


Validation of a Novel Semiautomated Segmentation Method for MRI Detection of Cartilage-Related Bone Marrow Lesions

Cartilage
1(4) 328–334
© The Author(s) 2010
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1947603510376819
http://cart.sagepub.com


A.J. Dijkstra¹, P. Anbeek², K.G. Auw Yang¹, K.L. Vincken²,
M.A. Viergever², R.M. Castelein¹, and D.B.F. Saris¹

Abstract

Objective: To determine the relationship of bone marrow lesions (BMLs) with phenomena such as clinical symptoms, histological subchondral bone damage, and development of osteoarthritis, a reliable and reproducible method to localize and quantify BMLs accurately is indispensable. Therefore, the goal of the current study was to develop and validate a novel semiautomated segmentation method based on the KNN classification technique on T2-weighted (T2w) SPIR and proton density-weighted (PDw) magnetic resonance images (MRIs), as this would provide an accurate, reliable, and reproducible tool. **Materials and Methods:** Twenty PDw and T2w SPIR MRIs were selected and manually segmented as a learning set for the software system. The manual segmentations were considered the gold standard. Automated segmentation based on the KNN classification technique was carried out on the same MRIs. To determine the accuracy and validity of the system, the automated segmentations were compared to the gold standard using the Dice Similarity Index (DSI). **Results:** The KNN classification system resulted both visually and statistically in an accurate segmentation of BMLs on T2w SPIR MRIs with an excellent mean optimal DSI of 0.702 (± 0.202 ; range, 0.409–0.908). Elimination of specific areas smaller than 10 voxels improved the accuracy. The accuracy was independent of BML size. The segmentation of BMLs on PDw MRIs was less reliable with a mean optimal DSI of 0.536 (± 0.156). **Conclusion:** Although the applicability of this method is limited on PDw MRIs, the KNN classification system provides an accurate, reliable, and reproducible tool for semiautomated segmentation of BMLs in T2w SPIR MRIs of the knee.

Keywords

bone marrow lesions, bone marrow edema, segmentation, quantitative method, cartilage, MRI, KNN classification

Introduction

Magnetic resonance imaging (MRI) is a valuable diagnostic tool for detecting and evaluating intra-articular pathology such as articular cartilage defects, meniscal lesions, and cruciate ligament ruptures. However, it remains difficult to uniformly evaluate aspects such as subchondral bone marrow lesions (BMLs).

BMLs are characterized by ill-defined hypointense zones on T1-weighted and proton density-weighted (PDw) MRIs, and ill-defined hyperintense zones on T2-weighted (T2w) fat-suppressed MRIs in the subchondral cancellous bone.¹ Histological studies of these subchondral bone areas have shown trabecular abnormalities, bone marrow necrosis, and fibrosis rather than bone marrow edema alone.

BMLs are unspecific and may be observed in conjunction with various disorders of the knee. For example, they are observed in the knee joints following trauma, with and without the occurrence of concomitant lesions such as articular

cartilage defects or ligamentous injuries. Furthermore, they can occur as a result of osteonecrosis, transient bone marrow edema syndrome, arthritis, osteoarthritis (OA), complex regional pain syndrome, and in tumors.²

In OA research, BMLs have been related to the progression of OA and are suggested to be related to clinical symptoms, such as pain.^{3–6} However, few studies are conclusive, and some are contradictory. This can be attributed in part to timing of MRI as well as to the fact that there is no gold standard for determining BMLs.

¹Department of Orthopaedics, University Medical Center Utrecht, the Netherlands

²Image Sciences Institute, Department of Radiology, University Medical Center Utrecht, the Netherlands

Corresponding Author:

D.B.F. Saris, Department of Orthopaedics, University Medical Center, Utrecht, PO Box 85500, 3508 GA Utrecht, the Netherlands
Email: d.saris@umcutrecht.nl

BML size has been assessed using semiquantitative as well as quantitative methods.⁷⁻¹⁰ Only a few methods for quantitative volumetric analysis have been validated. These methods use either manual or semiautomated segmentation of BML volume from surrounding normal bone marrow.^{7,11} Recently, a new method, based on the K-Nearest Neighbor (KNN) classification technique, was developed for fully automated segmentation of white matter lesions (WMLs) on cranial MRIs.¹² Both BMLs and WMLs appear as ill-defined hyperintense zones on the T2w image that gradually converts into normal tissue. Furthermore, BMLs as well as WMLs are restricted to certain areas on MRI, which is relevant for the detection method. Because of the similarities between the MRI aspect of BML and WML, it was our hypothesis that this MRI detection method and automated segmentation could also provide a useful tool for accurate determination of BML volumes in the knee.

Therefore, the goal of the current study was to develop a reliable and reproducible tool that is independent of the clinical interpretation of investigators and to validate such a novel semiautomated segmentation method based on the KNN classification technique to determine location and volume of a BML accurately on T2w SPIR and PDw MRIs.

Materials and Methods

Data Acquisition

The proposed segmentation method is based on the KNN classification, which is a statistical pattern recognition method. It uses a training set of MRI scans that are representative for the whole study group. Twenty MRI scans of the knee that show a BML from patients with solitary cartilage defects of the femoral condyle were selected as a training set and to validate the method. The MRIs were acquired, using the same protocol, preoperatively and at 6, 12, 24, or 36 months postoperatively after autologous chondrocyte implantation or microfracturing. The MRIs were selected based on spatial characteristics of the BML and the lesion volume with a considerable and clinically realistic variation in lesion size, such that they would represent the whole variety of lesions as much as possible.

MRI studies were performed on a Philips Gyroscan ACS-NT 1.5 Tesla whole-body system (Philips Medical Systems, Best, the Netherlands). All patients had the same magnetic resonance protocol of the knee consisting of a sagittal PDw scan without fat suppression and a sagittal T2w SPIR scan. All scans were performed with a 3-mm slice thickness, 0.3-mm slice gap, 24 slices, 0.29×0.29 in-plane resolution. The individual scan parameters were PD: TR/TE 2243/35 ms and T2: TR/TE 4084/70 ms. The entire acquisition time per session was less than 20 minutes. The only difference in acquisition of the MRI scans

was the use of different knee coils. In 9 MRIs, the Philips 1.5T Sense knee coil 8 elements was used; in 10 MRIs, the Philips 1.5T knee coil 3 elements was used; and in 1 patient, the Philips 1.5T Sense Flex M knee coil was used.

Manual Segmentation

Bone marrow edema lesions of the femoral condyle were manually segmented. The lesions were segmented on the T2w SPIR and the PDw image separately, resulting in different manual segmentations for the 2 images. The segmentations were composed using the defined standard procedure: In each slice containing a BML, a threshold on the signal intensity was set, such that the entire lesion was included, resulting in a binary image containing the BML and other tissue. Next, all other tissue that did not belong to the BML was discarded manually from this image. The final BML segmentations were evaluated in a consensus meeting by 3 investigators (an orthopedic surgeon, an orthopedic resident, and a medical student who was trained in recognizing BMLs on MRIs in advance of conducting the current study). To guarantee the quality of the manual segmentations, a musculoskeletal radiologist was asked to look at the manual segmentations afterward. This was considered the gold standard against which the automated segmentation was tested.

Mask Generation

Masks were created on the PDw images to define the region of interest for the segmentation. We used only the masks created on PDw images because the border between cartilage and subchondral bone was more clearly visible on PDw images than on T2w SPIR images. Only slices with a visible BML were included in the mask. In these slices, the exact boundaries of the femoral condyle were defined, and the region within these boundaries was considered as the mask. Comparison by an image-processing software tool showed that the PDw and T2w images were spatially similar in almost all cases because the patients had not moved during scanning. In the cases in which PDw and T2w images did not agree spatially, a rigid co-registration step (only translation), based on mutual information, was performed.

The major goal of these masks was to increase accuracy in the segmentation because only voxels in the affected regions were taken into account. Furthermore, they reduce computation time and computer memory utilization substantially.

KNN Classification

The aim of the method for semiautomated segmentation of the BMLs was to determine the lesion probability of each

voxel. For this purpose, the KNN classification method was used: a nonparametric procedure for estimation of local class conditional probability density functions from sample patterns. In general, KNN classification is based on the classification of samples, dependent on their features. In the proposed method, each image voxel is treated as a separate sample. A feature space is defined, in which each axis represents one of the voxel features. A set of preclassified learning voxels is used as learning set. The feature space is filled with these voxels, at coordinates corresponding to their feature values. Subsequently, an image voxel of a new patient is classified by inspection of the number of K learning voxels that are closest in the feature space. The new voxel is then classified according to the classes of those K neighbors. In many applications of KNN classification, the most frequent class of the K neighbors could be assigned to it.

In our study, which used 20 learning patients, the learning set for segmentation of one patient was built from the voxels of the other 19 patients (the so-called leave-one-out method). All voxels in the learning set were labeled with the value of 0 (nonlesion class) or 1 (lesion class), derived from the manual segmentations. Due to limitations in computer memory, 60% of the voxels were randomly selected for inclusion in the learning set. The segmentations were performed separately on the T2 and the PD image of the patients. This implies that the segmentations used separate training sets: one composed from the T2 images, the other from the PD images.

Three features were used in this study and can be divided into 2 categories: voxel intensities and spatial information. The 1st feature is defined by the signal intensity of a voxel in the T2w SPIR or the PDw image. The 2nd group of features incorporates the spatial location of a voxel in the femoral condyle. These were added because in some regions of the femoral condyle, a BML lesion is more likely to occur than in others. Using the spatial features has the effect that not only the signal intensity but also the spatial location of a voxel determines the probability of its being a part of the lesion. To identify these spatial feature values, the images of the left knees were first mirrored with respect to the y-axis in order to make the right and left knees comparable. Then, the in-plane x- and y-coordinates of the voxel were determined and defined as the 2nd and 3rd features. This resulted in a KNN classification with a 3-dimensional feature space.

Because different features have different ranges, a rescaling of the feature space was necessary to define a proper metric to compare distances in the feature space, which is essential to justify classification based on KNN. In many applications, this is commonly done by variance scaling: subtraction of the mean of the feature values and division of the outcome by the standard deviation. This approach

results in a mean of 0 and variance of 1 for every feature. However, the occurrence of a BML in the femoral condyle highly influences the intensity distribution of the image voxels. A Gaussian distribution, which is assumed in variance scaling, is not applicable anymore. Therefore, we determined the mode, which was the intensity reflecting the top of the histogram, and the 1-sided variance in the histogram at the side that did not include the lesion voxels. In the T2 SPIR image, this was the left side of the histogram; in the PD-image, it was the right side. The features were scaled by subtraction of the voxel values with the mode and division by this variance.

The choice of K in KNN classification depends on the number of features and the number of cases. When a small value of K is used, the obtained results are more influenced by individual cases. A larger value of K smoothens the outcome of the classification.¹³ In this study, we used a relatively small number of features in combination with a large number of cases. Therefore, we opted for a relatively large K. Taking computation time into account, we concluded that 100 was an acceptable choice for K.

The lesion probability of every voxel was determined by inspection of the K nearest neighbors of the examined voxel in the feature space. It was defined as the fraction of lesion voxels among those K neighbors. The voxel probabilities were presented in a so-called probability map, which is an image where each voxel intensity value is defined by the lesion probability of that voxel.

Volume Calculation

Lesion volumes were calculated by multiplication of the number of BML voxels by the voxel size. The ratio was defined as the segmentation volume divided by the gold standard volume. A ratio larger than 1 resembled an over-segmentation with respect to the gold standard volume.

Evaluation and Statistics

By applying different thresholds on the probability map, binary segmentations of the BMLs were produced. These segmentations were compared with the gold standard, where the number of correctly classified voxels (i.e., the true positives [TP] and true negatives [TN]), was counted as well as the number of false positives (FP) and false negatives (FN). The true positive fraction (TPF), which is the sensitivity, and the false positive fraction (FPF), which is 1 – specificity, was calculated for the threshold, running from 0 to 1. They are defined by

$$TPF = \frac{TP}{TP + FN},$$

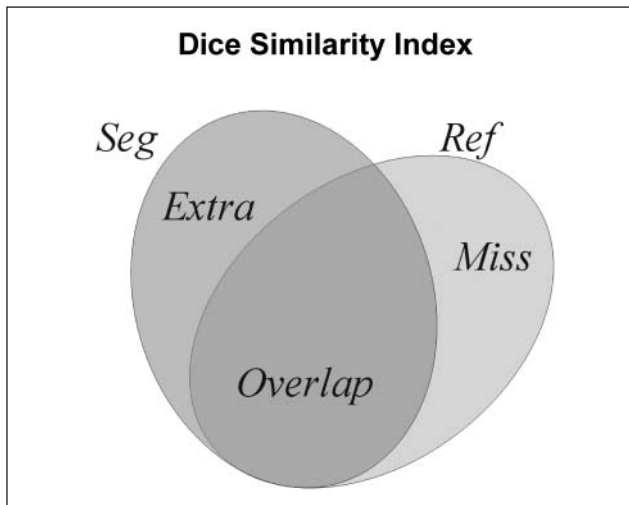


Figure 1. Comparison of a binary segmentation (*Seg*) with the manual segmentation (gold standard; *Ref*), with the correctly classified voxels (*Overlap*), the false positives (*Extra*), and the false negatives (*Miss*).

$$FPF = \frac{FP}{FP + TN} .$$

The TPF was represented in a receiver-operating characteristic (ROC) curve as a function of the FPF. Furthermore, the binary segmentations were evaluated with the Dice Similarity Index (DSI).¹⁴ The DSI is a measure for the correctly classified lesion area relative to the total area of the BML in both the reference (the gold standard) and the area of the segmented image and is applied in many segmentation studies, such as for brain tissue. The DSI is defined by

$$DSI = \frac{2 \times (Ref \cap Seg)}{Ref + Seg} .$$

In these definitions, *Ref* denotes the volume of the reference and *Seg* is the volume of the binary segmentation (**Fig. 1**). The DSI was represented in a graph as a function of the threshold, running from 0 to 1, for all feature sets. A high value represents better correlation with the reference, and 1 denotes that the segmentation equals the gold standard. In the literature, it is stated that a DSI value of 0.7 represents an excellent agreement.¹⁴ A series of binary segmentations was derived from the probability maps by applying thresholds, running from 0 to 1. The DSI was calculated for these binary segmentations, and the segmentation with the highest DSI was considered the optimal binary segmentation. The threshold corresponding to the optimal binary segmentation was called the optimal threshold. The optimal binary segmentations were adjusted, such that small groups of

voxels (blobs) in the image were discarded. This was done for blobs less than 5 and 10 voxels, resulting in 2 binary segmentations with minimum blob sizes of 5 and 10. For the optimal binary segmentation (before discarding smaller blobs), and the binary segmentations with blob size minimum 5 and 10, evaluation similarity measures were calculated as described above.

The correlation between the gold standard volume and the DSI of the segmentations with optimal threshold and minimal blobs of 5 and 10 voxels was determined by the Pearson correlation test. The gold standard volumes were compared with the segmentation volumes by the paired-samples *t*-test. Excel (Microsoft Corp., Redmond, WA) and SPSS (New York, NY) were used for statistical analysis, and the values shown are means \pm standard deviation.

Results

The 20 MRIs were derived from 19 patients. In 11 patients, the right knee was affected and scanned; in 7 patients, the left knee was affected; and in 1 patient, both knees were affected. The population average age was 32 years (± 8 ; range, 20-42); 5 patients were female, and 14 were male.

The KNN classification generated a probability map on which different thresholds were applied for the generation of the binary segmentation (**Fig. 2**). The mean optimal threshold appeared to be 0.27 for T2w SPIR and 0.29 for PDw MRIs (**Fig. 3**).

The KNN classification system resulted both visually and statistically in an accurate binary segmentation of BMLs on T2w SPIR MRIs with a mean optimal DSI of 0.702 (± 0.202 ; range, 0.409-0.908) using the mean optimal threshold (0.27), which represents an excellent agreement according to criteria formulated by Bartko (**Figs. 2 and 3**).¹⁵

Discarding of small blobs (e.g., segmented areas that contain less than 5 or 10 voxels) results in a higher mean DSI of 0.716 (± 0.198) and 0.721 (± 0.193), respectively. This shows that increasing the blob size until 10 voxels improves the accuracy of the T2w SPIR segmentations. Three MRIs had a low DSI value: 0.409, 0.401, and 0.076, respectively. Analysis of the signal-to-noise ratio (SNR) showed that these 3 MRIs had an SNR lower than 1.5, whereas the other MRIs had an SNR higher than 1.5 (**Fig. 4**). Leaving out the 3 MRIs with a low SNR results in a mean DSI of 0.752 (± 0.122). The ROC curve shows the relationship between the sensitivity and 1 – specificity for different thresholds for T2w SPIR and PDw segmentations (**Fig. 5**). With the optimal threshold of 0.27 for T2w SPIR, the sensitivity and specificity of the segmentations are 0.843 and 0.956, respectively. With the optimal threshold of 0.29 for PDw images, the sensitivity and specificity of the segmentations are 0.656 and 0.973, respectively.

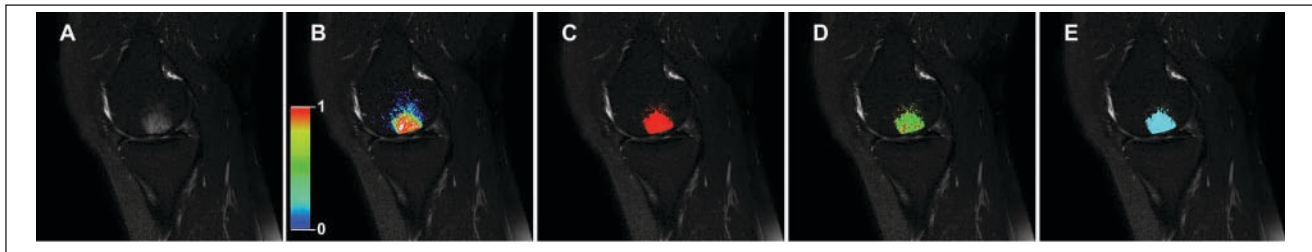


Figure 2. An example to illustrate the meaning of the probabilistic measures and to provide a better intuitive understanding of the KNN classification is presented. (A) Example of the original T2w SPIR magnetic resonance images (MRI) containing a bone marrow lesion (BML). Using the KNN classification, a probability map of the BML is generated (B). Accordingly, a binary segmentation is generated (red) after applying the mean optimal threshold (C). This segmentation can be compared with the gold standard (green; D). (E) Shows the result after discarding blobs of less than 10 voxels (cyan).

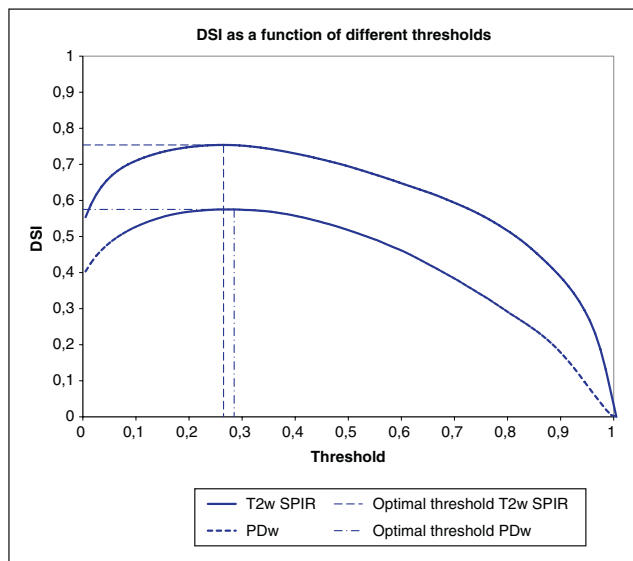


Figure 3. Dice Similarity Indices of binary bone marrow lesion (BML) segmentations of all patients for T2-weighted SPIR and proton density (PD)-weighted magnetic resonance images (MRIs) as function of the threshold.

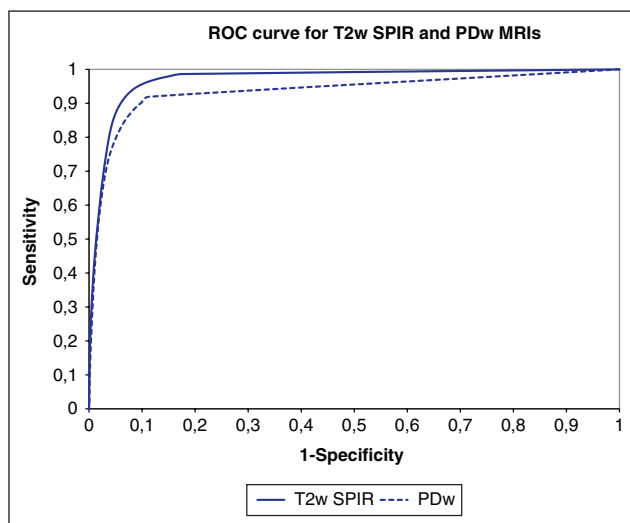


Figure 4. Dice Similarity Indices of binary bone marrow lesion (BML) segmentations of all patients for T2-weighted SPIR and proton density (PD)-weighted magnetic resonance images (MRIs) as function of the signal-to-noise ratio.

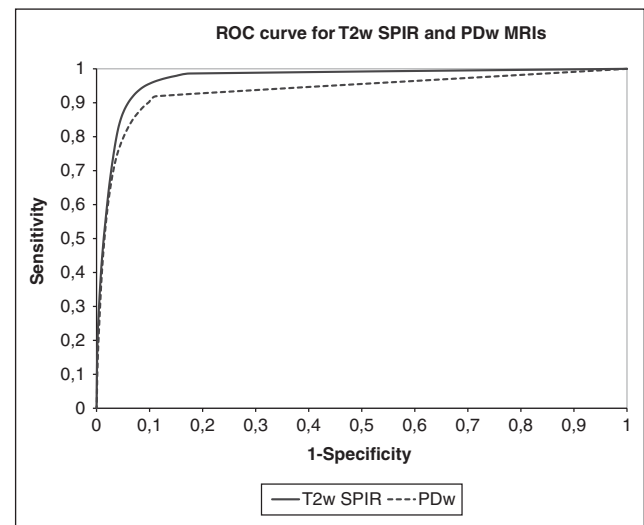


Figure 5. Receiver-operating characteristic (ROC) curves of classifications of all patients for T2w SPIR and proton density weighted (PDw) magnetic resonance images (MRIs) showing the relationship between the sensitivity and 1 – specificity for different thresholds.

Little or no association was observed between the gold standard volumes and the DSI (Fig. 6). Pearson correlation coefficients in the T2w-SPIR segmentation were 0.29 for the correlation with the optimal DSI, 0.31 for DSI with minimal blob size 5, and 0.29 for the DSI with minimal blob size 10. No significant differences were observed between the gold standard volumes and the segmented volumes by the paired-samples *t*-test ($P > 0.05$) for all segmentations.

The binary segmentation of BMLs by the software system on PDw MRI scans was less reliable, with a mean optimal DSI of $0.536 (\pm 0.156)$; Fig. 5).

Discussion

The goal of our study was to develop and validate a semiautomated segmentation method based on the KNN classification to determine BMLs on T2w SPIR and PDw MRI scans.

The combination of spatial information and signal intensities of MRIs in KNN classification provides an accurate,

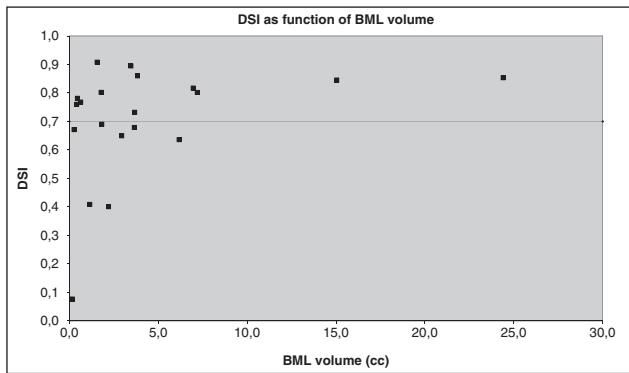


Figure 6. Dice Similarity Indices of binary bone marrow lesion (BML) segmentations of all patients for T2w SPIR as function of the BML volume (cc).

reliable, and reproducible tool for semiautomated segmentation of BMLs of the femoral condyles in T2w SPIR MRI scans of the knee with a high sensitivity and specificity and excellent DSI scores. Increasing the blob size to discard until 10 voxels even improves accuracy of the segmentations, which may be explained by the assumption that individual aspecific voxels should be eliminated from the segmentation, as BMLs on T2w SPIR MRI scans are represented by continuous lesions rather than individual abnormal voxels.

The Pearson correlation coefficients show that the accuracy of the method is independent of BML size. This means that the proposed method is applicable for all BML volumes.

In 3 MRI scans, the software system was unreliable; this may be explained by the fact that these scans showed a significant lower SNR compared to the other MRI scans, which may be influenced by operator-dependent and non-operator-dependent factors. The non-operator-dependent factors include field strength of the magnet and intrinsic molecular structure of the tissue being examined. Operator-dependent factors include the type of coil, the field of view, the number of acquisitions, sampling band width, matrix size, slice thickness, and the TR, TE, and flip angle. Of these operator-dependent factors, the type of coil was the only parameter that was variable among the MRI scans. The 3 unreliable MRI scans were acquired on the Philips 1.5T 3 elements knee coil, which suggests a significant factor of the coil type on the reliability of the system. On the other hand, various other MRI scans using this coil had good results. More likely, the lower SNR is caused by non-operator-dependent factors, such as obesity or clothing. A lower SNR likely results in a diffuse oversegmentation, which can be easily differentiated from the BML by visual inspection and should be excluded for automated segmentation in future studies.

To identify which images will result in suitable segmentations, we recommend performing an SNR analysis of the

image. The SNR is defined by $SNR = \text{mean}/\text{standard deviation}$, where mean and standard deviation denote the mean and standard deviation of the signal intensities of the background voxels (i.e., all voxels inside the mask that are non-BML). When the SNR is high enough (>1.5), the segmentation method performs well. A lower SNR (<1.5) results in a poorer segmentation.

In contrast to the excellent accuracy of the proposed method for the T2w SPIR MRI scans, the DSI found for segmentation of PDw MRIs without fat suppression demonstrated insufficient accuracy. The discrepancy in accuracy between T2w SPIR and PDw MRIs might be explained by the difference in grayscale contrast between the lesion and normal bone marrow, which is in favor of the T2w SPIR MRIs due to the application of fat suppression. In addition, on T2w SPIR MRIs, BMLs are much more homogeneous than on PDw MRIs. Therefore, it can be concluded that the proposed method is less suitable for segmentation of BMLs on PDw MRI scans without fat suppression.

The difficulty in segmenting BMLs lies in the fact that, on MRI, BMLs have a gradual transition into normal bone marrow. This makes segmentation methods prone to high interobserver and intraobserver variability. That is probably why most of the studies on BMLs used semiquantitative scoring methods instead of quantifying BMLs by segmentation methods. Only a few studies have validated a semi-automated segmentation method. Frobell *et al.* described and used a segmentation method quantifying BMLs, which still requires many manual steps. This makes their method time-consuming and more susceptible for higher intraobserver and interobserver variability.⁷ Mayerhoefer *et al.* described a computer-assisted quantitative analysis of BMLs of the knee that provides highly reproducible results and is largely observer independent.¹¹ However, the accuracy of this software system is not shown. Thereby, this method also requires manual interference to outline the contours of the examined bone marrow and to draw a region of interest, which makes the system highly time-consuming. Furthermore, this method is based only on grayscale intensity.

The KNN classification method, on the other hand, is not only based on grayscale intensity but also takes into account spatial features, which makes it less susceptible for segmentation of aspecific lesions at unlikely sites of the femoral condyle.

The reproducibility of this proposed method may be even further optimized by implementing the same learning set voxels rather than randomly selected learning set voxels for segmentation of the MRIs. Other variables influencing the reproducibility are the selection of those slices containing BMLs and the creation of a mask of that region.

The reproducibility of the proposed method is very high by definition, because this method is carried out almost

fully automatically and may be even further optimized by implementing the same learning set voxels rather than randomly selected learning set voxels for segmentation of the MRIs. Other variables are the selection of those slices containing BMLs and generating a mask of that region.

The proposed method is carried out almost fully automatically, which also makes it much less time-consuming than previously described semiautomated segmentation methods. Generating a mask of those slices containing a bone marrow lesion takes approximately 2 minutes per MRI, and the rest is done automatically by the software. Because of these advantages, we recommend using this method in future studies evaluating BMLs.

In conclusion, the KNN classification method in combination with the optimization techniques described in the current study provides an accurate, reliable, and reproducible tool for analysis of BML lesions on large numbers of T2w SPIR MRIs and can, therefore, be implemented in future studies analyzing the relationship of BMLs with phenomena such as clinical symptoms, histology of subchondral bone damage, and future development of OA.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

References

- Mink JH, Deutsch AL. Occult cartilage and bone injuries of the knee: detection, classification, and assessment with MR imaging. *Radiology*. 1989;170:823-9.
- Hofmann S, Kramer J, Breitsenseher M, Pietsch M, Aigner N. [Bone marrow edema in the knee. Differential diagnosis and therapeutic possibilities]. *Orthopade*. 2006;35:463-75.
- Felson DT, McLaughlin S, Goggins J, LaValley MP, Gale ME, Totterman S, et al. Bone marrow edema and its relation to progression of knee osteoarthritis. *Ann Intern Med*. 2003;139:330-6.
- Felson DT, Niu J, Guermazi A, Roemer F, Aliabadi P, Clancy M, et al. Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. *Arthritis Rheum*. 2007;56:2986-92.
- Hunter DJ, Zhang Y, Niu J, Goggins J, Amin S, LaValley MP, et al. Increase in bone marrow lesions associated with cartilage loss: a longitudinal magnetic resonance imaging study of knee osteoarthritis. *Arthritis Rheum*. 2006;54:1529-35.
- Roemer FW, Guermazi A, Javaid MK, Lynch JA, Niu J, Zhang Y, et al. Change in MRI-detected subchondral bone marrow lesions is associated with cartilage loss: the MOST Study. A longitudinal multicentre study of knee osteoarthritis. *Ann Rheum Dis*. 2009;68:1461-5.
- Frobell RB, Roos HP, Roos EM, Hellio Le Graverand MP, Buck R, Tamez-Pena J, et al. The acutely ACL injured knee assessed by MRI: are large volume traumatic bone marrow lesions a sign of severe compression injury? *Osteoarthritis Cartilage*. 2008;16:829-36.
- Kornaat PR, Ceulemans RY, Kroon HM, Riyazi N, Kloppenburg M, Carter WO, et al. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system. *Skeletal Radiol*. 2005;34:95-102.
- Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, et al. Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage*. 2004;12:177-90.
- Mayerhoefer ME, Breitsenseher MJ, Kramer J, Aigner N, Norden C, Hofmann S. STIR vs. T1-weighted fat-suppressed gadolinium-enhanced MRI of bone marrow edema of the knee: computer-assisted quantitative comparison and influence of injected contrast media volume and acquisition parameters. *J Magn Reson Imaging*. 2005;22:788-93.
- Mayerhoefer ME, Breitsenseher M, Hofmann S, Aigner N, Meizer R, Siedentop H, et al. Computer-assisted quantitative analysis of bone marrow edema of the knee: initial experience with a new method. *AJR Am J Roentgenol*. 2004;182:1399-403.
- Anbeek P, Vincken KL, van Osch MJ, Bisschops RH, van der Grond J. Probabilistic segmentation of white matter lesions in MR imaging. *Neuroimage*. 2004;21:1037-44.
- Bishop CM. *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press; 1995.
- Dice L. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297-302.
- Bartko J. Measurement and reliability: statistical thinking considerations. *Schizophr Bull*. 1991;17:483-489.