

Original paper

Bone marrow lesions of the femoral head: can radiomics distinguish whether it is reversible?

Halitcan Batur^{1,A,B,C,D,E,F}, Bokebatur Ahmet Rasit Mendi^{2,A,D,E,F}, Nurdan Cay^{3E,F}

¹Department of Pediatric Radiology, Ankara City Hospital, Ankara, Turkey

²Department of Radiology, Nigde Omer Halisdemir University Education and Research Hospital, Nigde, Turkey

³Department of Radiology, Ankara Yildirim Beyazıt University, Ankara, Turkey

Abstract

Purpose: Contrary to the self-limiting nature of reversible bone marrow lesions, irreversible bone marrow lesions require early surgical intervention to prevent further morbidity. Thus, early discrimination of irreversible pathology is necessitated. The purpose of this study is to evaluate the efficacy of radiomics and machine learning regarding this topic.

Material and methods: A database was scanned for patients who had undergone MRI of the hip for differential diagnosis of bone marrow lesions and had had follow-up images acquired within 8 weeks after the first imaging. Images that showed resolution of oedema were included in the reversible group. The remainders that showed progression into characteristic signs of osteonecrosis were included in the irreversible group. Radiomics was performed on the first MR images, calculating first- and second-order parameters. Support vector machine and random forest classifiers were performed using these parameters.

Results: Thirty-seven patients (seventeen osteonecrosis) were included. A total of 185 ROIs were segmented. Forty-seven parameters were accepted as classifiers with an area under the curve value ranging from 0.586 to 0.718. Support vector machine yielded a sensitivity of 91.3% and a specificity of 85.1%. Random forest classifier yielded a sensitivity of 84.8% and a specificity of 76.7%. Area under curves were 0.921 for support vector machine and 0.892 for random forest classifier.

Conclusions: Radiomics analysis could prove useful for discrimination of reversible and irreversible bone marrow lesions before the irreversible changes occur, which could prevent morbidities of osteonecrosis by guiding the decision-making process for management.

Key words: magnetic resonance imaging, osteonecrosis, texture analysis, radiomics, bone marrow oedema.

Introduction

Bone marrow lesions (BML) were originally known as 'bone marrow oedema' (BME) which was defined as a non-specific pattern of ill-defined area of low signal intensity on T1-weighted magnetic resonance images and high signal intensity on T2-weighted or short-tau inversion-recovery images [1]. First, it was thought that there was an increased

fluid in the bone marrow, but later histopathological studies have shown that variably there can be marrow oedema, necrosis, bleeding, fibrosis, ingrowth of fibrovascular tissue, lymphocytic infiltrates, and trabecular bone abnormalities [2-4]. BML can be caused by a broad spectrum of pathologies such as trauma, degenerative, inflammatory, ischaemic, infectious, metabolic/endocrine, iatrogenic, and neoplastic lesions [3]. With clinical findings, the differential diagnosis list can be narrowed, but it is still vital to

Correspondence address:

Dr. Halitcan Batur, Department of Pediatric Radiology, Ankara City Hospital, Turkey, e-mail: halitcanbatur@gmail.com

Authors' contribution:

A Study design · B Data collection · C Statistical analysis · D Data interpretation · E Manuscript preparation · F Literature search · G Funds collection

distinguish if the BML is reversible because the treatment options vary from pharmaceutical agents to joint preserving surgeries. At this point, advanced imaging techniques become important. Even with the use of magnetic resonance imaging (MRI), the differential diagnosis between reversible and irreversible BML can be challenging because bone marrow oedema can obscure other subtle changes. Furthermore, signs that are characteristic of osteonecrosis appear in MR images when irreversible damage settles [5,6]. Thus, an early diagnosis of irreversible BML before well-known imaging findings appear is of great importance.

Radiomics is a non-invasive and quantitative imaging method that evaluates the size, shape, and first- and second-order texture features in radiological imaging [7]. First-order texture parameters depict the intensity distribution within the segmented image, while second-order parameters demonstrate the statistical relationships between voxels in 3-dimensional space [8]. Diagnostic capabilities of radiomics in subjects that are challenging to radiologists have been demonstrated in previous literature [9]. The aim of this retrospective study is to evaluate the diagnostic capability of radiomics analysis and machine learning based on MRI regarding discrimination of bone marrow lesions before irreversible joint damage occurs.

Material and methods

This study was approved by Ankara City Hospital Institutional Review Board (decision number: E2-21-841, decision date: 20/09/2021). Necessity of written consent was waived by the IRB due to the retrospective design of the study and the anonymization of patient data.

Study population

The institutional database was scanned for patients who underwent hip MRI imaging for differential diagnosis of BML of the femoral head. The follow-up protocol of

our institution includes a second MRI, scheduled within 8 weeks of the first examination, which we used for the classification of patients. Patients whose images demonstrated resolution of oedema were included in the reversible group. Patients, whose images showed single- and double-line signs or crescent signs [5,6] were included in the irreversible group. Exclusion criteria were absence of follow-up images within 8 weeks, motion or susceptibility artifacts that hinder evaluation and segmenting process, bone marrow oedema related to another definitive diagnosis such as osteomyelitis, lymphoma, primary or metastatic bone masses, and the presence of clinical history preceding bone marrow reconversion. Patients who already had advanced stage osteonecrosis-related changes in first images were also excluded. The patient selection is summarized in Figure 1.

MRI acquisition and pre-processing

T1-weighted turbo spin echo images were utilized for analysis because radiomics features extracted from T1-weighted images were shown to be valuable tools in previous literature [10]. The images were acquired on a GE Signa Pioneer 3T scanner with a slice thickness of 4 mm, pixel size 0.57×0.57 mm, echo time of 12.4 ms, and repetition time of 688 ms. Relevant images were uploaded to IBEX software (University of Texas MD Anderson Cancer Center, Houston, TX, USA), and 2D segmentations were performed by 2 independent observers (H.B., 7 years of experience; B.A.R.M., 5 years of experience). Each lesion was segmented by choosing the largest single ROI per slice possible. In each section, the margins of the region of interest (ROI) containing the bone marrow oedema were determined manually using the free-hand method. Bone marrow oedema and normal marrow distinction was done based on the previous definition in the literature [1]. Because this definition includes an “ill-defined T1 hypointensity”, we used STIR images as a guide due to the conspicuity of oedema. Cortical bone, adjacent fat

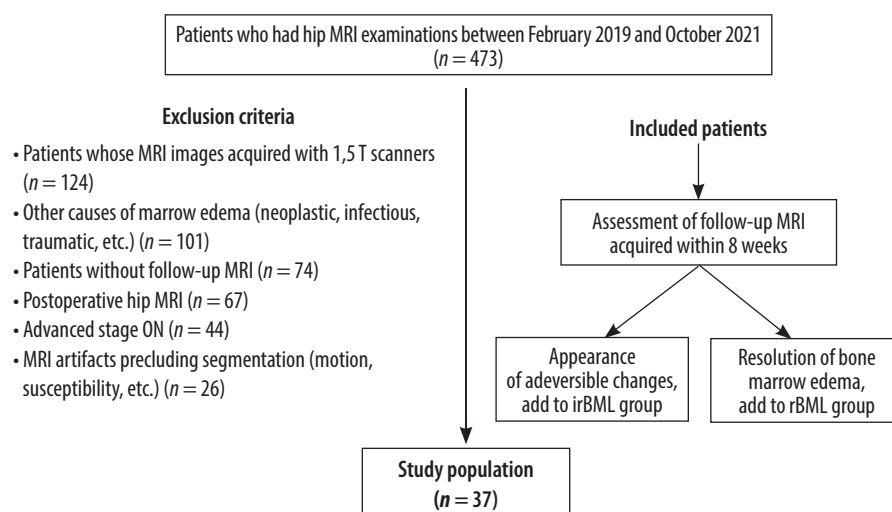


Figure 1. Summary of patient selection process

and muscle planes, and vascular structures were carefully excluded in each section while selecting the borders. We used 2D segmentation instead of 3D segmentation because the routine MRI protocol of our institution included only non-volumetric T1 sequences. An example of ROI acquisition is presented in Figure 2.

Texture analysis

Texture analysis was performed after the ROI of each lesion was obtained. The voxel size was re-sampled as $0.7 \times 0.7 \times 4$ mm using the bi-cubic interpolation method. The number of bins for histogram preparation and grey-level discretization was set to 64.

The relationship of 2 neighbouring voxels, which forms the basis for the calculation of second-order parameters,

was performed using offsets of one voxel and displacement vectors that isotropically separated from the centre voxel in 3D space as multiples of 45° in 13 different directions.

Because 2D ROIs were acquired, shape- and size-based parameters were not used for radiomics analysis.

After creating the presets, 74 radiomics parameters grouped as first and second order (Gray-Level Co-occurrence Matrix [GLCM], Grey-Level Run-Length Matrix [GLRLM], and Neighbouring Grey-Tone Difference Matrix [NGTDM]) were calculated using ROIs obtained by manual segmentation method.

Data analysis

Comparison of numerical variables among groups was done using the Mann-Whitney U (MWU) test.

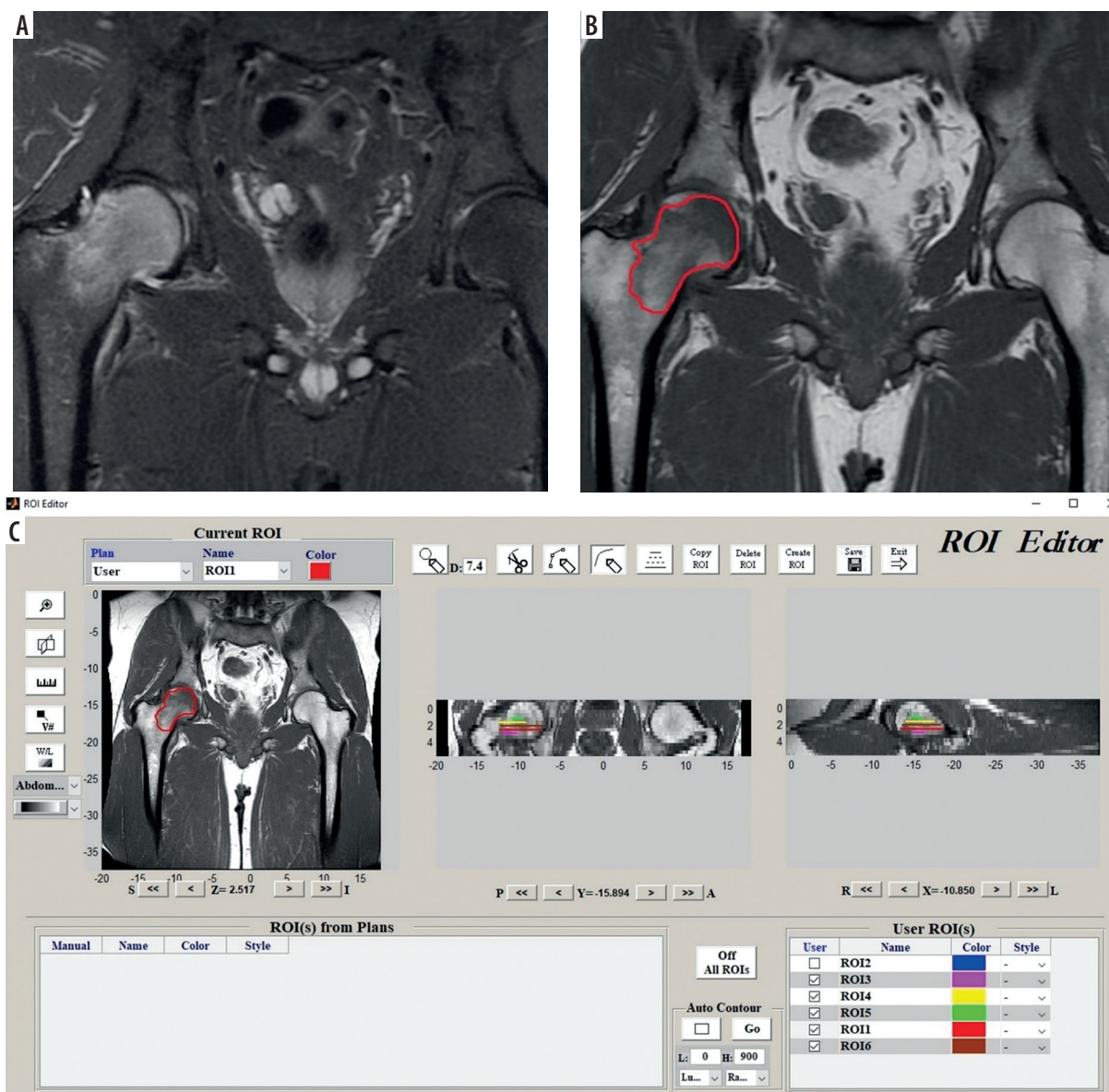


Figure 2. Depiction of segmentation process. A) Coronal STIR image, which is not primarily used for segmentation but is used for guiding owing to conspicuity of marrow oedema in STIR images. B) Coronal T1 image that is segmented. C) Software interface during segmentation process. Please note depiction of multiple ROIs in coronal and sagittal images, which are serrated because the images are not volumetric

The ability of the parameters to discriminate between the 2 groups was evaluated by receiver operating characteristic (ROC) analysis, calculating the area under the curve (AUC), sensitivity, and specificity. AUC values were reported with 95% confidence intervals and comparison to the AUC: 0.5 line. Sensitivity and specificity values were calculated for optimal thresholds, which were determined according to the Youden index.

Intraobserver reproducibility was assessed by Intraclass Correlation Coefficient (ICC). $ICC \geq 0.75$ was accepted as the cut-off for reproducibility [11].

Analyses were made using IBM SPSS v23 (IBM Corp, Armonk, NY), MedCalc v14.8.1 (MedCalc Software bvba, Ostend, Belgium), and XLStat statistical and data analysis add-on 2020.3.1 (Addinsoft, NY, USA) for Microsoft Excel 16.0.13029.

$P < 0.05$ was accepted as significant for all analyses.

Figure 3 summarizes the flowchart of the study.

Machine learning

Support vector machine (SVM) [12] and random forest classifier (RFC) analysis [13] for machine learning were used.

Support vector machine

Parameters for SVM are as follows; power kernel, C value: 1.0, Tolerance: 0.1. Because radiomics parameter values have a considerable range, standardisation of values was performed as preprocessing. The study population was classified into training and validation sets regarding a ratio of 70:30. To diminish effects of variability and selection bias on estimates of performance, 1000-fold bootstrapping was performed similarly to the method proposed by Vrigazova *et al.* [14]. Area under the curve (AUC), and sensitivity and specificity of SVM were reported as mean with 95% confidence intervals.

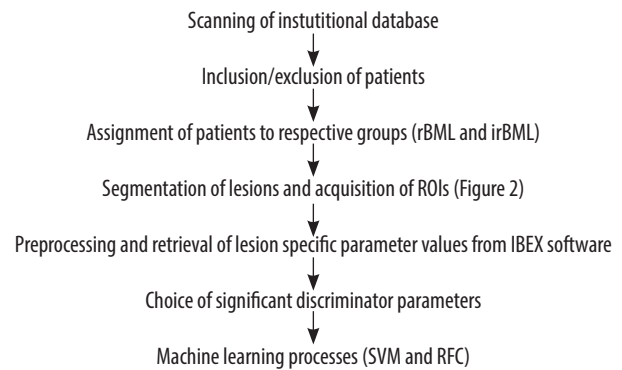
Random forest classifier

Parameters for RFC are as follows; method: bagging, sampling method: random with replacement, number of trees built: 1000. Maximum depth: 20. Diagnostic performance of RFC was reported with sensitivity, specificity, AUC, misclassification rate. An out-of-bag error evolution chart was supplied as well.

Results

Study population

A total of 37 patients were included. Seventeen demonstrated the aforementioned ON-related changes in follow-up images and were thus included in the irreversible group. Twenty demonstrated resolution of bone marrow



rBML – reversible bone marrow lesions, irBML – irreversible bone marrow lesions, SVM – support vector machine, RFC – random forest classification

Figure 3. Flowchart of the study

oedema and thus were included in the reversible group. A total of 185 ROIs were attained from this population (mean = 5, range = 2-9 ROIs per patient). In total, 104 ROIs belonged to the reversible group and 81 ROIs belonged to irreversible group. Ten (27%) patients were female. The mean age was 45.7 ± 12 years. There was no significant difference regarding distribution of sex ($p = 0.658$) amongst the 2 groups. The reversible group was significantly younger (43.2 ± 7.1 vs. 49.6 ± 6.3 , $p = 0.028$). Ten patients from the irreversible group and 10 patients from the reversible group had subchondral fractures that were visible on MR images. The mean interval between first and second examinations was 8.14 ± 0.36 weeks.

Inter-observer and Intra-observer reproducibility

Ten patients who produced 40 ROIs per observer during the segmentation process were chosen randomly. Then mean and range parameters from the first-order group were utilized for assessment of inter-observer correlation. According to both parameters, there was a significant correlation between the 2 observers (mean: $\rho = 0.886$, $p = 0.02$; range: $\rho = 0.863$, $p = 0.024$).

The same randomly chosen patients were used for assessment of intra-observer reproducibility. Two observers performed segmentation process twice. Then the “mean” and “range” parameters were utilized. According to both parameters, there was good intraobserver agreement for both observers. (Observer 1: mean ICC: 0.845 [range: 0.761-0.937], Observer 2: mean ICC: 0.867 [range: 0.777-0.940]).

Diagnostic performance of individual parameters

A total of 47 parameters had an AUC value significantly different from AUC:0.5 and were thus accepted as classifiers. AUC values ranged between 0.586 and 0.718. Their respective ROC values are presented in Table 1.

Table 1. ROC characteristics of individual parameters.

	AUC (95% CI)	Sensitivity	Specificity	Cut off ^{f1,2}
GLCM				
AutoCorrelation	0.693 (0.621-0.729)	0.49	0.88	≤ 370
ClusterProminence	0.649 (0.575-0.717)	0.56	0.74	≤ 33077
ClusterShade	0.589 (0.514-0.660)	0.75	0.52	> -190
ClusterTendency	0.652 (0.579-0.720)	0.54	0.79	≤ 111
Correlation	0.695 (0.624-0.761)	0.86	0.47	≤ 0.68
GLCMEntropy	0.617 (0.542-0.687)	0.60	0.64	≤ 7.9
Homogeneity2	0.586 (0.511-0.658)	0.60	0.61	> 0.25
InformationMeasureCorr1	0.653 (0.580-0.722)	0.37	0.90	> -0.23
InformationMeasureCorr2	0.666 (0.594-0.734)	0.55	0.71	≤ 0.93
InverseDiffMomentNorm	0.607 (0.532-0.677)	0.69	0.55	> 0.99
InverseDiffNorm	0.601 (0.526-0.672)	0.71	0.53	> 0.93
SumAverage	0.690 (0.618-0.756)	0.43	0.93	≤ 34
SumEntropy	0.649 (0.576-0.718)	0.54	0.75	≤ 5.23
SumVariance	0.695 (0.623-0.760)	0.50	0.87	≤ 1163
GLCMVariance	0.652 (0.579-0.720)	0.53	0.79	≤ 111
GLRLM				
HighGrayLevelRunEmphasis	0.692 (0.620-0.758)	0.50	0.86	≤ 403
LongRunHighGrayLevelEmphasis	0.705 (0.634-0.770)	0.53	0.90	≤ 1120
LongRunLowGrayLevelEmphasis	0.710 (0.639-0.774)	0.46	0.93	> 0.01
LowGrayLevelRunEmphasis	0.709 (0.637-0.773)	0.46	0.97	> 0.005
RunLengthNonUniformity	0.622 (0.548-0.692)	0.78	0.50	≤ 752
ShortRunEmphasis	0.594 (0.520-0.666)	0.94	0.27	≤ 0.93
ShortRunHighGrayLevelEmphasis	0.691 (0.619-0.756)	0.41	0.96	≤ 224
ShortRunLowGrayLevelEmphasis	0.714 (0.643-0.778)	0.47	0.97	> 0.003
First Order				
Energy	0.681 (0.609-0.747)	0.60	0.78	≤ 308210827
GlobalEntropy	0.613 (0.539-0.683)	0.51	0.75	≤ 4.8
GlobalMax	0.657 (0.584-0.725)	0.46	0.89	≤ 591
GlobalMean	0.690 (0.618-0.756)	0.50	0.87	≤ 379
GlobalMedian	0.691 (0.619-0.757)	0.49	0.89	≤ 375
GlobalMin	0.718 (0.647-0.782)	0.53	0.88	≤ 144
GlobalStd	0.629 (0.555-0.698)	0.53	0.76	≤ 130
GlobalUniformity	0.613 (0.539-0.684)	0.55	0.71	> 0.03
Kurtosis	0.616 (0.542-0.686)	0.70	0.53	> 1.95
LocalEntropyMax	0.600 (0.525-0.671)	0.42	0.83	≤ 4.5
LocalEntropyStd	0.636 (0.562-0.705)	0.76	0.50	≤ 0.42
MeanAbsoluteDeviation	0.630 (0.556-0.700)	0.51	0.77	≤ 105
10Percentile	0.712 (0.641-0.776)	0.43	0.97	≤ 179
25Percentile	0.701 (0.630-0.766)	0.43	0.97	≤ 240
50Percentile	0.691 (0.619-0.757)	0.50	0.89	≤ 375
75Percentile	0.685 (0.612-0.751)	0.48	0.89	≤ 471
90Percentile	0.676 (0.604-0.743)	0.52	0.85	≤ 569

Table 1. Cont.

	AUC (95% CI)	Sensitivity	Specificity	Cut off ^{1,2}
Range	0.610 (0.535-0.680)	0.50	0.78	≤ 487
RootMeanSquare	0.687 (0.615-0.753)	0.48	0.89	≤ 396
Skewness	0.623 (0.548-0.693)	0.46	0.78	> 0.10
Variance	0.629 (0.555-0.698)	0.53	0.77	≤ 16975
NGTDM				
Busyness	0.682 (0.610-0.748)	0.52	0.86	> 0.024
Complexity	0.597 (0.522-0.688)	0.40	0.84	≤ 67825
TextureStrength	0.618 (0.544-0.688)	0.54	0.68	≤ 236

¹For diagnosis of irreversible bone marrow lesions

²Determined according to Youden index

Diagnostic performance of machine learning

Support vector machine

Forty-seven parameters, listed in Table 1, were supplied to SVM. Sensitivity was found to be 91.3% (85.6-96.2%), and specificity was found to be 85.1% (80.1-89.7%) in the validation set. The AUC was 0.921 (95% CI: 0.888-0.964).

Random forest classifier

In total, 1000 trees were built with the aforementioned 47 parameters. The misclassification rate was 0.189. Sensitivity was found to be 84.8% (79.1-89.3%), and specificity was found to be 76.7% (70.3-82.4%). AUC was 0.811 (95% CI: 0.763-0.868). The out-of-bag (OOB) error evolution chart is shown in Figure 4.

Discussion

To the best of our knowledge, this is the only study that evaluates the diagnostic accuracy of radiomics in the

discrimination of reversible and irreversible BMLs of the hip before irreversible changes occur.

TBMES is a transient, self-limiting condition without a known aetiology. Although some mechanisms such as neurogenic and vascular hypotheses have been proposed, the exact pathogenesis remains uncertain [15-17]. Clinical presentation of TBMES includes a sudden-onset hip pain, effecting mostly males between the 4th and 6th decade of life and rarely females in the 3rd trimester of pregnancy [18]. The treatment approach mostly includes partial weight-bearing and administration of analgesics and anti-inflammatory agents [18,19]. In line with this information, we thought that all the reversible BMLs in our study were TBMES. Contrarily, ON is a more common aetiology of acute hip pain than TBMES. However, ON is a progressive disorder mostly due to secondary local or systemic conditions. Therefore, early surgical intervention can result in joint preservation [18].

In current clinical practice, MRI findings describing bone marrow oedema in TBMES are low signal intensity on T1-weighted images, high signal intensity on short-tau inversion recovery (STIR) and fat-suppressed T2-weighted

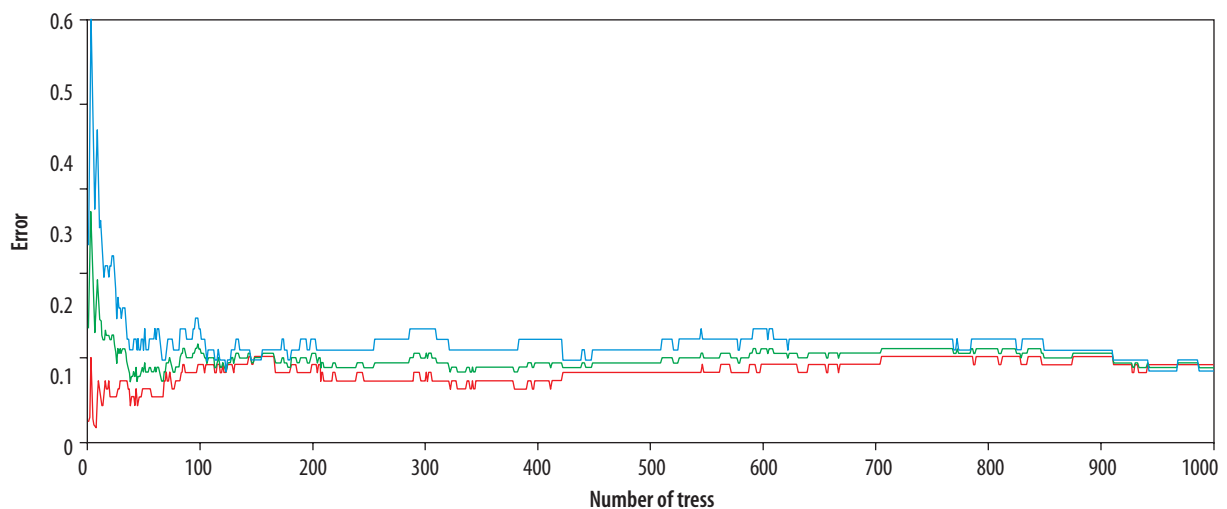


Figure 4. OOB evolution chart of RFC process X axis: number of trees generated, Y axis: error rate; red line represents diagnosis of reversible bone marrow lesion, blue line represents diagnosis of irreversible bone marrow lesion, and green line represents OOB errors

images, and delayed enhancement in contrast-enhanced images with the lack of additional subchondral changes [20–22]. The existence of subchondral low signal intensity regions > 4 mm thick and > 12.5 mm long on T2-weighted and contrast-enhanced T1-weighted images has been found to be associated with irreversible lesions such as ON. A subchondral low signal intensity band-like lesion on T1-weighted images and the double-line sign seen on non-fat-suppressed T2-weighted SE or TSE images have been highly diagnostic criteria for ON [5,6]. Recently, quantitative dynamic contrast-enhanced MRI with high temporal resolution demonstrated promising results in differentiating TBMES from ON [23]. Nonetheless, the differential diagnosis is sometimes challenging because bone marrow oedema is a common condition with both disorders, and imaging findings can be unclear.

Diagnostic performance of individual parameters was subpar in the present study. The highest AUCs that could be achieved were 0.718 (GlobalMin), 0.714 (ShortRunLowGreyLevelEmphasis), 0.712 (10Percentile), 0.710 (LongRunLowGreyLevelEmphasis), and 0.709 (LowGreyLevelRunEmphasis). GlobalMin and 10Percentile are first-order parameters which state that more hypointense voxels are abundant in the tissue that was sampled. In the case of T1 images, this could be translated as more non-fat voxels. Lower values of these parameters suggested a diagnosis of irreversible BMLs, and consequently more non-fat areas. The remainder of the aforementioned parameters belong to the GLRLM group. GLRLM parameters assess uniformity of neighbouring voxels that run along vectors. Thus, bigger values suggest voxels with similar intensities lie along a longer vector, which could be translated as a more homogenous structure. In our study, we found that ON images had lower values, which depict heterogenous organisation of voxel intensities. In conclusion, based on these findings, we speculate that BML of ON contains non-fat areas such as blood or necrotic materials contrary to TBMES, which lacks these components.

A recent study that assessed utility of radiomics and interpretation of conventional MR images by radiologists reported an AUC value up to 93.7% for machine learning techniques and 90.6% for musculoskeletal radiologists [24]. The authors reported that the diagnostic performance of radiomics was similar to MSK radiologists. However, points of the paper are well taken, included ON cases demonstrating signs such as band-like sign, single/double line signs, and crescent signs, which state that articular damage has already settled to some extent. Moreover, images were acquired with either 1.5T or 3T scanners. Because the field strength affects the signal properties of voxels, contrary to previous studies, we aimed for the utility of standardized and uniform hardware in order to eliminate the effect of field strength on the radiomics parameters.

Even though it has been suggested by some authors that BML is a late manifestation of ON which is secondary

to subchondral fractures [18,25], our findings were not concordant with these statements. Despite showing BML on MR images, 7 of the 17 ON patients lacked a subchondral fracture on CT. Interestingly, 10 of 20 TBMES patients had a subchondral fracture, as well. There was no significant difference regarding parameters when patients with subchondral fractures were compared to those without visible fractures within their respective group (Table 2). Thus, we concluded that the presence of subchondral fractures does not cause a significant alteration in parameter measurements.

As stated before, propagation of BML in ON patients has been a dilemma [23,24,26,27], and we refrain from commenting further on this topic. Furthermore, it is worth noting that the presence of a subchondral fracture did not affect the radiomics parameters. The main purpose of our work is to suggest a ML-based model that differentiates irreversible BMLs from reversible BMLs before articular damage occurs.

The retrospective design of our study could be considered as a limitation. Second, due to partial missing data regarding patient characteristics, we failed to implement a clinical model that could have included known risk factors for ON such as history of corticosteroids, bisphosphonate, vasculitis, and chronic alcohol consumption. Throughout the study, we were careful to use the definition of irreversible or reversible BMLs instead of ON or TBMES as much as possible due to different opinions in the literature. This might be another limitation, but our aim in this study was to interpret the prognosis of BMLs observed at femoral head with ML, not to make a pathological diagnosis or staging the disease. Third, due to the limited sample size, we opted for an internal validation process. Notwithstanding that external validation produces more robust and generalizable results, our sample size was far from the minimum (at least 100 samples) or optimal (at least 200 samples) that were suggested in the current literature [28]. Also, we used 2D segmentation because the routine imaging protocol in our institution includes non-volumetric images only. This precluded utility of shape- and size-based radiomics parameters. Finally, evaluation of subchondral fractures with MR images could also be accepted as a limitation because CT has been shown to be more sensitive regarding this issue [18].

Conclusions

Our findings show that radiomics analysis could prove useful for discrimination of reversible and irreversible BMLs. While we achieved an AUC value of 0.921 for SVM and 0.892 for RFC, the specificity was rather low (85.1% for SVM and 76.7% for RFC). Thus, we are far from the recommendation of surgical intervention based solely on radiomics. Studies with larger sample sizes and predefined 3D protocols and models combined with clinical

Table 2. Comparison of parameters between patients with and without subchondral fractures

	<i>p</i> -value (irreversible group) ^a	<i>p</i> -value (reversible group) ^b
GLCM		
AutoCorrelation	0.735	0.318
ClusterProminence	0.813	0.614
ClusterShade	0.457	0.589
ClusterTendency	0.632	0.248
Correlation	0.545	0.075
GLCMEntropy	0.589	0.789
Homogeneity2	0.891	0.108
InformationMeasureCorr1	1	0.918
InformationMeasureCorr2	1	0.935
InverseDiffMomentNorm	0.932	0.745
InverseDiffNorm	0.857	0.842
SumAverage	0.884	0.773
SumEntropy	0.547	0.764
SumVariance	0.689	0.639
GLCMVariance	0.679	0.479
GLRLM		
HighGrayLevelRunEmphasis	0.914	0.843
LongRunHighGrayLevelEmphasis	0.905	0.792
LongRunLowGrayLevelEmphasis	0.933	0.682
LowGrayLevelRunEmphasis	0.897	0.541
RunLengthNonUniformity	0.894	0.167
ShortRunEmphasis	0.868	0.091
ShortRunHighGrayLevelEmphasis	0.892	0.921
ShortRunLowGrayLevelEmphasis	0.911	0.854
First Order		
Energy	0.734	0.121
GlobalEntropy	0.495	0.621
GlobalMax	0.624	0.543
GlobalMean	0.598	0.761
GlobalMedian	0.651	0.813
GlobalMin	0.647	0.772
GlobalStd	0.716	0.542
GlobalUniformity	0.794	0.617
Kurtosis	0.567	0.389
LocalEntropyMax	0.738	0.846
LocalEntropyStd	0.659	0.095
MeanAbsoluteDeviation	0.638	0.971
10Percentile	0.915	0.745
25Percentile	0.928	0.729
50Percentile	0.944	0.851
75Percentile	0.931	0.768
90Percentile	0.908	0.698
Range	0.892	0.647
RootMeanSquare	0.759	0.597
Skewness	0.641	0.791
Variance	0.539	0.613
NGTDM		
Busyness	0.761	0.478
Complexity	0.441	0.543
TextureStrength	0.496	0.591

^aComparison of irreversible bone marrow lesion patients with subchondral fractures and without subchondral fractures. Mann-Whitney *U* test.

^bComparison of reversible bone marrow lesion patients with subchondral fractures and without subchondral fractures. Mann-Whitney *U* test.

cal features could provide better diagnostic accuracy and guide the decision-making process for clinical management of patients with BMLs of the hip.

Conflict of interest

The authors report no conflict of interest.

References

- Wilson A, Murphy W, Hardy D, Totty W. Transient osteoporosis: transient bone marrow edema? *Radiology* 1988; 167: 757-760.
- Bonadio MB, Ormond Filho AG, Helito CP, et al. Bone marrow lesion: image, clinical presentation, and treatment. *Magnetic Resonance Insights* 2017; 10: 1178623X17703382. doi: 10.1177/1178623x17703382.
- Eriksen EF. Treatment of bone marrow lesions (bone marrow edema). *Bonekey Rep* 2015; 4: 755. doi: 10.1038/bonekey.2015.124.
- Thiryayi W, Thiryayi S, Freemont A. Histopathological perspective on bone marrow oedema, reactive bone change and haemorrhage. *Eur J Radiol* 2008; 67: 62-67.
- Mitchell D, Rao V, Dalinka M, et al. Femoral head avascular necrosis: correlation of MR imaging, radiographic staging, radionuclide imaging, and clinical findings. *Radiology* 1987; 162: 709-715.
- Malizos KN, Karantanas AH, Varitimidis SE, et al. Osteonecrosis of the femoral head: etiology, imaging and treatment. *Eur J Radiol* 2007; 63: 16-28.

7. Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture analysis of imaging: what radiologists need to know. *Am J Roentgenol* 2019; 212: 520-528.
8. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol* 2019; 25: 485-495.
9. Gülbay M, Özbay BO, Mendi BAR, et al. A CT radiomics analysis of COVID-19-related ground-glass opacities and consolidation: is it valuable in a differential diagnosis with other atypical pneumonias? *PLoS One* 2021; 16: e0246582. doi: 10.1371/journal.pone.0246582.
10. Juntu J, Sijbers J, De Backer S, et al. Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. *J Magn Res Imaging* 2010; 31: 680-689.
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155-163.
12. Borstelmann SM. Machine learning principles for radiology investigators. *Acad Radiol* 2020; 27: 13-25.
13. Breiman L. Random forests. *Machine Learning* 2001; 45: 5-32.
14. Vrigazova B, Ivanov I. Tenfold bootstrap procedure for support vector machines. *Comput Sci* 2020; 21.
15. Curtiss PH Jr, Kincaid WE. Transitory demineralization of the hip in pregnancy: a report of three cases. *JBJS* 1959; 41: 1327-1333.
16. Rosen RA. Transitory demineralization of the femoral head. *Radiology* 1970; 94: 509-512.
17. Lequesne M. Transient osteoporosis of the hip. A nontraumatic variety of Südeck's atrophy. *Ann Rheum Dis* 1968; 27: 463-471. doi: 10.1136/ard.27.5.463.
18. Karantanas AH, Drakonaki EE. The role of MR imaging in avascular necrosis of the femoral head. *Semin Musculoskelet Radiol* 2011; 15: 281-300.
19. Korompilias AV, Karantanas AH, Lykissas MG, Beris AE. Bone marrow edema syndrome. *Skeletal Radiol* 2009; 38: 425-436.
20. Malizos KN, Zibis AH, Dailiana Z, et al. MR imaging findings in transient osteoporosis of the hip. *Eur J Radiol* 2004; 50: 238-244.
21. Vande Berg BC, Malghem JJ, Lecouvet FE, et al. Idiopathic bone marrow edema lesions of the femoral head: predictive value of MR imaging findings. *Radiology* 1999; 212: 527-535.
22. Eustace S, Keogh C, Blake M, et al. MR imaging of bone oedema: mechanisms and interpretation: pictorial review. *Clin Radiol* 2001; 56: 4-12.
23. Geith T, Stellwag AC, Müller PE, et al. Is bone marrow edema syndrome a precursor of hip or knee osteonecrosis? Results of 49 patients and review of the literature. *Diagn Interv Radiol* 2020; 26: 355-362.
24. Klontzas ME, Manikis GC, Nikiforaki K, et al. Radiomics and machine learning can differentiate transient osteoporosis from avascular necrosis of the hip. *Diagnostics (Basel)* 2021; 11: 1686. doi: 10.3390/diagnostics11091686.
25. Meier R, Kraus TM, Schaeffeler C, et al. Bone marrow oedema on MR imaging indicates ARCO stage 3 disease in patients with AVN of the femoral head. *Eur Radiol* 2014; 24: 2271-2278.
26. Klontzas ME, Zibis AH, Vassalou EE, Karantanas AH. MRI of the hip: current concepts on bone marrow oedema. *Hip Int* 2017; 27: 329-335.
27. Trevisan C, Ortolani S, Monteleone M, Marinoni E. Regional migratory osteoporosis: a pathogenetic hypothesis based on three cases and a review of the literature. *Clin Rheumatol* 2002; 21: 418-425.
28. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016; 35: 214-226.