Differentiation of inflammatory from degenerative changes in the sacroiliac joints by machine learning supported texture analysis

Kepp, Felix H ; Huber, Florian A ; Wurnig, Moritz C ; Mannil, Manoj ; Kaniewska, Malwina ; Guglielmi, Riccardo ; Del Grande, Filippo ; Guggenberger, Roman

Abstract: PURPOSE To compare the diagnostic performance of texture analysis (TA) against visual qualitative assessment in the differentiation of spondyloarthritis (SpA) from degenerative changes in the sacroiliac joints (SIJ). METHOD Ninety patients referred for suspected inflammatory lower back pain from the rheumatology department were retrospectively included at our university hospital institution. MRI at 3 T of the lumbar spine and SIJ was performed with oblique coronal T1-weighted (w), fluid-sensitive fat-saturated (fs) TIRM and fsT1w intravenously contrast-enhanced (CE) images. Subjects were divided into three age- and gender-matched groups (30 each) based on definite clinical diagnosis serving as clinical reference standard with either degenerative, inflammatory (SpA) or no changes of the SIJ. SIJ were rated qualitatively by two independent radiologists and quantitatively by region-of-interest-based TA with 304 features subjected to machine learning logistic regression with randomized ten-fold selection of training and validation data. Qualitative and quantitative results were evaluated for diagnostic performance and compared against clinical reference standard. RESULTS Agreement of radiologist's diagnose with clinical reference was fair for both readers (= 0.32 and 0.44). ROC statistics revealed significant outperformance of TA compared to qualitative ratings for differentiation of SpA from remainder (AUC = 0.89 vs. 0.75), SpA from degenerative (AUC = 0.91 vs. 0.67) and TIRM-positive SpA (i.e. with bone marrow edema) from remainder cases (AUC = 0.95 vs. 0.76). T1w-CE images were the most important discriminator for detection of SpA. CONCLUSIONS TA is superior to qualitative assessment for the differentiation of inflammatory from degenerative changes of the SIJ. Intravenous CE-images increase diagnostic yield in quantitative TA.

DOI: https://doi.org/10.1016/j.ejrad.2021.109755

degenerative changes in the sacroiliac joints by machine learning supported texture analysis. European Journal of Radiology, 140:109755.
DOI: https://doi.org/10.1016/j.ejrad.2021.109755

European Journal of Radiology 140 (2021) 109755

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Differentiation of inflammatory from degenerative changes in the sacroiliac joints by machine learning supported texture analysis

Felix H. Kepp [a,b,1], Florian A. Huber [a,b,1,*], Moritz C. Wurnig [a,b], Manoj Mannil [a,b], Malwina Kaniewska [a,b], Riccardo Guglielmi [c], Filippo Del Grande [d], Roman Guggenberger [a,b]

[a] Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Raemistrasse 100, 8091 Zurich, Switzerland
[b] Faculty of Medicine, University of Zurich, Switzerland
[c] Institute of Radiology, Spital Thurgau AG, Cantonal Hospital Münsterlingen, Spitalcampus 1, 8596 Münsterlingen, Switzerland
[d] Istituto di imaging della Svizzera Italiana, Regional Hospital of Lugano, Via Tesserete 46, 6900 Lugano, Switzerland

ARTICLE INFO

Keywords:
Magnetic resonance imaging
Ankylosing spondylitis
Computer-assisted image analysis
Sacroiliac joint
Lower back pain

ABSTRACT

Purpose: To compare the diagnostic performance of texture analysis (TA) against visual qualitative assessment in the differentiation of spondyloarthritis (SpA) from degenerative changes in the sacroiliac joints (SIJ).
Method: Ninety patients referred for suspected inflammatory lower back pain from the rheumatology department were retrospectively included at our university hospital institution. MRI at 3 T of the lumbar spine and SIJ was performed with oblique coronal T1-weighted (w), fluid-sensitive fat-saturated (fs) TIRM and fsT1w intravenously contrast-enhanced (CE) images. Subjects were divided into three age- and gender-matched groups (30 each) based on definite clinical diagnosis serving as clinical reference standard with either degenerative, inflammatory (SpA) or no changes of the SIJ. SIJ were rated qualitatively by two independent radiologists and quantitatively by region-of-interest-based TA with 304 features subjected to machine learning logistic regression with randomized ten-fold selection of training and validation data. Qualitative and quantitative results were evaluated for diagnostic performance and compared against clinical reference standard.
Results: Agreement of radiologist's diagnose with clinical reference was fair for both readers ($\kappa = 0.32$ and 0.44). ROC statistics revealed significant outperformance of TA compared to qualitative ratings for differentiation of SpA from remainder (AUC = 0.89 vs. 0.75), SpA from degenerative (AUC = 0.91 vs. 0.67) and TIRM-positive SpA (i.e. with bone marrow edema) from remainder cases (AUC = 0.95 vs. 0.76). T1w-CE images were the most important discriminator for detection of SpA.
Conclusions: TA is superior to qualitative assessment for the differentiation of inflammatory from degenerative changes of the SIJ. Intravenous CE-images increase diagnostic yield in quantitative TA.

1. Introduction

In patients with inflammatory lower back pain due to spondyloarthritis (SpA) plain radiographs and CT may demonstrate advanced structural bone changes such as erosive changes and osseous proliferation [1,2]. However, MRI can identify pathologic changes in non- or pre-radiographic stages and may reveal signs of both, acute or chronic inflammation [3].

Criteria required for definition of sacroiliitis by MRI in patients with SpA published by the European League Against Rheumatism (EULAR) are based on qualitative, i.e. visually perceivable presence of subcortical bone marrow edema (BME) as an indicator of active inflammation [4]. Additional structural lesions such as bone erosion, new bone formation, sclerosis and/or ankyloses with fatty bone marrow infiltration are regarded as sequelae of chronic inflammation and only serve as supportive findings [4]. However, both BME and structural changes are also seen in degenerative changes or even asymptomatic patients [5] leading to rather low specificity of MRI alone in patients with inflammatory symptoms [6]. In addition, interreader agreement of qualitative imaging findings about the sacroiliac joints (SIJ) may show substantial variability depending on reader`s experience and clinical information [7,8].

Texture analysis (TA) as part of ambitions to extract quantitative

* Corresponding author at: Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Raemistrasse 100, 8051 Zurich, Switzerland.
E-mail address: florian.huber@usz.ch (F.A. Huber).
[1] Shared co-first authors.

https://doi.org/10.1016/j.ejrad.2021.109755
Received 25 January 2021; Received in revised form 9 April 2021; Accepted 3 May 2021

information from medical images characterizes signal patterns of pixels or voxels in a region or volume of interest mostly imperceptible to the human eye and is able to quantify parameters with high reproducibility [9]. TA has proven to be successful in characterization of lesion malignancy as well as in quantification of degenerative musculoskeletal disorders [10–12] and could potentially help in differentiating MR findings in SIJs.

The purpose of this study was to compare the diagnostic performance of TA against visual qualitative assessment in the differentiation of SpA from degenerative changes in the SIJ.

2. Material and methods

2.1. Study population

This retrospective study was performed in accordance with the Declaration of Helsinki and waived by the local ethic board (Cantonal Ethics Committee Zurich). Patients were over 18 years old and only included if a written general informed consent was signed and documented in the hospital information system. Patients were excluded if suffering from chronic diseases possibly affecting bone marrow composition, from diseases of bone metabolism, prior operation, trauma or fracture to the sacroiliac region. Furthermore, MRI of insufficient quality were excluded (e.g. due to motion- or metal-induced artifacts, insufficient oblique coronal angulation and/or insufficient fat-saturation).

Of 620 total referrals from the rheumatology outpatient clinic to our institution between 2015 and 2017, we primarily included 150 first visits with inflammatory lower back pain for an MRI examination at 3 T of the lumbar spine and/or the SIJ. Ninety patients who met all inclusion and exclusion criteria were eventually included in the study (STARD flowchart in Fig. 1).

2.2. Study cohorts

A total of 30 patients (out of the 150 primarily included subjects) were identified as valid cases of SpA, which was considered as suitable for the further study design based on prior power calculations (refer to

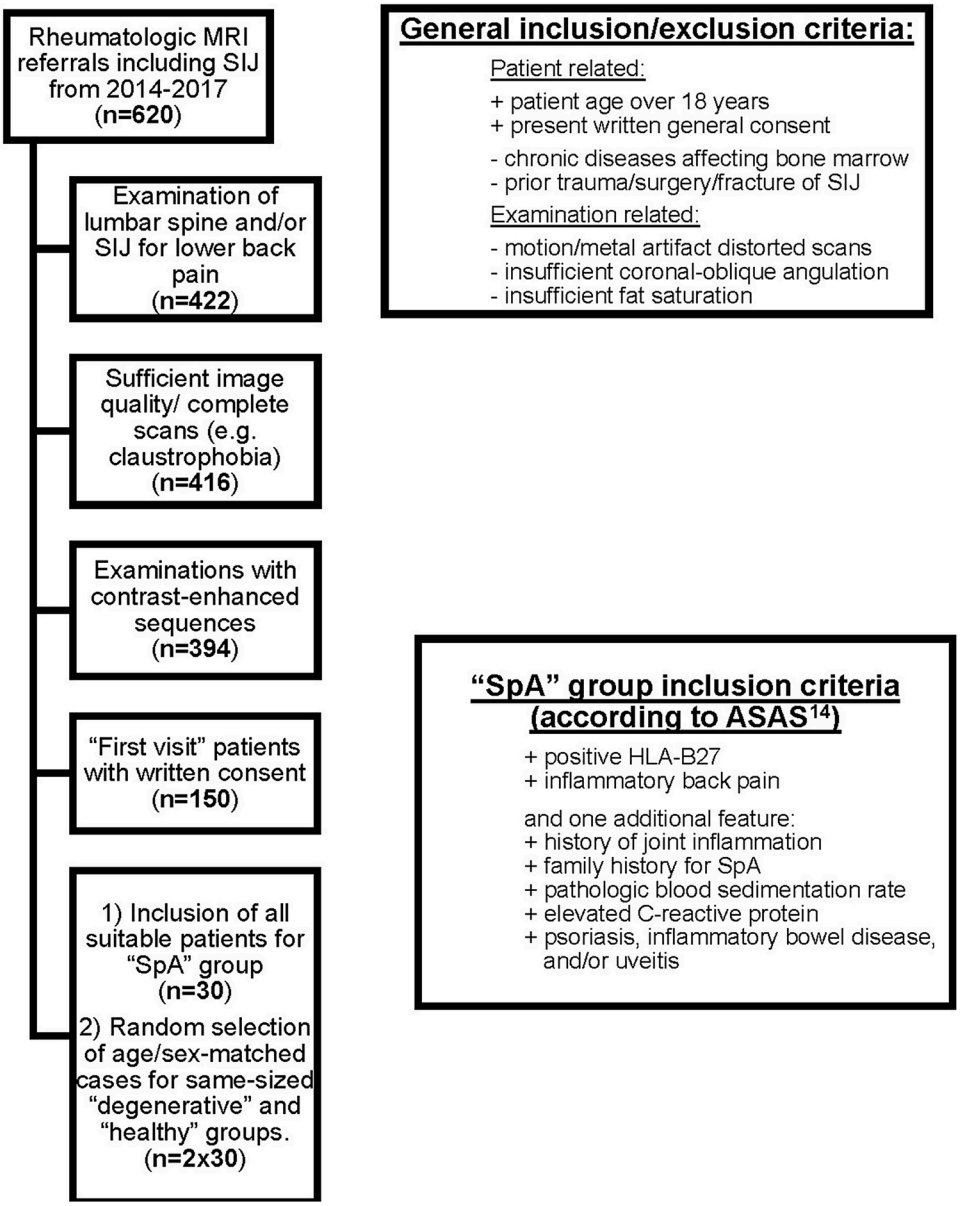

Fig. 1. Flowchart diagram of patient selection algorithm, in accordance with STARD checklist. (MRI = magnetic resonance imaging; SIJ = sacroiliac joint; SpA = spondylarthritis).

"Statistical analysis"). Out of the 120 remaining image datasets, we built two equally sized groups with either degenerative changes or normal SIJ by identifying the best age- and gender-matching cases. The remaining 60 cases were neglected intentionally for the purpose of balanced-groups statistics. Each group was formed based on clinical diagnosis established by our rheumatology department and served as the clinical reference standard.

The normal group suffered from unspecific lower back pain and showed no or minor radiological SIJ changes; hence after excluding any underlying rheumatologic condition by the treating rheumatologists it served as healthy control cohort. If present, follow-up visits in the patient record were screened in order to confirm clinical diagnosis and rule out possible transition to a rheumatoid disorder.

The degenerative group consisted of patients with lower back pain that showed SIJ changes on MRI but without evidence of underlying inflammatory disease based on EULAR [4] and ASAS [13] criteria in the standard diagnostic work-up (excluding history of arthritis, enthesitis, dactylitis; family history of SpA; HLA-B27 positivity; pathologic erythrocyte sedimentation rate or levels of C-reactive protein; manifestation of psoriasis, inflammatory bowel disease or uveitis; good response to NSAR) and was thus labelled as degenerative in nature.

The SpA group comprised patients with clinical symptoms of inflammatory lower back pain and in addition showed acute with/-out chronic inflammatory changes of the SIJ on MRI scans. Eventual positive clinical diagnosis was established based on aforementioned standard diagnostic work-up in accordance with ASAS and EULAR criteria [14].

2.3. Image acquisition and post-processing

All examinations were performed on two identical 3.0 T MR scanners (Magnetom Skyra, Siemens Healthineers, Erlangen, Germany) using the standard spine coil (204-element Integrated Spine coil, Tim Dockable Table, ditto). An institutional protocol was used for imaging of the SIJ consisting of the following three oblique coronal turbo spin-echo sequences (parallel with the long axis of the sacral bone): T1-weighted (w) (TR/TE 476/11 ms, matrix 487 × 487, field of view (FOV) 190 × 190 mm, slice thickness and increment (SL) 3 mm, parallel acquisition technique (PAT) GRAPPA, acceleration factor (AF) 2, number of excitations (NEX) 1, acquisition time (AT) 135 s) ; T2w fat-saturated (fs) TIRM (TR/TE 3490/40 ms, matrix 487 × 487, FOV 190 × 190 mm, SL 3 mm, PAT GRAPPA, AF 2, NEX 2, AT 320 s); fsT1w contrast-enhanced (fsT1wCE; TR/TE 652/11 ms, matrix 487 × 487, FOV 190 × 190 mm, SL 3 mm, PAT GRAPPA, AF 2, NEX 1, AT 162 s). Standard PACS (Impax 6, Agfa-Gevaert NV, Mortsel, Belgium) was used for study-related qualitative image interpretation.

Image selection was performed by the more experienced qualitative reader. For each of the 90 patients, the three most representative consecutive images with the most significant SIJ findings regarding BME (hyperintensity on TIRM and fsT1wCE) and/or structural changes (hypo-/hyperintensity on T1w from sclerosis/fatty infiltration), if present, were exported as DICOM of each of the respective sequence stacks. Exemplary images of the healthy cases were exported likewise. Of each of the reduced three image stacks (T1w, TIRM and fsT1wCE), one single image with the subjectively greatest amount of pathologic SIJ changes on the corresponding TIRM/T1w images was selected by both readers in consensus for TA, resulting in 3 × 90 = 270 images and 540 SIJ ROIs (one per side) for analysis.

Image size was rescaled to a pixel spacing of 0.39 mm.

2.4. Qualitative analysis

The qualitative readout was performed on reduced image stacks (three images per sequence) per patient, by two musculoskeletal-fellowship trained independent radiologists (F.H.K., M.C.W.) with four and five years of experience in musculoskeletal radiology. Both were blinded regarding radiology report and patient history. The readers scrolled through the images and rated each SIJ on a four-point Likert scale (0–3 for “normal appearance”, “mild”, “moderate” and “severe” alterations) for BME, enthesitis/joint effusion or synovitis/capsulitis and structural changes such as fatty infiltration, erosions, subchondral sclerosis and ankylosis, as disease-typic findings [4]. Prior to study readout, the readers performed a dedicated training session on study-unrelated exemplary images. In order to take into account different parameters of acute and chronic SIJ changes beyond BME, a non-weighted cumulative score was generated, summing up all Likert-ratings per patient. Furthermore, each reader had to assign a patient to either the normal, degenerative or inflammatory group, respectively.

2.5. Texture analysis

TA was performed by two blinded radiologists (F.A.H., M.K.) on each of the 180 single SIJ in the three MR sequences on single images prior identified to show the most severe SIJ changes (Fig. 2). Readers independently placed respective regions of interest (ROI) for TA using open source software (MaZda 4.6, Institute of Electronics, Technical University of Lodz, Lodz, Poland) [16]. Therefore, a circle-shaped two-dimensional ROI at a defined size of 3921 pixels (around 596 mm^2) was placed on each SIJ (Fig. 3). Criteria of placement included a) the SIJ joint space preferably crossing the center of each ROI, b) covering as much of abnormal findings as possible and c) excluding any nearby anatomical confounders like vessels, nerves and muscles. The definite ROI size was tested on a large number of study-unrelated images, where it was seen to be the largest possible size complying all criteria, throughout all cases. If no pathology was present in the respective SIJ, the readers placed the ROI randomly in the upper or lower image part of the synovial SIJ portions, considering the remainder quality criteria of ROI placement. Image normalization was performed in the TA software between histogram’s mean and three standard deviations to take care of technical intra- and interscanner differences [15]. A total 305 features were automatically calculated in the TA software. Feature reduction was first performed by exclusion of all parameters with an interreader agreement below 0.8, measured with intra-class correlation (ICC). Second, all highly redundant features (i.e. with a Pearson’s correlation coefficient of 0.75 or higher) were neglected, and only one parameter with the highest ICC was used for further analysis, respectively.

2.6. Statistical analysis

Statistical analysis was performed with SPSS (version 25.0; IBM, Armonk, New York, USA).

Our statistical rationale relied on power estimations with the qualitative cumulative scores. The presumption was a difference in mean scores between “SpA pos.” and “degenerative patients of at least 2, with a maximum assumed standard deviation of 4, which resulted in a minimum group of 30 patients per group ($\alpha = 0.05$, $\beta = 0.2$). To compensate exclusions due to sex/age-imbalance, we aimed to find around 35 suitable patients per study cohort within the 150 available cases who met all general study criteria.

Interreader agreement was determined for all qualitative variables using weighted kappa (κ) with linear weighting. Furthermore, Cohen’s κ between reader diagnosis and clinical reference standard was calculated for both readers. According to Landis and Koch, agreement levels were considered poor ($\kappa < 0$), slight ($0 < \kappa < 0.2$), fair ($0.21 < \kappa < 0.4$), moderate ($0.41 < \kappa < 0.6$), substantial ($0.61 < \kappa < 0.8$), and (almost) perfect ($0.81 < \kappa$), respectively [17]. Additionally, Mann-Whitney-U-Test was performed to reveal differences of mean ratings from all qualitative parameters (e.g. BME) between clinically degenerative or SpA patients. This analysis was performed using the qualitative results of the more experienced radiologist.

First, analysis of variance (ANOVA) was used to test for statistical

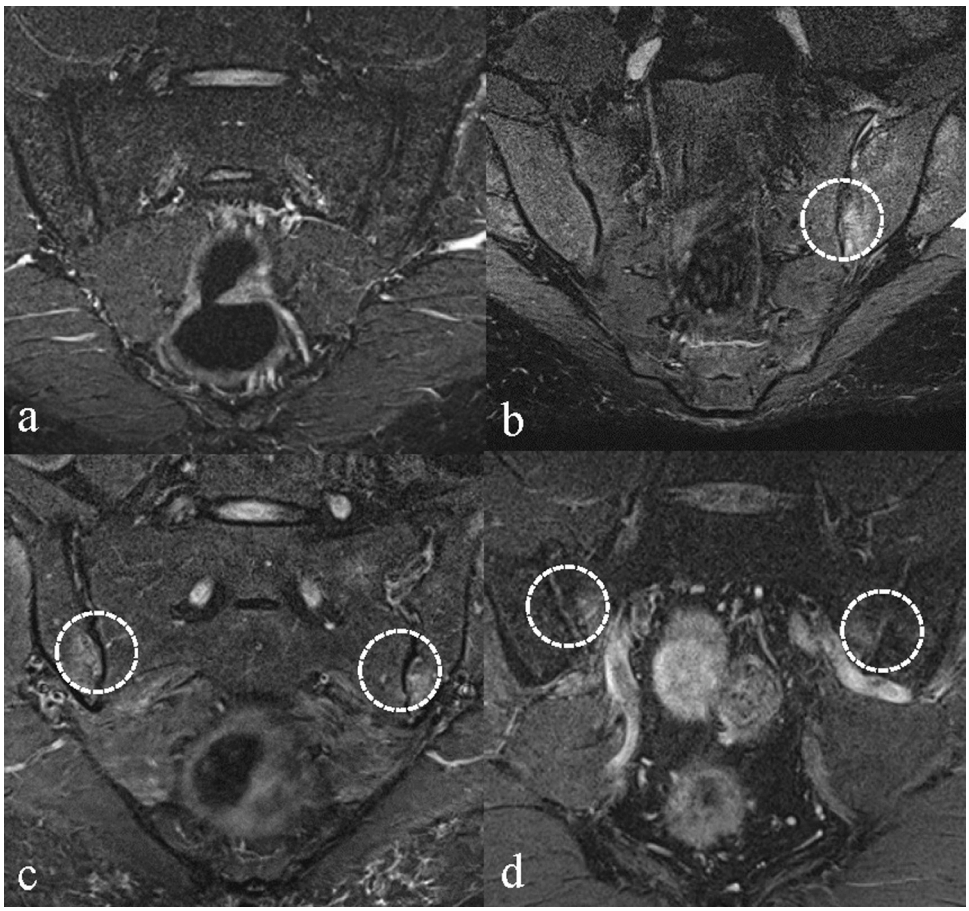

Fig. 2. Variety of BME adjacent to SIJs (white arrows). Panels present representative oblique coronal TIRM images from a) a healthy control patient, b) an individual with degenerative findings and from c) and d) patients with clinically confirmed SpA. While SpA was diagnosed by both readers in c), case d) was considered as degenerative by qualitative analysis of both readers but was correctly identified as inflammatory by texture analysis.

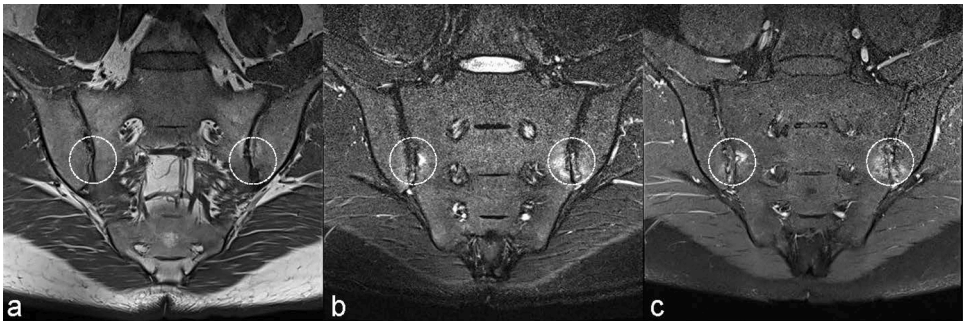

Fig. 3. Representative slices of a 34-year-old, male exemplary patient with acute findings in diagnosed SpA of both SIJ, in oblique coronal a) T1w, b) TIRM and c) fsT1wCE sequences, respectively. The dotted circle represents the identical size of placement of region of interest (ROI) for texture analysis (TA) measurements. For better visualization, the original color-filled ROIs from the TA software were post-processed for this manuscript.

differences of cumulative scores between the three clinically labeled groups. Second, dichotomization of the population was performed for further analysis, according to the following differentiation tasks of 1) SpA positive SIJ from remainder, 2) SpA positive SIJ from degenerative and 3) SpA positive SIJ with BME (i.e. rating > 1 for BME in TIRM by the more experienced reader) from remainder SIJ.

Subsequently, receiver operating characteristic (ROC) analysis was performed for assessing the diagnostic performance of the cumulative qualitative score, separately per reader. Furthermore, the same aforementioned three dichotomization scenarios were applied for the calculation of accuracy tables based on overall patient diagnoses of each reader, respectively.

Final TA parameters were further analyzed, whereas an algorithm had to solve the same three binary classification tasks as for qualitative ratings, e.g. classification of “SpA” or “remainders”. Therefore, a machine learning algorithm using logistic regression with subsequent ten-fold cross validation, i.e. random selection of training and validation data (1:9 fraction), was performed with R (v3.6.1, R Foundation for Statistical Computing, Vienna, Austria). Eventually, a ROC analysis of TA results was performed in a similar fashion to the qualitative analysis. P-values below 0.05 were considered statistically significant.

3. Results

3.1. Demographics

Out of total 90 patients (50 female (55,6 %) and 40 male (44,4 %) individuals) three groups of 30 patients each were built. The mean age was 46 years (range 18–82) without significant age differences among the equally sized groups ($p = 0.59$).

3.2. Qualitative analysis

Qualitative ratings per category were distributed as shown in Table 1 and were significantly less often pathologic for BME ($p < 0.01$), enthesitis/effusion ($p < 0.01$), erosions ($p < 0.05$), and ankylosis ($p < 0.05$) in patients with degenerative disease compared with SpA.

The interreader agreement ranged from "fair" ($\kappa = 0.33$) for fatty infiltration to "substantial" ($\kappa = 0.77$) for ankylosis. The majority of variables showed moderate agreement (Table 1). For both radiologists, there was fair ($k = 0.32$ and 0.4) agreement between their overall impression and the clinical reference diagnosis of a patient.

ANOVA demonstrated significant differences of cumulative qualitative scores among the three clinical labels for both readers, respectively (both $p < 0.001$; see Fig. 4).

ROC-analysis of the qualitative analysis revealed almost identical performance of both readers. This finding was consistent for the three dichotomized differentiation tasks (Table 2): While there were acceptable results for distinguishing SpA from the remainder SIJ (both AUC = 0.75), readers performed significantly poorer in differentiating between SpA and degeneration (AUC = 0.67 and 0.66). For the detection of SpA, there was no significant improvement if only TIRM-positive SIJ were included as SpA and compared against the remainder (both AUC = 0.76).

Diagnostic performance (sensitivity, specificity, and accuracy) was comparable between both readers (Table 2) and showed accuracies ranging from 65 % (SpA vs. degenerative) to 77.8 % (TIRM-pos. SpA vs. remainder).

3.3. Quantitative texture analysis

Feature selection revealed fourteen (fsT1wCE) and ten (T1w, TIRM) TA parameters for further analysis. The respective correlation matrices are given in Fig. 5.

For the detection of SpA, AUC values of the logistic regression algorithm after ten-fold cross validation ranged from 0.74 in TIRM to 0.81 in fsT1wCE images. Even higher results were achieved for the differentiation between SpA and degenerative cases (highest AUC 0.87 in fsT1wCE). Other than with qualitative ratings, SpA detection rate was notably higher if only TIRM-positive cases were counted as inflammatory cases. ROC analysis of TIRM-positive SpA versus the remainder cases then ranged from AUC 0.69 (T1w) to 0.91 (fsT1wCE).

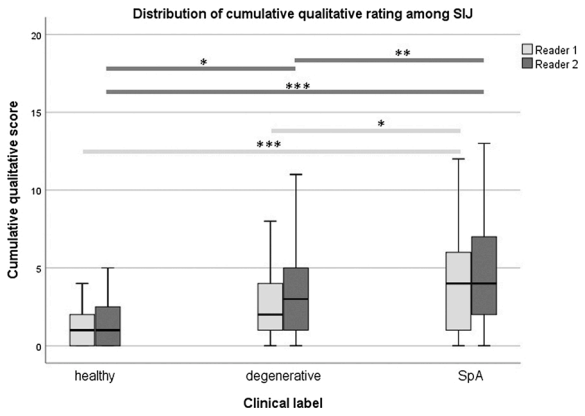

Fig. 4. Cumulative scores of both SIJs among three different patient groups (healthy, degenerative and SpA) for both readers. Mean qualitative overall scores differed significantly between groups. Asterisks indicate significant differences in post-hoc comparisons at significance levels of (*) $p < 0.05$, (**) $p < 0.01$ and (***) $p < 0.001$.

ROC-performances in all tasks with logistic regression for combined TA derived from all MR sequences delivered AUC-values of 0.89 for SpA vs. remainder, 0.91 for SpA vs. degenerative, and 0.95 for TIRM-positive SpA vs. remainder cases (Fig. 6).

In comparison, the best TA-derived AUC-values for the different tasks were always remarkably higher than those for cumulative qualitative scores (Table 3).

4. Discussion

In this investigation, we compared qualitative and quantitative image finfings in SIJ of SpA patients, degenerative disease and healthy controls. Throughout all calculations, quantitative TA outperformed qualitative radiologist performance in correctly labeling imaging of a patient to their clinical reference standards, as defined by relevant guidelines. The best performing discriminator in TA was the utilization of CE sequences.

There is limited work assessing the diagnostic performance of TA for differentiation of inflammatory from degenerative changes in joints and bones. This is the first study to do so in order to differentiate SpA patients from degenerative SIJ changes.

Since the introduction of MRI as the most sensitive tool for the detection of early changes in SpA [18], the EULAR guidelines [4] have focused on the identification of BME in order to diagnose active sacroiliitis/SpA. As this task not only depends on the clinical context but also on the reader`s experience it is associated with substantial interreader variability. Numerous studies have tried to improve accuracy for detection of acute or chronic SpA, e.g. by reassessing clinical history [13] or utilizing CE-MRI [19], yet with limited success.

Table 1
Distribution of qualitative assessment of the more experienced reader for each parameter, separate per body side and per patient population. Mann-Whitney-U-test was performed between single ratings, and post-hoc ANOVA tests for total score, in degenerative and SpA-SIJ, respectively. Inter-reader agreements were calculated overall subgroups, the values represent p-levels. Total amount of pathologic ratings (1 – 3, i.e. mild – severe), and medians in parenthesis. Means, standard deviations and 95 % confidence intervals of the respective cumulative scores are given in the last line.

	Controls		Degenerative		Axial SpA		p-value degenerative vs. SpA	Inter-reader agreement (κ)
	Left	Right	Left	Right	Left	Right	Both	Both
Bone marrow edema (TIRM)	8 (0)	6 (0)	16 (1)	20 (1)	23 (1)	25 (1)	<0.01	0.742
Enthesitis/Effusion (TIRM)	2 (0)	4 (0)	3 (0)	2 (0)	3 (0)	8 (0)	<0.01	0.507
Synovitis/Capsulitis (fsT1wCE)	1 (0)	0 (0)	4 (0)	4 (0)	4 (0)	8 (0)	0.05	0.462
Fatty infiltration (T1w)	9 (0)	9 (0)	9 (0)	11 (0)	12 (0)	12 (0)	0.885	0.327
Erosions (T1w)	4 (0)	2 (0)	9 (0)	13 (0)	15 (0.5)	16 (1)	<0.05	0.484
Subchondral sclerosis (T1w)	8 (0)	8 (0)	14 (0)	18 (1)	19 (1)	20 (1)	0.167	0.416
Ankylosis yes/no (T1w)	2 (0)	2 (0)	1 (0)	2 (0)	1 (0)	0 (0)	<0.05	0.766
Total score	3.1 ± 4.99 [1.34;5.06]		6.1 ± 4.16 [4.55;7.65]		9.47 ± 6.17 [7.16;11.77]		<0.01	

Table 2
Diagnostic performance metrics of both readers' diagnoses. The respective performances are listed for the three differentiation scenarios of 1) SpA vs. non-SpA, 2) SpA vs. degenerative, and for 3) TIRM-positive SpA vs. the remainder patients, as well as calculated overall for both readers. Furthermore, area under the curve (AUC)-values from receiver operating characteristic (ROC) analysis based on cumulative scores are stated separately for each reader.

	SpA vs. non-SpA		SpA vs. degenerative		TIRM pos. SpA vs. remainder		Overall
	Reader 1	*Reader 2*	*Reader 1*	*Reader 2*	*Reader 1*	*Reader 2*	*Both readers*
Sensitivity	66.7 %	70 %	66.7 %	70 %	71.4 %	71.4 %	69.4 %
Specificity	80 %	76.7 %	66.7 %	60 %	80.6 %	75.8 %	73.3 %
Accuracy	75.6 %	74.4 %	66.7 %	65 %	77.8 %	74.4 %	72.3 %
AUC	0.75	0.75	0.67	0.66	0.76	0.76	–

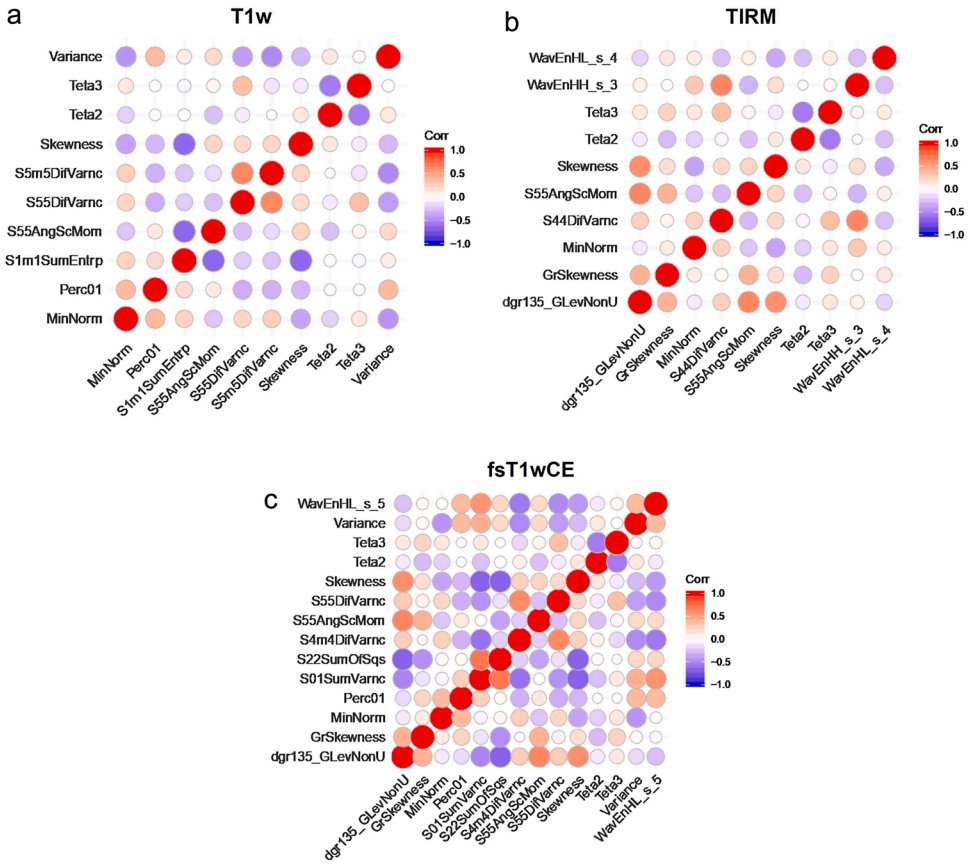

Fig. 5. Correlation plot of all used texture analysis (TA) features in a) T1w, b) TIRM and c) fsT1wCE images, after primary and secondary feature reduction. Color and size of circles represent Pearson's correlation coefficient from −1 (dark blue) to +1 (red). The respective descriptions are common abbreviations of standard TA features and listed in detail in Supplementary files.

Our approach of including quantitative results for reporting SIJ changes is associated with inherent high interreader agreement as only variables with high reproducibility were included (ICC > 0.8) compared to mere qualitative assessment where heterogeneous agreement among variables was seen ("fair" (κ = 0.33) for fatty infiltration to "substantial" (κ = 0.77) for ankylosis). Also in accordance with clinical experience and literature [20], qualitative labels correlated only fairly with clinical reference standard for both readers (κ = 0.32 and 0.4), and diagnostic performance based on reading physician`s labels were fair at best, considering metrics of diagnostic accuracy around 70 %. The latter may be due to an inconsistent rating of fatty infiltration and an over-diagnosis of inflammatory changes based on BME alone which was present in 73 % of all cases. It is however well known that not all BME or associated structural changes about the SIJ are caused by inflammation.

While based on current ASAS and EULAR recommendations for the imaging arm in the diagnosis pathway of SpA, BME is a prerequisite for diagnosing a patient with active sacroiliitis. Structural lesions such as bone erosion, new bone formation, sclerosis and fatty infiltration should also be taken into account but are also not specific for inflammation [21]. Given the importance of MRI for ruling out SpA in the SIJ, radiologists are tempted to over-diagnose BME or structural changes in patients with unspecific lower back pain as inflammatory. Despite overall significant differences in incidence of certain key qualitative parameters among different patient groups, e.g. BME, effusion or erosions, they were all associated with weak interreader agreement and only fair agreement with clinical diagnosis.

Beyond qualitative assessment, a positive effect of quantitative TA on diagnostic accuracy has been shown in different musculoskeletal applications, e.g. in pattern-based quantification of fatty muscle infiltration [22,23]. TA as a quantitative approach can not only help to inherently increase interreader agreement but also to refine radiologic diagnosis in order to deliver higher specificity for the underlying cause.

In this study combined TA of all available sequences was able to identify SpA at remarkable diagnostic performance (AUC = 0.91), while qualitative interpretation (AUC = 0.67) and TA of traditional SpA-sensitive sequences (AUC = 0.71 for TIRM) both performed

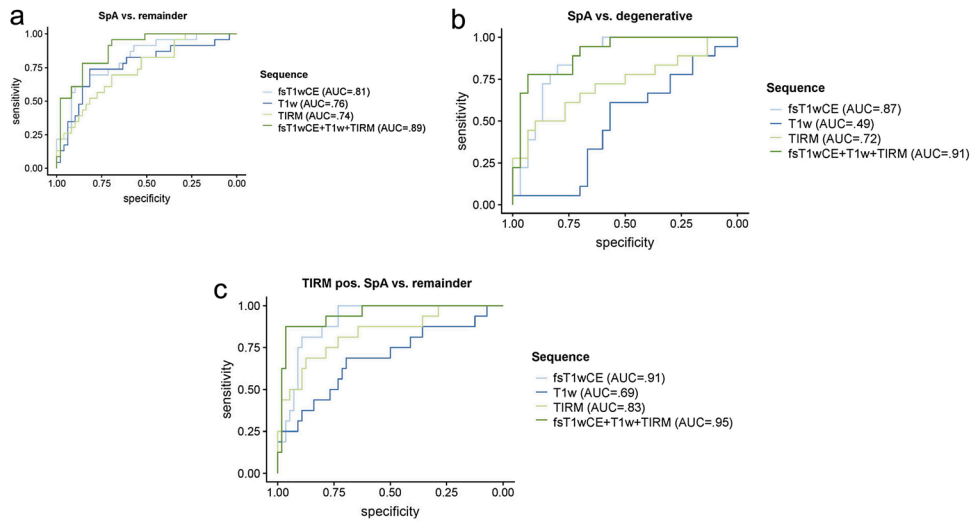

Fig. 6. ROC statistic of quantitative texture analysis (TA). The ROC-curves represent separate results for TA in each sequence (fsT1wCE, T1w and TIRM) as well as for regression analysis from combined TA data of all three sequences. The results are given for differentiation between a) SpA vs. remainder, c) SpA vs. degenerative, and for c) TIRM-positive SpA vs. the remainder patients.

Table 3
Texture analysis (TA) performance. Comparison of performance of TA for the three different tested scenarios and overall, compared to cumulative qualitative scores. Values indicate area under curve (AUC).

AUC	SpA vs. non-SpA	SpA vs. degenerative	TIRM pos. SpA vs. remainder	Overall
TA	0.89	0.91	0.95	0.92
Score	0.75	0.67	0.76	0.73

significantly inferior to the combined approach. While the relation between certain photographic patterns and TA features have been described in detail, the links between cellular/compositional bone marrow changes and their influence on different TA parameters is debatable [9]. However, certain discriminative TA parameters in this study were found to be of importance in other tasks on CE-MRI like hepatocellular carcinoma detection [24,25] or background parenchymal enhancement of the breasts [26]. There have been attempts to increase diagnostic accuracy by the use of CE-MRI in SpA [27,28] but the results are not convincing and thus the use of intravenous contrast material is currently not recommended by the EULAR taskforce. According to our data, unlike EULAR recommendations, fsT1wCE was the most important discriminator in classifying/detection SpA. One reason might be that the increased contrast uptake of BME in inflammatory SIJ may be more pronounced or structurally different on a microscopic level, thus perceivable by TA when compared to degenerative causes [29,30]. Most notably, fsT1wCE was essential for differentiation between SpA and degenerative cases, which is probably the most frequent dilemma that musculoskeletal radiologists face when reading SIJ MR.

The following study limitations should be acknowledged. First, this was a retrospective single center study. To create standardized conditions patients were scanned on identical MR scanners using identical coils and protocols. Different scanners, field-strengths and protocols might impact qualitative and TA results. Moreover, TA was performed in two-dimensional fashion only and may be of limited reproducibility/applicability for 3D-acquired sequences. Second, our study included individuals with optimal MR scans devoid of technical flaws, e.g. motion artifacts. Artifacts in clinical routine may affect the applicability of our findings to large-scale use. Another limitation is the inclusion of patients with minor MR findings into the healthy control group. While our group was defined mostly by clinical criteria and follow-ups did not support the clinical relevance of minimal radiologic findings, it is nonetheless arguable that an inclusion bias may be present. Finally, our results based on qualitative and quantitative readouts performed by musculoskeletal-fellowship trained radiologists with a minimum of five years of experience, as we aimed to represent the expertise in most radiologic service facilities. Experienced dedicated expert musculoskeletal radiologists with specific rheumatologic know-how may perform better, but in the authors` experience even then diagnostic accuracy usually lags behind expectations.

5. Conclusions

In conclusion, TA from MRI was superior to qualitative assessment for differentiation of inflammatory SpA from SIJ degeneration. Our results indicate that the use of i.v. contrast in addition to native sequences may have a potentional additive value in this task, when combined with quantitative TA.

Further research is needed to investigate the role of TA as promising auxiliary diagnostic tool in unveiling subvisual changes in early-stage SpA.

Author contributions

R.G.: Guarantor of integrity of the entire study. F.H.K., F.A.H., R.G., F.D.G.: Study concepts and design. F.H.K.: Literature research. F.H.K., M.C.W., M.M., M.K.: Experimental studies/data analysis. F.A.H., M.C.W., M.M.: Statistical analysis. F.A.H.: Manuscript preparation. All authors: Manuscript editing.

Funding

The authors state that this work has not received any funding.

Guarantor

The scientific guarantor of this publication is Roman Guggenberger, MD.

Statistics and biometry

Statistical analysis was performed under supervision of Mr. Wurnig, who holds a graduate degree in mathematics.

Informed consent

A written general consent for research purposes was present for all included patients.

Ethical approval

This research was done in accordance with the standards of the responsible institutional and national ethics committee. Approval from the institutional review board and the respective ethics committee was given for this retrospective study.

Methodology

- retrospective
- experimental
- performed at one institution

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ejrad.2021.109755.

References

[1] J. Braun, J. Sieper, M. Bollow, Imaging of sacroiliitis, Clin. Rheumatol. 19 (1) (2000) 51–57.

[2] A. Lacout, B. Rousselin, J.P. Pelage, CT and MRI of spine and sacroiliac involvement in spondyloarthropathy, AJR Am. J. Roentgenol. 191 (4) (2008) 1016–1023, https://doi.org/10.2214/AJR.07.3446.

[3] N. Inanc, P. Atagündüz, F. Sen, T. Biren, H.T. Turoğlu, H. Direskeneli, The investigation of sacroiliitis with different imaging techniques in spondyloarthropathies, Rheumatol. Int. 25 (8) (2005) 591–594, https://doi.org/10.1007/s00296-004-0490-9.

[4] P. Mandl, V. Navarro-Compán, L. Terslev, P. Aegerter, D. van der Heijde, M. A. D'Agostino, X. Baraliakos, S.J. Pedersen, A.G. Jurik, E. Naredo, C. Schueller-Weidekamm, U. Weber, M.C. Wick, P.A. Bakker, E. Filippucci, P.G. Conaghan, M. Rudwaleit, G. Schett, J. Sieper, S. Tarp, H. Marzo-Ortega, M. Østergaard, (EULAR) ELAR, EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice, Ann. Rheum. Dis. 74 (7) (2015) 1327–1339, https://doi.org/10.1136/annrheumdis-2014-206971.

[5] U. Weber, R.G. Lambert, S.J. Pedersen, J. Hodler, M. Østergaard, W. P. Maksymowych, Assessment of structural lesions in sacroiliac joints enhances diagnostic utility of magnetic resonance imaging in early spondylarthritis, Arthritis Care Res. (Hoboken) 62 (12) (2010) 1763–1771, https://doi.org/10.1002/acr.20312.

[6] U. Weber, R.G. Lambert, M. Østergaard, J. Hodler, S.J. Pedersen, W. P. Maksymowych, The diagnostic utility of magnetic resonance imaging in spondylarthritis: an international multicenter evaluation of one hundred eighty-seven subjects, Arthritis Rheum. 62 (10) (2010) 3048–3058, https://doi.org/10.1002/art.27571.

[7] C. Lukas, J. Braun, D. van der Heijde, K.G. Hermann, M. Rudwaleit, M. Østergaard, A. Oostveen, P. O'Connor, W.P. Maksymowych, R.G. Lambert, A.G. Jurik, X. Baraliakos, R. Landewé, Group AOMiAW, Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment, J. Rheumatol. 34 (4) (2007) 862–870.

[8] B. Arnbak, T.S. Jensen, C. Manniche, A. Zejden, N. Egund, A.G. Jurik, Spondyloarthritis-related and degenerative MRI changes in the axial skeleton–an inter- and intra-observer agreement study, BMC Musculoskelet. Disord. 14 (2013) 274, https://doi.org/10.1186/1471-2474-14-274.

[9] M.G. Lubner, A.D. Smith, K. Sandrasegaran, D.V. Sahani, P.J. Pickhardt, CT texture analysis: definitions, applications, biologic correlates, and challenges, Radiographics 37 (5) (2017) 1483–1503, https://doi.org/10.1148/rg.2017170056.

[10] C.S. Lisson, C.G. Lisson, K. Flosdorf, R. Mayer-Steinacker, M. Schultheiss, A. von Baer, T.F.E. Barth, A.J. Beer, M. Baumhauer, R. Meier, M. Beer, S.A. Schmidt, Diagnostic value of MRI-based 3D texture analysis for tissue characterisation and discrimination of low-grade chondrosarcoma from enchondroma: a pilot study, Eur. Radiol. 28 (2) (2018) 468–477, https://doi.org/10.1007/s00330-017-5014-6.

[11] F.A. Huber, S. Stutz, I. Vittoria de Martini, M. Mannil, A.S. Becker, S. Winklhofer, J. M. Burgstaller, R. Guggenberger, Qualitative versus quantitative lumbar spinal stenosis grading by machine learning supported texture analysis-Experience from the LSOS study cohort, Eur. J. Radiol. 114 (2019) 45–50, https://doi.org/10.1016/j.ejrad.2019.02.023.

[12] M. Mannil, J.M. Burgstaller, U. Held, M. Farshad, R. Guggenberger, Correlation of texture analysis of paraspinal musculature on MRI with different clinical endpoints: lumbar Stenosis Outcome Study (LSOS), Eur. Radiol. 29 (1) (2019) 22–30, https://doi.org/10.1007/s00330-018-5552-6.

[13] M. Rudwaleit, A. Metter, J. Listing, J. Sieper, J. Braun, Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria, Arthritis Rheum. 54 (2) (2006) 569–578, https://doi.org/10.1002/art.21619.

[14] J. Sieper, M. Rudwaleit, X. Baraliakos, J. Brandt, J. Braun, R. Burgos-Vargas, M. Dougados, K.G. Hermann, R. Landewé, W. Maksymowych, D. van der Heijde, The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis, Ann. Rheum. Dis. 68 (Suppl. 2) (2009), https://doi.org/10.1136/ard.2008.104018 ii1–44.

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

[16] P.M. Szczypiński, M. Strzelecki, A. Materka, A. Klepaczko, MaZda–a software package for image texture analysis, Comput. Methods Programs Biomed. 94 (1) (2009) 66–76, https://doi.org/10.1016/j.cmpb.2008.08.005.

[17] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977) 159–174.

[18] A.N. Bennett, D. McGonagle, P. O'Connor, E.M. Hensor, F. Sivera, L.C. Coates, P. Emery, H. Marzo-Ortega, Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-B27 status in early inflammatory back pain predict radiographically evident ankylosing spondylitis at eight years, Arthritis Rheum. 58 (11) (2008) 3413–3418, https://doi.org/10.1002/art.24024.

[19] C. Lin, J.D. MacKenzie, J.L. Courtier, J.T. Gu, D. Milojevic, Magnetic resonance imaging findings in juvenile spondyloarthropathy and effects of treatment observed on subsequent imaging, Pediatr. Rheumatol. Online J. 12 (2014) 25, https://doi.org/10.1186/1546-0096-12-25.

[20] D. van der Heijde, J. Sieper, W.P. Maksymowych, M.A. Brown, R.G. Lambert, S. S. Rathmann, A.L. Pangan, Spinal inflammation in the absence of sacroiliac joint inflammation on magnetic resonance imaging in patients with active nonradiographic axial spondyloarthritis, Arthritis Rheumatol. 66 (3) (2014) 667–673, https://doi.org/10.1002/art.38283.

[21] T. Diekhoff, K.G. Hermann, J. Greese, C. Schwenke, D. Poddubnyy, B. Hamm, J. Sieper, Comparison of MRI with radiography for detecting structural lesions of the sacroiliac joint using CT as standard of reference: results from the SIMACT study, Ann. Rheum. Dis. 76 (9) (2017) 1502–1508, https://doi.org/10.1136/annrheumdis-2016-210640.

[22] H. Alizai, L. Nardo, D.C. Karampinos, G.B. Joseph, S.P. Yap, T. Baum, R. Krug, S. Majumdar, T.M. Link, Comparison of clinical semi-quantitative assessment of muscle fat infiltration with quantitative assessment using chemical shift-based water/fat separation in MR studies of the calf of post-menopausal women, Eur. Radiol. 22 (7) (2012) 1592–1600, https://doi.org/10.1007/s00330-012-2404-7.

[23] M. Mannil, J.M. Burgstaller, A. Thanabalasingam, S. Winklhofer, M. Betz, U. Held, R. Guggenberger, Texture analysis of paraspinal musculature in MRI of the lumbar spine: analysis of the lumbar stenosis outcome study (LSOS) data, Skeletal Radiol. 47 (7) (2018) 947–954, https://doi.org/10.1007/s00256-018-2919-3.

[24] W. Zhou, L. Zhang, K. Wang, S. Chen, G. Wang, Z. Liu, C. Liang, Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced MR images, J. Magn. Reson. Imaging 45 (5) (2017) 1476–1484, https://doi.org/10.1002/jmri.25454.

[25] D. Stocker, H.P. Marquez, M.W. Wagner, D.A. Raptis, P.A. Clavien, A. Boss, M. A. Fischer, M.C. Wurnig, MRI texture analysis for differentiation of malignant and benign hepatocellular tumors in the non-cirrhotic liver, Heliyon 4 (11) (2018), e00987, https://doi.org/10.1016/j.heliyon.2018.e00987.

[26] Y. Amano, J. Woo, M. Amano, F. Yanagisawa, H. Yamamoto, M. Tani, MRI texture analysis of background parenchymal enhancement of the breast, Biomed Res. Int. 2017 (2017), 4845909, https://doi.org/10.1155/2017/4845909.

[27] C.E. Althoff, E. Feist, E. Burova, I. Eshed, M. Bollow, B. Hamm, K.G. Hermann, Magnetic resonance imaging of active sacroiliitis: do we really need gadolinium? Eur. J. Radiol. 71 (2) (2009) 232–236, https://doi.org/10.1016/j.ejrad.2009.04.034.

[28] M. de Hooge, R. van den Berg, V. Navarro-Compán, F. van Gaalen, D. van der Heijde, T. Huizinga, M. Reijnierse, Magnetic resonance imaging of the sacroiliac joints in the early detection of spondyloarthritis: no added value of gadolinium compared with short tau inversion recovery sequence, Rheumatology (Oxford) 52 (7) (2013) 1220–1224, https://doi.org/10.1093/rheumatology/ket012.

[29] H.A. Elshabrawy, Z. Chen, M.V. Volin, S. Ravella, S. Virupannavar, S. Shahrara, The pathogenic role of angiogenesis in rheumatoid arthritis, Angiogenesis 18 (4) (2015) 433–448, https://doi.org/10.1007/s10456-015-9477-2.

[30] G. Bhatnagar, J. Makanyanga, B. Ganeshan, A. Groves, M. Rodriguez-Justo, S. Halligan, S.A. Taylor, MRI texture analysis parameters of contrast-enhanced T1-weighted images of Crohn's disease differ according to the presence or absence of histological markers of hypoxia and angiogenesis, Abdom. Radiol. (NY) 41 (7) (2016) 1261–1269, https://doi.org/10.1007/s00261-016-0657-3.