraction of BME positive participants by

289   SPARCC scoring is above the reported 30-50% of PsA patients with axial involvement [4].

290   More positive findings by SPARCC scoring than by the ASAS criteria has been reported

291   previously [33]. The difference of BME positive participants between the two methods is

292   probably caused by different readers and principal differences in the methods. Standardized

293   methods for scoring of axial spondyloarthritis (axSpA) are subject to some variation between

294   readers [34]. Images of little active inflammation is more subject to low inter-reader

295   correlation, and the mild to moderate disease burden in the cohort of this study is thus suspected

296   to contribute to the difference of the two methods. Both methods rely on hyperintensity in STIR

297   images, where edema related to inflammation can be detected as a bright signal on a dark

298   background in subchondral bone marrow [35]. The use of T1-weighted images to support the

299   radiological evaluation is prone to rejection of positive findings in STIR images, which may

300   explain fewer positive cases by the radiological evaluation. The radiological evaluation

301   identified changes in BME from baseline to week 11 for three patients, which agreed with

302   higher and lower SPARCC scores for these patients. SPARCC scoring identified changes from

303   baseline to week 11 for more patients, but in general of minor magnitude. Applying a SPARCC

304   score threshold of five for minimally important change [32] reduced number of patients with

305   changes from baseline to week 11 from 23 to seven.

306

307   Quantitative methods for analysis of STIR MR images have been proven to discriminate

308   between active therapy and placebo after 12 weeks of treatment in clinical trials of ankylosing

309   spondylitis [36, 37]. Changes occurring in the spine due to the HIIT should thus be detectable

310   with the current MR imaging protocol and methods for image analysis. MR imaging has

311   previously identified BME in healthy individuals and in athletes, suggesting that mechanical

312   strain contribute to BME [33, 38]. Two studies contradict that HIIT may increase disease

313   burden. For this cohort of PsA patients, it has been shown that HIIT have beneficial effects on

314   fatigue and cardiac risk factors, without increased joint affection and pain [11, 12]. It has also

315   been shown that Ankylosing Spondylitis Disease Activity Score (ASDAS) and BASDAI were

316   significantly reduced after 3 months of HIIT in patients with axSpA [8]. Our current study

317   showed no significant changes in BME in the spine from HIIT, by radiological evaluation,

318   SPARCC scoring or texture analysis of MR images (Table 2 – 5), which supports that HIIT is

319   safe to recommend to patients with PsA.

320

321   Mean values of textural features were different in voxels from BME compared to voxels from

322   healthy bone marrow (Table 4). These observed differences are consistent with a previous

323   study, where textural features of MR images have been applied in machine learning to classify

324   active inflammation in sacroiliac joints [24]. Choice of textural features are also important for

325   successful tissue discrimination [26]. We surveyed intensity, gradient and GLCM textural

326   features, which partly has been utilized in other studies with classification of BME [23, 25,

327   39]. These studies also included histogram and run-length matrix features. In studies of

328   osteoarthiritis in the knee [23, 40, 41], most of texture, histogram and run-length matrix

329   features were all significantly different between the patient groups. When discriminating the

330   post-radiation lesions edema, fatty conversion and hemorrhage, Romanos *et al.* found that

331   GLCM textural features comprised four out of five features in the optimal design of the

332   classification scheme [39]. Classification of sacroiliitis based on feature extraction has been

333   demonstrated, based on 203 extracted image features [24]. The features maximum pixel value

334 and Haar wavelet from LH on level 2 were important to discriminate instances. The high -

335 intensity gray level in inflammation is a plausible cause for the impact of maximum pixel value

336 for classification of sacroiliitis. Quantitative textural analysis has also been suggested to detect

337 bone structure changes due to exercise [42, 43]. However, findings in these studies are not

338 directly comparable to textural features of BME.

339

340 There was no detectable effect of HIIT in the textural features from MR images (Table 5). This

341 is in accordance with the radiological evaluation and SPARCC scoring of the MR images.

342 However, four textural features were significantly decreased for the whole cohort (before

343 Bonferroni correction). These changes in values for textural features in voxels of BME towards

344 values of voxels in healthy bone marrow are similar to the reduction in DAS44 for both HIIT

345 and control groups (Table 1). Participants in both groups thus experienced a measurable

346 reduction in disease burden over the 11 weeks, as previously described [11]. This finding

347 indicates that textural features could be a more sensitive method to detect changes in BME than

348 methods based on visual inspection of images. DAS44 is a subjective measure of disease

349 burden, and the changes in textural features may be accidental. Further studies that investigate

350 the correlation of textural features with disease activity and include more patients with axial

351 PsA are necessary.

352

353 Limitations of this study include the small disease burden in the included patients and few

354 patients with manifested PsA in the spine. A future longitudinal study including treatment

355 follow-up, in a larger cohort with a larger disease burden, would further enable assessment of

356 the methods sensitivity to change, and the potential for classification based on textural features

357 It has previously been reported beneficial effects of HIIT in this patient cohort, with increased

358 VO_2max, reduced truncal fat, and less fatigue [11, 12]. Importantly, joint affection and pain

359   did not differ from the control group. The negative findings also in MR images strongly

360   indicate no structural changes. The evidence of HIIT being safe to conduct for patients with

361   PsA is thus stronger.

362

363   MR imaging showed that 11 weeks of HIIT in PsA patients led to no changes in spinal bone

364   marrow edema, supporting that HIIT is safe for these patients. This study indicate that texture

365   analysis of spinal MR images could be more sensitive to changes in BME than radiological

366   evaluation and SPARCC scoring.

References

368   1.   Wright V, Moll JMH. Psoriatic arthritis. In: Wright V, Moll JMH, editors.
369   Seronegative polyarthritis. Amsterdam: North Holland Publishing Co; 1976. p. 169-223.
370   2.   Hoff M, Gulati AM, Romundstad PR, Kavanaugh A, Haugeberg G. Prevalence and
371   incidence rates of psoriatic arthritis in central Norway: data from the Nord-Trondelag health
372   study (HUNT). Ann Rheum Dis. 2015;74(1):60-4. Epub 2013/08/22. doi:
373   10.1136/annrheumdis-2013-203862. PubMed PMID: 23962458.
374   3.   Ibrahim G, Waxman R, Helliwell PS. The prevalence of psoriatic arthritis in people
375   with psoriasis. Arthritis Rheum. 2009;61(10):1373-8. Epub 2009/10/01. doi:
376   10.1002/art.24608. PubMed PMID: 19790120.
377   4.   Baraliakos X, Coates LC, Braun J. The involvement of the spine in psoriatic arthritis.
378   Clin Exp Rheumatol. 2015;33(5 Suppl 93):S31-5. Epub 2015/10/17. PubMed PMID:
379   26471338.
380   5.   Sveaas SH, Smedslund G, Hagen KB, Dagfinrud H. Effect of cardiorespiratory and
381   strength exercises on disease activity in patients with inflammatory rheumatic diseases: a
382   systematic review and meta-analysis. Br J Sports Med. 2017;51(14):1065-72. Epub
383   2017/04/30. doi: 10.1136/bjsports-2016-097149. PubMed PMID: 28455366.
384   6.   Coates LC, Kavanaugh A, Mease PJ, Soriano ER, Laura Acosta-Felquer M,
385   Armstrong AW, et al. Group for Research and Assessment of Psoriasis and Psoriatic Arthritis
386   2015 Treatment Recommendations for Psoriatic Arthritis. Arthritis Rheumatol.
387   2016;68(5):1060-71. Epub 2016/01/11. doi: 10.1002/art.39573. PubMed PMID: 26749174.
388   7.   Sandstad J, Stensvold D, Hoff M, Nes BM, Arbo I, Bye A. The effects of high
389   intensity interval training in women with rheumatic disease: a pilot study. Eur J Appl Physiol.
390   2015;115(10):2081-9. Epub 2015/05/28. doi: 10.1007/s00421-015-3186-9. PubMed PMID:
391   26013051.
392   8.   Sveaas SH, Bilberg A, Berg IJ, Provan SA, Rollefstad S, Semb AG, et al. High
393   intensity exercise for 3 months reduces disease activity in axial spondyloarthritis (axSpA): a
394   multicentre randomised trial of 100 patients. Br J Sports Med. 2020;54(5):292-7. Epub
395   2019/02/13. doi: 10.1136/bjsports-2018-099943. PubMed PMID: 30745314.
396   9.   Schett G, Lories RJ, D'Agostino MA, Elewaut D, Kirkham B, Soriano ER, et al.
397   Enthesitis: from pathophysiology to treatment. Nat Rev Rheumatol. 2017;13(12):731-41.
398   Epub 2017/11/22. doi: 10.1038/nrrheum.2017.188. PubMed PMID: 29158573.
399   10.   Debusschere K, Cambre I, Gracey E, Elewaut D. Born to run: The paradox of
400   biomechanical force in spondyloarthritis from an evolutionary perspective. Best Pract Res
401   Clin Rheumatol. 2017;31(6):887-94. Epub 2018/12/05. doi: 10.1016/j.berh.2018.07.011.
402   PubMed PMID: 30509447.
403   11.   Thomsen RS, Nilsen TIL, Haugeberg G, Bye A, Kavanaugh A, Hoff M. Impact of
404   High-Intensity Interval Training on Disease Activity and Disease in Patients With Psoriatic
405   Arthritis: A Randomized Controlled Trial. Arthritis Care Res (Hoboken). 2019;71(4):530-7.
406   Epub 2018/06/09. doi: 10.1002/acr.23614. PubMed PMID: 29882634.
407   12.   Thomsen RS, Nilsen TIL, Haugeberg G, Bye A, Kavanaugh A, Hoff M. Effect of
408   high-intensity interval training on cardiovascular disease risk factors and body composition in
409   psoriatic arthritis: a randomised controlled trial. RMD Open. 2018;4(2):e000729. Epub
410   2018/11/08. doi: 10.1136/rmdopen-2018-000729. PubMed PMID: 30402265; PubMed
411   Central PMCID: PMCPMC6203095.
412   13.   Poggenborg RP, Sorensen IJ, Pedersen SJ, Ostergaard M. Magnetic resonance
413   imaging for diagnosing, monitoring and prognostication in psoriatic arthritis. Clin Exp
414   Rheumatol. 2015;33(5 Suppl 93):S66-9. Epub 2015/10/17. PubMed PMID: 26470678.

415   14.   Wiell C, Szkudlarek M, Hasselquist M, Moller JM, Vestergaard A, Norregaard J, et
416   al. Ultrasonography, magnetic resonance imaging, radiography, and clinical assessment of
417   inflammatory and destructive changes in fingers and toes of patients with psoriatic arthritis.
418   Arthritis Res Ther. 2007;9(6):R119. Epub 2007/11/16. doi: 10.1186/ar2327. PubMed PMID:
419   18001463; PubMed Central PMCID: PMCPMC2246238.
420   15.   Braun J, Baraliakos X, Golder W, Brandt J, Rudwaleit M, Listing J, et al. Magnetic
421   resonance imaging examinations of the spine in patients with ankylosing spondylitis, before
422   and after successful therapy with infliximab: evaluation of a new scoring system. Arthritis
423   Rheum. 2003;48(4):1126-36. Epub 2003/04/11. doi: 10.1002/art.10883. PubMed PMID:
424   12687557.
425   16.   Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Krishnananthan R, Stone M,
426   et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index
427   for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum.
428   2005;53(4):502-9. doi: 10.1002/art.21337 [doi].
429   17.   Fritz J, Henes JC, Thomas C, Clasen S, Fenchel M, Claussen CD, et al. Diagnostic
430   and interventional MRI of the sacroiliac joints using a 1.5-T open-bore magnet: a one-stop-
431   shopping approach. AJR Am J Roentgenol. 2008;191(6):1717-24. doi: 191/6/1717
432   [pii];10.2214/AJR.08.1075 [doi].
433   18.   Rudwaleit M, Jurik AG, Hermann KG, Landewe R, van der Heijde D, Baraliakos X,
434   et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of
435   axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Ann
436   Rheum Dis. 2009;68(10):1520-7. doi: ard.2009.110767 [pii];10.1136/ard.2009.110767 [doi].
437   19.   Tuceryan M, Jain AK. Texture analysis. In: Chen CH, Pau LF, Wang PSP, editors.
438   The Handbook of Pattern Recognition and Computer Vision. 2nd ed: World Scientific
439   Publishing Co.; 1998. p. 207-48.
440   20.   Shapiro LG, Stockman GC. Computer Vision. California: Prentice Hall; 2001. 580 p.
441   21.   Mayerhoefer ME, Breitenseher MJ, Kramer J, Aigner N, Norden C, Hofmann S.
442   STIR vs. T1-weighted fat-suppressed gadolinium-enhanced MRI of bone marrow edema of
443   the knee: computer-assisted quantitative comparison and influence of injected contrast media
444   volume and acquisition parameters. J Magn Reson Imaging. 2005;22(6):788-93. Epub
445   2005/11/05. doi: 10.1002/jmri.20439. PubMed PMID: 16270290.
446   22.   Heilmeier U, Wamba JM, Joseph GB, Darakananda K, Callan J, Neumann J, et al.
447   Baseline knee joint effusion and medial femoral bone marrow edema, in addition to MRI-
448   based T2 relaxation time and texture measurements of knee cartilage, can help predict
449   incident total knee arthroplasty 4-7 years later: data from the Osteoarthritis Initiative. Skeletal
450   Radiol. 2019;48(1):89-101. Epub 2018/07/02. doi: 10.1007/s00256-018-2995-4. PubMed
451   PMID: 29961091.
452   23.   Chuah TK, Van Reeth E, Sheah K, Poh CL. Texture analysis of bone marrow in knee
453   MRI for classification of subjects with bone marrow lesion - data from the Osteoarthritis
454   Initiative. Magn Reson Imaging. 2013;31(6):930-8. Epub 2013/02/26. doi:
455   10.1016/j.mri.2013.01.014. PubMed PMID: 23434436.
456   24.   Faleiros MC, Nogueira-Barbosa MH, Dalto VF, Junior JRF, Tenorio APM, Luppino-
457   Assad R, et al. Machine learning techniques for computer-aided classification of active
458   inflammatory sacroiliitis in magnetic resonance imaging. Advances in rheumatology
459   (London, England). 2020;60(1):25. PubMed PMID: Medline:32381053.
460   25.   Chuah TK, Poh CL, Sheah K. Quantitative texture analysis of MRI images for
461   detection of cartilage-related bone marrow edema. Conf Proc IEEE Eng Med Biol Soc.
462   2011;2011:5112-5. Epub 2012/01/19. doi: 10.1109/IEMBS.2011.6091266. PubMed PMID:
463   22255489.

464   26.   Mayerhoefer ME, Breitenseher MJ, Kramer J, Aigner N, Hofmann S, Materka A.
465   Texture analysis for tissue discrimination on T1-weighted MR images of the knee joint in a
466   multicenter study: Transferability of texture features and comparison of feature selection
467   methods and classifiers. J Magn Reson Imaging. 2005;22(5):674-80. Epub 2005/10/11. doi:
468   10.1002/jmri.20429. PubMed PMID: 16215966.
469   27.   Srikantan G, Lam SW, Srihari SN. Gradient-based contour encoding for character
470   recognition. Pattern Recogn. 1996;29(7):1147-60. doi: Doi 10.1016/0031-3203(95)00146-8.
471   PubMed PMID: WOS:A1996UU38800006.
472   28.   Chronaiou I, Thomsen RS, Huuse EM, Euceda LR, Pedersen SJ, Hoff M, et al.
473   Quantifying bone marrow inflammatory edema in the spine and sacroiliac joints with
474   thresholding. BMC Musculoskelet Disord. 2017;18(1):497. Epub 2017/11/29. doi:
475   10.1186/s12891-017-1861-1. PubMed PMID: 29179748; PubMed Central PMCID:
476   PMCPMC5704357.
477   29.   Lambert RG, Bakker PA, van der Heijde D, Weber U, Rudwaleit M, Hermann KG, et
478   al. Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by
479   the ASAS MRI working group. Ann Rheum Dis. 2016;75(11):1958-63. Epub 2016/01/16.
480   doi: 10.1136/annrheumdis-2015-208642. PubMed PMID: 26768408.
481   30.   Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK:
482   improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310-20. Epub
483   2010/04/10. doi: 10.1109/TMI.2010.2046908. PubMed PMID: 20378467; PubMed Central
484   PMCID: PMCPMC3071855.
485   31.   Perona P, Malik J. Scale-Space and Edge-Detection Using Anisotropic Diffusion. Ieee
486   T Pattern Anal. 1990;12(7):629-39. doi: Doi 10.1109/34.56205. PubMed PMID:
487   WOS:A1990DK89400002.
488   32.   Maksymowych WP, Lambert RG, Brown LS, Pangan AL. Defining the Minimally
489   Important Change for the SpondyloArthritis Research Consortium of Canada Spine and
490   Sacroiliac Joint Magnetic Resonance Imaging Indices for Ankylosing Spondylitis. J
491   Rheumatol. 2012;39(8):1666-74. Epub 2012/07/04. doi: 10.3899/jrheum.120131. PubMed
492   PMID: 22753648.
493   33.   de Winter J, de Hooge M, van de Sande M, de Jong H, van Hoeven L, de Koning A,
494   et al. Magnetic Resonance Imaging of the Sacroiliac Joints Indicating Sacroiliitis According
495   to the Assessment of SpondyloArthritis international Society Definition in Healthy
496   Individuals, Runners, and Women With Postpartum Back Pain. Arthritis Rheumatol.
497   2018;70(7):1042-8. Epub 2018/03/08. doi: 10.1002/art.40475. PubMed PMID: 29513924;
498   PubMed Central PMCID: PMCPMC6032910.
499   34.   Lukas C, Braun J, van der Heijde D, Hermann KG, Rudwaleit M, Ostergaard M, et al.
500   Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing
501   spondylitis: a multireader experiment. J Rheumatol. 2007;34(4):862-70. doi: 0315162X-34-
502   862 [pii].
503   35.   Maksymowych WP, Lambert RG, Ostergaard M, Pedersen SJ, Machado PM, Weber
504   U, et al. MRI lesions in the sacroiliac joints of patients with spondyloarthritis: an update of
505   definitions and validation by the ASAS MRI working group. Ann Rheum Dis.
506   2019;78(11):1550-8. Epub 2019/08/20. doi: 10.1136/annrheumdis-2019-215589. PubMed
507   PMID: 31422357.
508   36.   Hu Z, Xu M, Li Q, Lin Z, Liao Z, Cao S, et al. Adalimumab significantly reduces
509   inflammation and serum DKK-1 level but increases fatty deposition in lumbar spine in active
510   ankylosing spondylitis. Int J Rheum Dis. 2012;15(4):358-65. Epub 2012/08/18. doi:
511   10.1111/j.1756-185X.2012.01734.x. PubMed PMID: 22898215.
512   37.   Lambert RG, Salonen D, Rahman P, Inman RD, Wong RL, Einstein SG, et al.
513   Adalimumab significantly reduces both spinal and sacroiliac joint inflammation in patients

514 with ankylosing spondylitis: a multicenter, randomized, double-blind, placebo-controlled
515 study. Arthritis Rheum. 2007;56(12):4005-14. Epub 2007/12/01. doi: 10.1002/art.23044.
516 PubMed PMID: 18050198.
517 38.   Baraliakos X, Richter A, Feldmann D, Ott A, Buelow R, Schmidt CO, et al. Which
518 factors are associated with bone marrow oedema suspicious of axial spondyloarthritis as
519 detected by MRI in the sacroiliac joints and the spine in the general population? Ann Rheum
520 Dis. 2020. Epub 2020/11/27. doi: 10.1136/annrheumdis-2020-218669. PubMed PMID:
521 33239274.
522 39.   Romanos O, Solomou E, Georgiadis P, Kardamakis D, Siablis D. Magnetic resonance
523 imaging and image analysis of post - radiation changes of bone marrow in patients with
524 skeletal metastases. J BUON. 2013;18(3):788-94. Epub 2013/09/26. PubMed PMID:
525 24065500.
526 40.   MacKay JW, Murray PJ, Low SB, Kasmai B, Johnson G, Donell ST, et al.
527 Quantitative analysis of tibial subchondral bone: Texture analysis outperforms conventional
528 trabecular microarchitecture analysis. J Magn Reson Imaging. 2016;43(5):1159-70. Epub
529 2015/11/26. doi: 10.1002/jmri.25088. PubMed PMID: 26606692.
530 41.   MacKay JW, Murray PJ, Kasmai B, Johnson G, Donell ST, Toms AP. MRI texture
531 analysis of subchondral bone at the tibial plateau. Eur Radiol. 2016;26(9):3034-45. Epub
532 2015/12/19. doi: 10.1007/s00330-015-4142-0. PubMed PMID: 26679180.
533 42.   Sikio M, Harrison LC, Nikander R, Ryymin P, Dastidar P, Eskola HJ, et al. Influence
534 of exercise loading on magnetic resonance image texture of thigh soft tissues. Clin Physiol
535 Funct Imaging. 2014;34(5):370-6. Epub 2013/11/22. doi: 10.1111/cpf.12107. PubMed
536 PMID: 24256409.
537 43.   Harrison LC, Nikander R, Sikio M, Luukkaala T, Helminen MT, Ryymin P, et al.
538 MRI texture analysis of femoral neck: Detection of exercise load-associated differences in
539 trabecular bone. J Magn Reson Imaging. 2011;34(6):1359-66. Epub 2011/09/29. doi:
540 10.1002/jmri.22751. PubMed PMID: 21954096.
541

542

543 # Supporting information

544 **S1 Table. Mathematical description of extracted image features.** Description of the features

545 that were extracted and quantified from the MR images. Gradient features $g_1 - g_{10}$ were

546 computed from the two-dimensional directional gradients for x-axis (G_x) and y-axis (G_y);

547 $L1_{norm} = |G_x| + |G_y|$, and $L2_{norm} = \sqrt{G_x^2 + G_y^2}$.

548

549   Figure 1
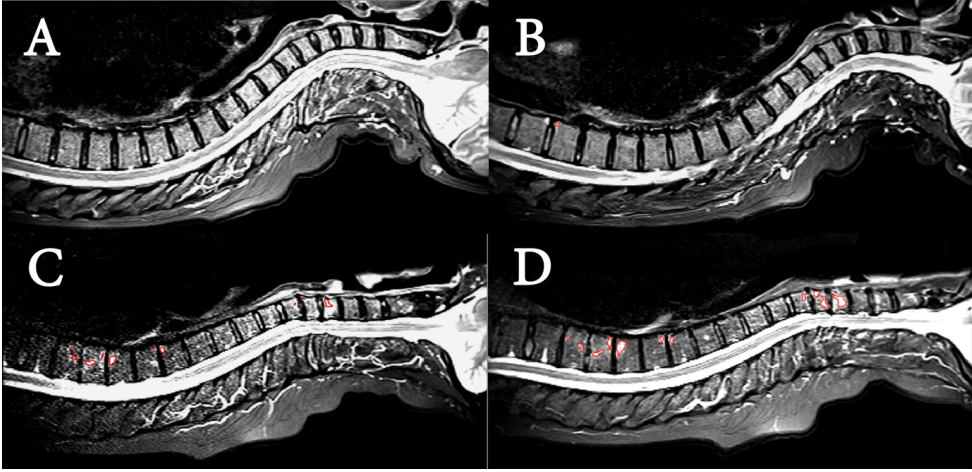
550

Paper III

Original Article

Feasibility of contrast-enhanced MRI derived textural features to predict overall survival in locally advanced breast cancer

Ioanna Chronaiou[1], Guro Fanneløb Giskeødegård[1],
Pål Erik Goa[2], Jose Teruel[3], Roja Hedayati[4,5],
Steinar Lundgren[4,5], Else Marie Huuse[6], Martin D. Pickles[7],
Peter Gibbs[8], Beathe Sitter[1] and Tone Frost Bathen[1,6]

Abstract
Background: The prognosis for women with locally advanced breast cancer (LABC) is poor and there is a need for better treatment stratification. Gray-level co-occurrence matrix (GLCM) texture analysis of magnetic resonance (MR) images has been shown to predict pathological response and could become useful in stratifying patients to more targeted treatments.
Purpose: To evaluate the ability of GLCM textural features obtained before neoadjuvant chemotherapy to predict overall survival (OS) seven years after diagnosis of patients with LABC.
Material and Methods: This retrospective study includes data from 55 patients with LABC. GLCM textural features were extracted from segmented tumors in pre-treatment dynamic contrast-enhanced 3-T MR images. Prediction of OS by GLCM textural features was assessed and compared to predictions using traditional clinical variables.
Results: Linear mixed-effect models showed significant differences in five GLCM features (f_1, f_2, f_5, f_{10}, f_{11}) between survivors and non-survivors. Using discriminant analysis for prediction of survival, GLCM features from 2 min post-contrast images achieved a classification accuracy of 73% ($P < 0.001$), whereas traditional prognostic factors resulted in a classification accuracy of 67% ($P = 0.005$). Using a combination of both yielded the highest classification accuracy (78%, $P < 0.001$). Median values for features f_1, f_2, f_{10}, and f_{11} provided significantly different survival curves in Kaplan–Meier analysis.
Conclusion: This study shows a clear association between textural features from post-contrast images obtained before neoadjuvant chemotherapy and OS seven years after diagnosis. Further studies in larger cohorts should be undertaken to investigate how this prognostic information can be used to benefit treatment stratification.

[1]Department of Circulation and Medical Imaging, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
[2]Department of Physics, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
[3]Department of Radiation Oncology, NYU Langone Health, New York, NY, USA
[4]Cancer clinic, St. Olavs University Hospital, Trondheim, Norway
[5]Department of Clinical and Molecular Medicine, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
[6]Clinic of Radiology and Nuclear Medicine, St. Olavs University Hospital, Trondheim, Norway

[7]Radiology Department, Hull University Teaching Hospitals NHS Trust, Hull, UK
[8]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Corresponding author:
Tone Frost Bathen, Department of Circulation and Medical Imaging, NTNU – Norwegian University of Science and Technology, Trondheim, Norway.
Email: tone.f.bathen@ntnu.no

Introduction

Breast cancer is currently expected to constitute one-third of all new cancer diagnoses in women (1). With emergence of better and more targeted treatments, mortality rates for female breast cancer has dropped. However, prognosis for patients diagnosed with locally advanced breast cancer (LABC) is still poor (2). Current treatment for patients with LABC is multi-modal and includes chemotherapy, irradiation, endocrine therapy, and surgical resection. Neoadjuvant chemotherapy (NAC) is often offered before surgery to downstage tumors and improve operability, and to eradicate distant micro-metastases (3). NAC has been linked with improved long-term prognosis in patients who achieve pathological complete response (4,5), but not all patients respond to treatment. Identifying prognostic factors, even before initiating NAC, would be beneficial for improved treatment stratification and overall survival (OS) in patients with LABC.

Dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) involves the acquisition of a baseline image and a series of images after the intravenous administration of a contrast agent. DCE-MRI allows three-dimensional (3D) visualization of angiogenic properties in breast cancer, making it a powerful tool for detecting changes before morphological alterations (6). Pre-treatment MRI of the breast provides characteristics of the whole lesion and holds more comprehensive information on heterogeneity than biopsies, which represent only a small proportion of the tumor. DCE-MRI derived pharmacokinetic model parameters have been used in evaluating and predicting early response to NAC (7–12) as well as OS for patients treated with NAC (13). Another approach is gray-level co-occurrence matrix (GLCM) texture analysis, which quantifies spatial variations in gray-level intensity and provides information about intuitive qualities of the images, such as sharpness and homogeneity. Texture analysis of DCE-MRI has been used in breast cancer staging (14–19), for prediction of invasive disease (20) and monitoring response to treatment (21–25), while a recent study (26) examined the potential to predict breast cancer survival using textural features from pre-treatment DCE-MRI along with shape features and traditional survival factors.

The main aim of this study was to assess the feasibility of GLCM features for prediction of seven-year OS in patients with LABC, using the whole DCE time-series.

Material and Methods

Patients and treatment

This retrospective study includes 56 Caucasian women with LABC treated with NAC, with available status for OS at least seven years after diagnosis. The study was approved by the Mid-Norway Regional Committee for Medical and Health Research Ethics (2009/112) and written informed consent was obtained from all patients. Patients were treated at St. Olav's University Hospital, Trondheim, Norway in the period 2007–2010 according to national guidelines (24). In brief, patients received four cycles of FEC, followed by four cycles of docetaxel, every three weeks. After the fourth cycle of FEC, the response to NAC was evaluated; if a decrease in the tumor longest diameter ≥80% was achieved, the cycles of docetaxel were either cancelled or replaced by two cycles of FEC. After the last cycle of NAC, patients were appointed for mastectomy and axillary lymph node dissection with postoperative treatment according to national guidelines (24). Histopathological analysis of the resected breast mass and axillary nodes provided pathological response status. The hormone receptor status of the tumors were determined from the diagnostic biopsy obtained before onset of NAC, and classified as estrogen (ER)-positive or progesterone (PgR)-positive if ≥10% of the cells stained positive. Human epidermal growth factor receptor 2 (HER-2) status was defined as positive by immunohistochemistry (score 3+) or by in situ hybridization (gene amplification ratio >2) (24). We followed REporting recommendations for tumor MARKer (REMARK) prognostic criteria (27).

Imaging acquisition and processing

Patients were imaged before NAC on a 3-T MR scanner (Siemens Tim Trio, Erlangen, Germany, software platform VB13A and VB15A) using a dedicated four-channel bilateral breast coil. The MRI acquisition and texture analysis are previously described (24). In brief; T1-weighted DCE-MR images were acquired using a 3D radiofrequency-spoiled gradient echo sequence. One baseline image was acquired before administration of contrast (0.1 mmol/kg) (Omniscan, GE Healthcare, Norway [47 patients]; Magnevist, Bayer Healthcare Pharmaceuticals, USA [9 patients]). Seven post-contrast images were acquired with a temporal resolution of 1 min. The imaging protocol changed after scanning 17 patients due to scanner software platform upgrade, larger flip angle (from 6° to 10°), and shorter repetition time (from 3.50 ms to 3.22 ms). All images were acquired with an in-plane resolution of 1.1 mm^2 and slice thickness in the range of 1.1–1.5 mm. Images were corrected for motion artefacts (FSL package, Oxford FMRIB Centre, University of Oxford, Oxford, UK). The tumors were manually segmented on motion-corrected subtraction images. Texture analysis was performed on the segmented tumors from all the slices
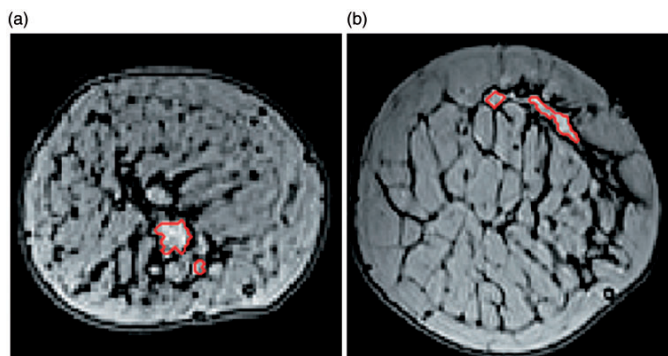
Fig. 1. MR image examples. Two minutes post-contrast dynamic contrast-enhanced MR images of (a) a survivor and (b) a non-survivor of LABC. The tumor region is outlined in red. While the segmentation was performed on the subtracted images, the figure shows the tumor volume in non-subtracted 2 min post-contrast images for better visualization.

containing tumors, from the baseline and post-contrast images, using two-dimensional GLCMs. Textural features from all slices were averaged, resulting in 16 GLCM features (f_1 through f_{16}) per patient per image. Processing of the images is described in detail by Teruel et al NMR Biomed (24). Representative MR images are shown in Fig. 1.

Statistical analysis

Linear mixed-effects models (LMMs) were used to assess time-related changes in individual GLCM features related to OS, considering all post-contrast time points simultaneously. LMMs (28) were built by random-intercept models (R 3.1.1, function lme; 'nlme' package (29)) employing restricted maximum likelihood. Log-transformed GLCM features (individually for f_1–f_{16}) from post-contrast images were used as response variables, while patient ID was modelled as a random effect. Fixed effects included survival status, log-transformed pre-contrast GLCM features, MR protocol used (1 or 2), and time-point (minutes post-contrast). Multiple testing correction was performed using the Benjamini–Hochberg method.

Multivariate analysis was performed by principal component analysis (PCA) and orthogonalized partial least squares discriminant analysis (OPLSDA) (PLS toolbox version 8.2.1, Eigenvector Research, Inc., Seattle, WA, USA) in MATLAB (MathWorks, Natick, MA, USA). PCA of autoscaled data was used to assess natural clusters of the data and to search for possible outliers. The benefit of applying OPLSDA is its capability to handle datasets with large numbers of variables compared to samples and datasets where variables are correlated (collinearity), typical characteristics of variables derived from texture

analysis. For discriminating survivors from non-survivors, separate OPLSDA models were built using GLCM features from the different time-points as input. Models were also built based on clinical prognostic factors alone and on the combination of clinical prognostic factors and GLCM features from 2-min post-contrast images (the time-point which was found to best discriminate survivors/non-survivors by OPLSDA based on GLCM features alone). Clinical prognostic factors included were age at diagnosis, largest tumor diameter (measured by calipers before treatment), TNM stage, ER and PgR receptor status, HER2 status, type (invasive ductal carcinoma, invasive lobular carcinoma, or other), and histological grade. For all models, the response variable was seven-year survival status. GLCM features and clinical prognostic factors were autoscaled before modelling. Models were validated by 10-fold cross-validation, where the dataset was split into 10 random subsets that were each kept out of model-building and used for validation. The whole procedure was repeated with 20 iterations, with average classification results presented. The significance of the resulting model was evaluated using permutation testing with 1000 permutations; models with P values (p_{perm}) < 0.05 were considered significant. OPLSDA loading plots were colored according to the variable's importance in the projection scores (VIP), reflecting the variable's influence on the classification. T-tests were used to assess whether classification accuracies of OPLSDA models built using GLCM features only, clinical variables only, and a combination of both were significantly different, by comparing the distribution of classification accuracies from the 20 iterations.

The mean values of the individual GLCM features at the time-point that best discriminated between survivors and non-survivors (2 min post-contrast) were

calculated. Differences in GLCM features between survivors/non-survivors at 2-min post-contrast were assessed by multiple linear regressions, correcting for the use of two acquisition sequences for DCE-MRI (MATLAB, MathWorks, Natick, MA, USA). GLCM features were log transformed before analysis to conform to normality.

Kaplan–Meier analysis (MATLAB, MathWorks, Natick, MA, USA) was used to depict differences in OS among patients with GLCM values above or below median derived from 2 min post-contrast images. GLCM features tested were significant in LMM analysis and/or important in OPLSDA analysis (VIP score->1.5); angular second moment (f_1), contrast (f_2), correlation (f_3), inverse difference moment (f_5), entropy (f_9), difference variance (f_{10}), difference entropy (f_{11}), and information measure of correlation 1 (f_{12}). Log-rank tests were performed to assess if the survival curves were significantly different.

To assess correlations between GLCM features, Pearson's correlation coefficients were calculated. A heat-map was established to visualize the correlations.

Results

Patient characteristics

One patient with multiple small and diffuse lesions was an outlier with high residual variation in PCA of GLCM features was excluded (Suppl. Fig. 1), resulting in a final cohort of 55 patients. Of these, 38 patients were survivors, while 17 patients were deceased within seven years and classified as non-survivors. Table 1 presents clinical characteristics for the patient cohort. Median follow-up time was 106 months for survivors and 41 months for non-survivors.

Characteristics of textural features

Mean values and SD from 2 min post-contrast are shown in Suppl. Table 1. A heat-map visualizing the Pearson's correlation coefficients between the individual GLCM features at 2 min post-contrast is shown in Fig. 2. A strong and positive correlation was observed of GLCM textural features f1 and f5, and of f2, f9, f10, and f11. Simultaneously, f1 and f5 were negatively correlated with f2, f9, f10, and f11.

Differences in GLCM values according to survival status

Table 2 shows the calculated *P* values from LMM analysis assessing differences in GLCM values between survivors and non-survivors. LMMs included GLCM textural features from all time-points simultaneously

for assessment of significant differences using the whole DCE time-series. The GLCM features angular second moment (f_1), contrast (f_2), inverse difference moment (f_5), difference variance (f_{10}), and difference entropy (f_{11}) were significantly different between survivors and non-survivors before but not after multiple testing correction. Additionally, entropy (f_9) and cluster shade (f_{15}) approached significance. The change in the MR protocol was found to significantly impact two of the textural features: variance (f_4) and cluster shade (f_{15}).

Textural features show added value to clinical prognostic factors in predicting overall survival

Classification accuracies for survival based on multivariate analysis of GLCM textural features from different time-points post-contrast are shown in Fig. 3. Textural features provided classification accuracies in the range of 60–73%, with the highest accuracy (73.1%) obtained 2 min post-contrast (sensitivity 65.3%, specificity 80.8%, $p_{perm} < 0.001$). Scores and loadings from the OPLSDA model for predicting survival status from 2 min post-contrast images are shown in Fig. 4. Higher levels of features representing correlation (f_3) and inverse difference moment (f_5), and lower levels of features representing contrast (f_2), entropy (f_9), difference variance (f_{10}), difference entropy (f_{11}), and information measure of correlation 1 (f_{12}) from 2 min post-contrast in non-survivors were the most important differentiators. Clinical parameters provided a classification accuracy for prediction of survival of 67.2% (sensitivity 82.8%, specificity 51.7%, $p_{perm} = 0.005$). The combination of clinical prognostic factors and GLCM features from 2 min post-contrast yielded a significantly higher ($P < 0.001$) classification accuracy for survival compared to clinical parameters alone (accuracy 77.8%, sensitivity 79.0%, specificity 76.7%, $p_{perm} < 0.001$).

Kaplan–Meier analysis shows correlation between GLCM textural features and survival

Survival curves for f_1, f_2, f_9, f_{10}, and f_{11} were significantly different when dividing the patients in two groups using median values for GLCM textural features from 2 min post-contrast images as cut-off (*P* values 0.015, 0.001, 0.041, 0.007, and 0.0015, respectively). Survival curves for f1, f2, f10, and f11 remained significant after multiple testing corrections (q values <0.05; Fig. 5).

Discussion

In this paper, we have demonstrated the possibility to assess patient prognosis from pre-treatment DCE-MRI

Table 1. Description of the patient cohort: clinical characteristics before administration of NAC for survivors (n = 38) and non-survivors (n = 17) of LABC.

	Survivors (n = 38)	Non-survivors (n = 17)	P
Mean age (range) at diagnosis (years)	56 (35–82)	52 (35–73)	0.239*
Largest tumor dimension (mm)	62 (30–100)	70 (30–100)	**0.010***
Clinical stage[†]			**<0.001[‡]**
IIB (T3 N0 M0)	17 (45)	5 (30)	
IIIA (T2–3 N1–2 M0)	12 (32)	2 (11)	
IIIB (T4a–d N0–2 M0)	9 (23)	3 (18)	
IIIC (T1–2 N3 M0)	0	1 (6)	
IV (T3–4c N1–2 M1)	0	6 (35)	
T			0.069[‡]
2	1 (3)	1 (6)	
3	28 (74)	10 (59)	
4	9 (23)	6 (35)	
N			**0.008[‡]**
0	21 (55)	6 (35)	
1	13 (34)	6 (35)	
2	4 (11)	4 (24)	
3	0	1 (6)	
M			**<0.001[‡]**
0	38 (100)	11 (65)	
1	0	6 (35)	
Receptor status			
ER+	25 (66)	13 (76)	0.749[‡]
PgR+	17 (45)	7 (41)	0.777[‡]
HER2+	14 (37)	3 (18)	0.209[‡]
Unknown	1 (3)	0	
Histopathological type of breast cancer			0.784[‡]
IDC	35 (92)	15 (88)	
ILC	2 (5)	1 (6)	
Mucinous carcinoma	1 (3)	1 (6)	
Grade of breast cancer			0.127[‡]
Grade 1	3 (8)	1 (6)	
Grade 2	19 (50)	4 (23)	
Grade 3	16 (42)	12 (71)	
Response to treatment			0.543[‡]
Stable disease	13 (34)	5 (30)	
Partial response	12 (32)	8 (47)	
Complete response	13 (34)	4 (23)	
Presence of lymph node metastasis	17 (45)	11 (65)	0.245[‡]

Values are given as n (%) unless otherwise specified.

Significant P values (p≤0.01) are given in bold.

*P value calculated using independent t-test for two variables.

[†]T2, tumor is >2 cm but <5 cm across; T3, tumor is >5 cm across; T4, tumor has either spread to the skin, the chest wall or both, or inflammatory carcinoma is present; N0, no nearby lymph node metastasis; N1–2, metastasis to movable/fixed ipsilateral axillary nodes; N3, metastasis to ipsilateral internal mammary lymph nodes; M0, no distant metastasis; M1, distant metastasis is present.

[‡]P value calculated using Fisher's exact test.

ER+, estrogen receptor negative; HER2+, herceptin receptor positive; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; LABC, locally advanced breast cancer; NAC, neoadjuvant chemotherapy; PgR+, progesterone receptor positive.

using GLCM textural features. Five out of 16 GLCM textural features were significantly different (before multiple testing) between survivors and non-survivors at seven years of follow-up using LMMs; angular second moment (f_1), contrast (f_2), inverse difference moment (f_5), difference variance (f_{10}), and difference entropy (f_{11}), while entropy (f_9) approached significance. Importantly, OPLSDA provided significant classification of survival status from 2 min post-contrast images, with the same GLCM textural features in addition to correlation (f_3) and information measure of correlation 1 (f_{12}) being the most important
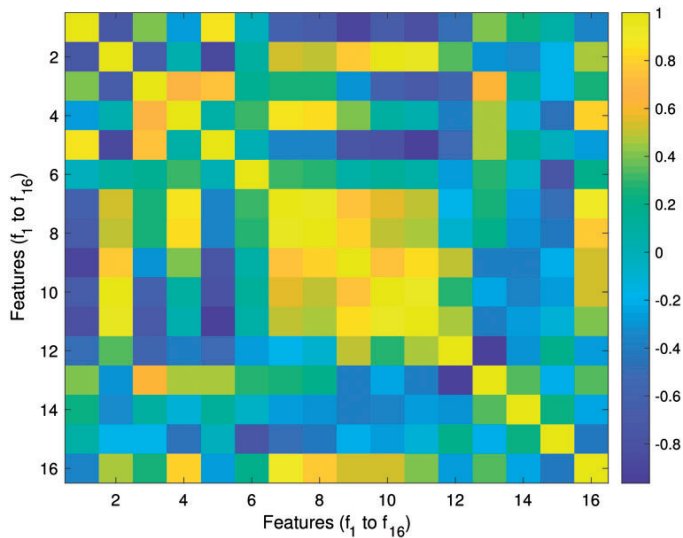
Fig. 2. Heat-map of correlations between textural features. The heat-map based on Pearson correlations between the textural features illustrates correlations, ranging from strong negative (-1) to strong positive ($+1$), between GLCM features f1–f16 obtained from 2 min post-contrast images.

Table 2. GLCM textural features with significantly different levels between survivors and non-survivors.

Feature	P value
Angular second moment (f_1)	**0.008**
Contrast (f_2)	**0.034**
Correlation (f_3)	0.308
Variance (f_4)	0.787*
Inverse difference moment (f_5)	**0.008**
Sum average (f_6)	0.122
Sum variance (f_7)	0.470
Sum entropy (f_8)	0.366
Entropy (f_9)	0.050
Difference variance (f_{10})	**0.024**
Difference entropy (f_{11})	**0.018**
Information measure of correlation 1 (f_{12})	0.347
Information measure of correlation 2 (f_{13})	0.777
Maximal correlation coefficient (f_{14})	0.358
Cluster shade (f_{15})	0.087*
Cluster prominence (f_{16})	0.477

Linear mixed-effects model *P* values for the relationship of GLCM features from all time-points to survival in patients with LABC are shown. *P* values < 0.05 are presented in bold. No features were significant after Benjamini–Hochberg correction for multiple testing (p<0.05).
*Feature affected by change in MR protocol.
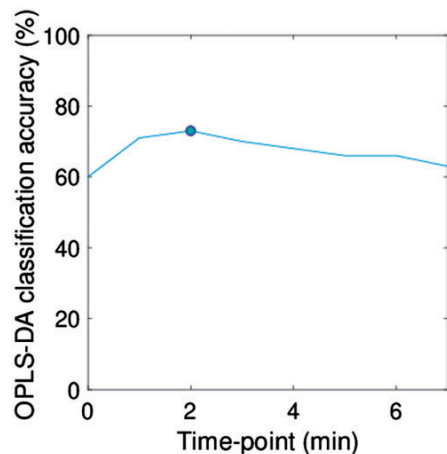GLCM, gray-level co-occurrence matrix; LABC, locally advanced breast cancer; MR, magnetic resonance.

Fig. 3. Classification accuracy for the prediction of survival using orthogonalized partial least squares discriminant analysis at each timepoint. Classification accuracy by partial least squares discriminant analysis using all GLCM features obtained from post-contrast images of patients with LABC (n = 55) at each time-point (0–7 min post-contrast). The circle marks the maximum classification accuracy, which is at 2 min post-contrast.
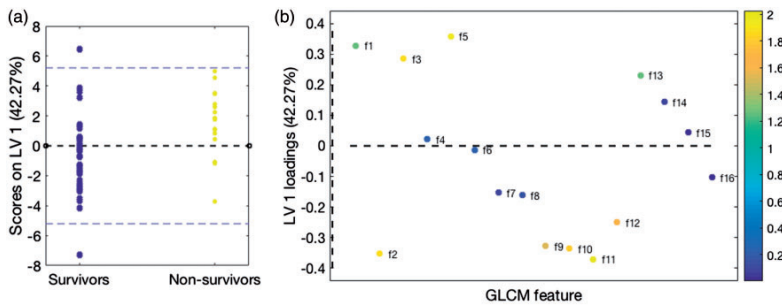
Fig. 4. Scores and loadings from orthogonalized partial least squares discriminant analysis (OPLSDA) for OS. Scores (a) and loadings (b) on latent variable (LV) 1 from OPLSDA predicting survival using as input all GLCM features obtained from 2 min post-contrast images of patients with LABC (n = 55). In the loadings, variables are colored according to their importance for projection (VIP) score, reflecting the variable's importance for the classification. For this cohort, OPLSDA classified survivors and non-survivors with an accuracy of 73.1% ($p_{perm} <$ 0.001).
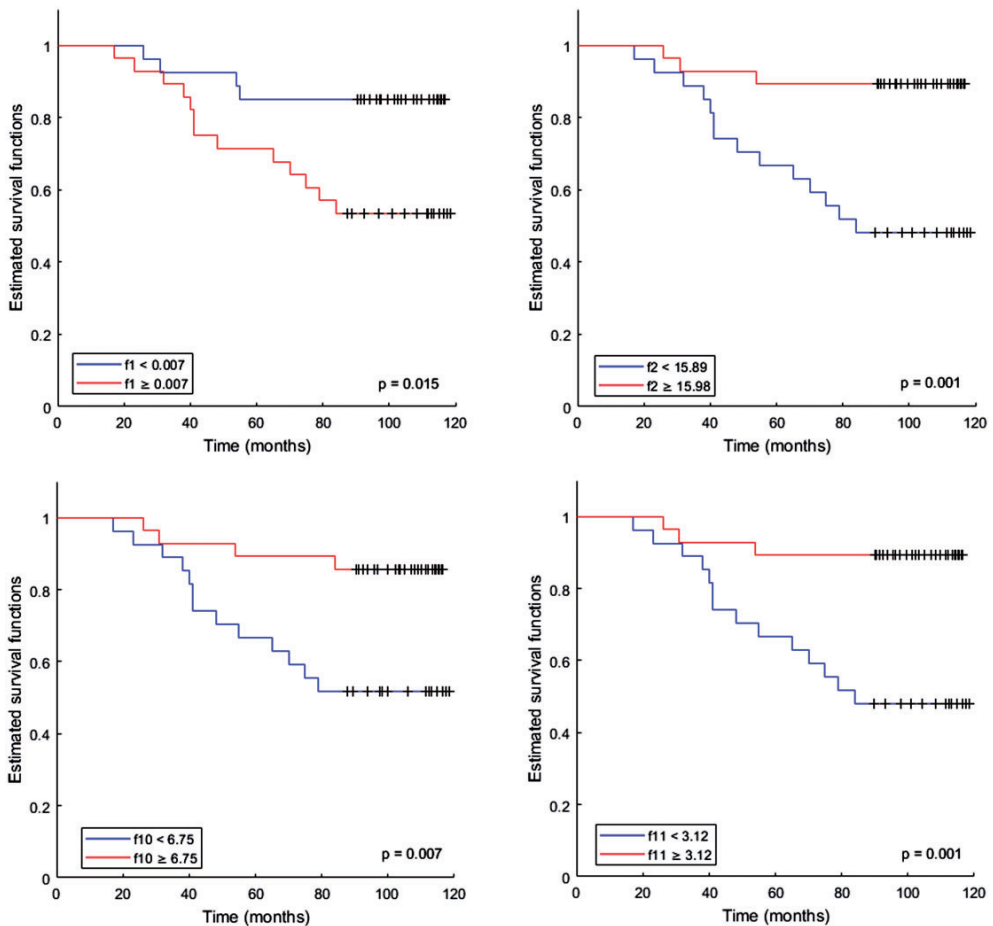
Fig. 5. Kaplan–Meier survival curves for patients with high and low values of selected GLCM features. The recurrence-free proportions are plotted against OS time for GLCM features dichotomized to above or below median values. Only survival curves for features that remained significant after multiple testing corrections are shown (f1, f2, f10, and f11).

variables. GLCM textural features predicted OS by OPLSDA with accuracy comparable to clinical prognostic factors. Interestingly, while classification using clinical prognostic factors provided high sensitivity, GLCM textural features provided classification with high specificity. Combining GLCM features and clinical parameters provided the most accurate classification of survival, with both sensitivity and specificity approaching 80%. Grouping patients based on high and low values of textural features resulted in significant Kaplan–Meier survival curves for four of the same features, even when correcting for multiple testing.

Describing the underlying physiology that GLCM textural features represent is challenging. The highly positively correlated textural features f_1 and f_5 (Fig. 2) are, in general, associated with uniformity of gray levels, with higher levels indicating a more homogeneous signal distribution within the tumor. In our cohort, non-survivors were associated with higher uniformity. The textural features f_9, f_{10}, and f_{11}, also positively correlated, are associated with heterogeneity, and the non-survivors were associated with lower values of these features. In terms of physiology, this could point towards a fast, high and uniform uptake of contrast in the more aggressive tumors, resulting in more uniform signal distribution.

A few studies have assessed prognostic outcomes in breast cancer using pre-treatment DCE MRI data (13,26,30,31). A strength of the current study is the seven-year follow-up, which allows the assessment of long-term outcomes. Our results are in accordance with the study of Kim et al. (31), who found that patients with breast cancers that appeared with less heterogeneity (lower entropy) on contrast-enhanced T1-weighted subtraction images exhibited poorer outcomes. Further, Pickles et al. (26) evaluated the prognostic values of pre-treatment DCE-MRI-based parameters for breast cancer patients, including the same textural features as explored in our study. GLCM textural features representing heterogeneity of contrast enhancement; sum variance (f_7) and sum entropy (f_8), and symmetry of contrast enhancement; cluster shade (f_{15}) and cluster prominence (f_{16}) were consistently higher in patients with shorter survival (26). None of these features appeared as significantly different between the patient groups in our study; however, cluster shade (f_{15}) was found to be significantly affected by the change in the MR protocol, which could mask differences. Interestingly, one of the features with most significant association to survival in our cohort, f_{10} (also related to heterogeneity), was previously shown to correlate with treatment response in the same dataset, with lower values reported for the non-responders (24). An association between f_{10} and treatment response was also detected in an independent patient cohort (21), but then with higher values for the non-responders reported.

Several factors related to acquisition parameters and image processing can affect the robustness, reproducibility, and reliability of textural features (32–34). This may explain why studies report different textural features as important for similar classification purposes. Statistical analysis (LMM) showed differences by MR protocol in two of the features (Table 2), but these had limited importance in the classification of survival. This is supported by previous research showing that textural features are increasingly sensitive to acquisition parameters variation with increasing spatial resolution, but that the effect on pattern discrimination still is minimal provided sufficiently high spatial resolution (35). However, software platform upgrades are natural and necessary events in maintaining clinical scan protocols. Identifying robust features less affected by such changes could improve clinical translation of texture analysis. A different contrast agent was used for nine of the patients (24). However, differences in image contrast due to different contrast agents have been found to be negligible (36), and we assumed this difference had no significant impact on the texture feature values. Before the extraction of textural features, we performed histogram equalization to minimize the effects of the differences in acquisition protocol. A previous study by Sikiö et al. (37) has shown some textural features to depend on the size of the segmented area used for the analysis. Larger tumor size is associated with a poorer prognosis in breast cancer and the largest tumor diameter was significantly higher in the non-survivors of our cohort. However, none of the textural features correlated to tumor size in our study (results not shown).

Our study is based on a small patient cohort ($n = 55$). A larger patient cohort would allow to construct more robust models, including validation using independent data. However, this study still demonstrates the feasibility of predicting patient prognosis in LABC using texture analysis of baseline DCE-MRI, before all clinical prognostic factors can be determined. Our results should be verified using larger cohorts and reproducibility of the textural features should be assessed, preferably using data from a multicenter cohort to address the impact from differences in acquisition parameters.

A strength of our study is that the use of LMMs allowed information from the whole enhancement time-curve to be included in the analysis. However, the most significant associations for textural features and survival were still observed in the 2 min post-contrast images (Fig. 3), as also reported when associating textural features to treatment response (21,24) demonstrating the clinical importance of using the

textural features from the time-point where the contrast uptake is at its peak. In addition, previous studies pinpoint the potential of the enhancement curve type to predict survival (30), where tumors with a high proportion of the aggressive curve type III voxels (38) were associated with a poorer prognosis. The fast and high contrast uptake in addition to the rapid wash out is the major features of this curve type.

In conclusion, textural features derived from contrast-enhanced MRI before NAC in breast cancer patients are associated with a seven-year survival outcome. Furthermore, the textural features show added value to the clinical prognostic factors in the prediction of long-term survival. This is in accordance with previous studies, where textural features have been associated with long-term outcome and treatment response, and show a valuable potential to obtain more personalized treatment for breast cancer patients. However, our results need validation in a larger, independent cohort.

ORCID iD

Tone Frost Bathen (iD) https://orcid.org/0000-0002-8582-6965

Supplemental material

Supplemental material for this article is available online.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7–34.
2. Klein J, Tran W, Watkins E, et al. Locally advanced breast cancer treated with neoadjuvant chemotherapy and adjuvant radiotherapy: a retrospective cohort analysis. BMC Cancer 2019;19:306.
3. Gianni L, Eiermann W, Semiglazov V, et al. Neoadjuvant and adjuvant trastuzumab in patients with HER2-positive locally advanced breast cancer (NOAH): follow-up of a randomised controlled superiority trial with a parallel HER2-negative cohort. Lancet Oncol 2014;15:640–647.
4. Wang M, Hou L, Chen M, et al. Neoadjuvant chemotherapy creates surgery opportunities for inoperable locally advanced breast cancer. Sci Rep 2017;7:44673.
5. Kong X, Moran MS, Zhang N, et al. Meta-analysis confirms achieving pathological complete response after neoadjuvant chemotherapy predicts favourable prognosis for breast cancer patients. Eur J Cancer 2011;47:2084–2090.
6. Turnbull LW. Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. NMR Biomed 2009;22:28–39.
7. Drew PJ, Kerin MJ, Mahapatra T, et al. Evaluation of response to neoadjuvant chemoradiotherapy for locally advanced breast cancer with dynamic contrast-enhanced MRI of the breast. Eur J Surg Oncol 2001;27:617–720.
8. Cheung YC, Chen SC, Su MY, et al. Monitoring the size and response of locally advanced breast cancers to neoadjuvant chemotherapy (weekly paclitaxel and epirubicin) with serial enhanced MRI. Breast Cancer Res Treat 2003;78:51–58.
9. Hylton NM, Blume JD, Bernreuter WK, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy–results from ACRIN 6657/I-SPY TRIAL. Radiology 2012; 263:663–672.
10. Li X, Arlinghaus LR, Ayers GD, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: pilot study findings. Magn Reson Med 2014;71:1592–1602.
11. Drisis S, Metens T, Ignatiadis M, et al. Quantitative DCE-MRI for prediction of pathological complete response following neoadjuvant treatment for locally advanced breast cancer: the impact of breast cancer subtypes on the diagnostic accuracy. Eur Radiol 2016; 26:1474–1484.
12. Drew PJ, Kerin MJ, Mahapatra T, et al. Evaluation of response to neoadjuvant chemoradiotherapy for locally advanced breast cancer with dynamic contrast-enhanced MRI of the breast. Eur J Surg Oncol 2001;27:617–620.
13. Heldahl MG, Bathen TF, Rydland J, et al. Prognostic value of pretreatment dynamic contrast-enhanced MR imaging in breast cancer patients receiving neoadjuvant chemotherapy: overall survival predicted from combined time course and volume analysis. Acta Radiol 2010;51:604–612.
14. Cai H, Peng Y, Ou C, et al. Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach. PLoS One 2014;9: e87387.
15. Chen W, Giger ML, Li H, et al. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. Magn Reson Med 2007;58:562–571.
16. Karahaliou A, Vassiou K, Arikidis NS, et al. Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. Br J Radiol 2010;83:296–309.
17. Milenkovic J, Hertl K, Kosir A, et al. Characterization of spatiotemporal changes for the classification of dynamic

contrast-enhanced magnetic-resonance breast lesions. Artif Intell Med 2013;58:101–114.

18. Nagarajan MB, Huber MB, Schlossbauer T, et al. Classifying small lesions on breast MRI through dynamic enhancement pattern characterization. Mach Learn Med Imaging 2011;7009:352–359.

19. Nagarajan MB, Huber MB, Schlossbauer T, et al. Classification of small lesions in dynamic breast MRI: eliminating the need for precise lesion segmentation through spatio-temporal analysis of contrast enhancement. Mach Vis Appl 2013;24:1371–1381.

20. Harowicz MR, Saha A, Grimm LJ, et al. Can algorithmically assessed MRI features predict which patients with a preoperative diagnosis of ductal carcinoma in situ are upstaged to invasive breast cancer? J Magn Reson Imaging 2017;46:1332–1340.

21. Ahmed A, Gibbs P, Pickles M, et al. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. J Magn Reson Imaging 2013;38:89–101.

22. Golden DI, Lipson JA, Telli ML, et al. Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer. J Am Med Inform Assoc 2013;20:1059–1066.

23. Michoux N, Van den Broeck S, Lacoste L, et al. Texture analysis on MR images helps predicting non-response to NAC in breast cancer. BMC Cancer 2015;15:574–587.

24. Teruel JR, Heldahl MG, Goa PE, et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. NMR Biomed 2014;27:887–896.

25. Wu J, Gong G, Cui Y, et al. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. J Magn Reson Imaging 2016; 44:1107–1115.

26. Pickles MD, Lowry M, Gibbs P. Pretreatment Prognostic Value of Dynamic Contrast-Enhanced Magnetic Resonance Imaging Vascular, Texture, Shape, and Size Parameters Compared With Traditional Survival Indicators Obtained From Locally Advanced Breast Cancer Patients. Invest Radiol 2016;51:177–185.

27. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 2006;100:229–235.

28. Pinheiro J, Bates D. Linear Mixed-Effects Model: Basic Concepts and Examples. In: Pinheiro J, Bates D, editors. Mixed-Effects Models in S and S-PLUS. New York: Springer, 2000.

29. Pinheiro J, Bates D, DebRoy S, et al. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117. 2014.

30. Tuncbilek N, Tokatli F, Altaner S, et al. Prognostic value DCE-MRI parameters in predicting factor disease free survival and overall survival for breast cancer patients. Eur J Radiol 2012;8:863–867.

31. Kim JH, Ko ES, Lim Y, et al. Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. Radiology 2017;282:665–675.

32. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. Br J Radiology 2017; 90:20160642.

33. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004;22:81–91.

34. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–762.

35. Mayerhoefer ME, Szomolanyi P, Jirak D, et al. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. Med Phys 2009;36:1236–1243.

36. Kanal E, Maravilla K, Rowley HA. Gadolinium contrast agents for CNS imaging: current concepts and clinical evidence. AJNR Am J Neuroradiol 2014;35:2215–2226.

37. Sikio M, Holli KK, Harrison LCV, et al. Parkinson's Disease: Interhemispheric Textural Differences in MR Images. Acad Radiol 2011;18:1217–1224.

38. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology 1999;211:101–110.

NTNU
Norwegian University of
Science and Technology