

## RESEARCH ARTICLE

# T<sub>2</sub>-weighted magnetic resonance imaging texture as predictor of low back pain: A texture analysis-based classification pipeline to symptomatic and asymptomatic cases

Juuso H. J. Ketola<sup>1</sup>  | Satu I. Inkinen<sup>1</sup>  | Jaro Karppinen<sup>2,3,4</sup>  |  
Jaakko Niinimäki<sup>1,2,5</sup>  | Osmo Tervonen<sup>1,2,5</sup> | Miika T. Nieminen<sup>1,2,5</sup> 

<sup>1</sup>Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland

<sup>2</sup>Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland

<sup>3</sup>Department of Physical and Rehabilitation Medicine, Rehabilitation Services of South Karelia Social and Health Care District, Lappeenranta, Finland

<sup>4</sup>Department of Occupational Health, Finnish Institute of Occupational Health, Oulu, Finland

<sup>5</sup>Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland

## Correspondence

Juuso H. J. Ketola, Research Unit of Medical Imaging, Physics and Technology, University of Oulu/Faculty of Medicine, P.O. Box 5000, FI-90014 University of Oulu, Finland.  
Email: [juuso.ketola@oulu.fi](mailto:juuso.ketola@oulu.fi)

## Funding information

Tauno Tönnöngin Säätiö; Jane ja Aatos Erkon Säätiö; Teknologiateollisuuden 100-Vuotisjuhlasäätiö

## Abstract

Low back pain is a very common symptom and the leading cause of disability throughout the world. Several degenerative imaging findings seen on magnetic resonance imaging are associated with low back pain but none of them is specific for the presence of low back pain as abnormal findings are prevalent among asymptomatic subjects as well. The purpose of this population-based study was to investigate if more specific magnetic resonance imaging predictors of low back pain could be found via texture analysis and machine learning. We used this methodology to classify T<sub>2</sub>-weighted magnetic resonance images from the Northern Finland Birth Cohort 1966 data to symptomatic and asymptomatic groups. Lumbar spine magnetic resonance imaging was performed using a fast spin-echo sequence at 1.5 T. Texture analysis pipeline consisting of textural feature extraction, principal component analysis, and logistic regression classifier was applied to the data to classify them into symptomatic (clinically relevant pain with frequency  $\geq 30$  days and intensity  $\geq 6/10$ ) and asymptomatic (frequency  $\leq 7$  days, intensity  $\leq 3/10$ , and no previous pain episodes in the follow-up period) groups. Best classification results were observed applying texture analysis to the two lowest intervertebral discs (L4-L5 and L5-S1), with accuracy of 83%, specificity of 83%, sensitivity of 82%, negative predictive value of 94%, precision of 56%, and receiver operating characteristic area-under-curve of 0.91. To conclude, textural features from T<sub>2</sub>-weighted magnetic resonance images can be applied in low back pain classification.

## KEYWORDS

low back pain, lumbar spine, machine learning, magnetic resonance imaging, texture analysis

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Orthopaedic Research*® published by Wiley Periodicals LLC on behalf of Orthopaedic Research Society.

## 1 | INTRODUCTION

Low back pain (LBP) is a complex condition in which biological, psychological, and social factors impact on both the experience of back pain and associated disability.<sup>1</sup> LBP is a very common symptom and the leading cause of disability throughout the world.<sup>1</sup> Consequently, LBP amounts to a considerable amount of annual costs worldwide when healthcare costs and indirect costs from sick leaves are considered.<sup>2</sup>

LBP may result from an injury or degenerative process of the lumbar innervated tissues such as facet joints, intervertebral discs (IVDs), ligaments, or muscles. Several studies have shown that LBP is related to annular tears,<sup>3,4</sup> disc height narrowing,<sup>3</sup> facet (or apophyseal) joint degeneration,<sup>5</sup> and endplate lesions such as Schmorl's nodes, fractures, erosion, and calcifications.<sup>6</sup>

Magnetic resonance imaging (MRI) is the preferred imaging modality for most spinal diseases, as it allows illustration of vertebrae, IVDs, musculature, nerve roots, foramina, and facet joints with good contrast.<sup>7</sup> MRI studies are often used to confirm IVD herniation, nerve root entrapment, spinal canal stenosis, and more serious pathologies such as trauma or tumor metastases.<sup>8</sup> MRI can also show IVD degeneration and vertebral endplate changes that have been associated with clinically relevant LBP.<sup>9-11</sup> However, these abnormalities are also common among asymptomatic subjects as imaging studies have revealed that up to 87% of asymptomatic people have lumbar IVD abnormalities seen in MRI.<sup>12-14</sup> Thus, lesions or degenerative changes revealed in MRI studies may not be representative of clinical symptoms.<sup>15,16</sup> LBP that cannot be attributed to any known pathology is called nonspecific LBP.

Despite these diagnostic challenges of LBP, substantial effort has been made to find a connection between LBP and MRI findings. Disc degeneration is categorized into five grades, also known as Pfirrmann grades, correlating loss of IVD signal intensity and height in T<sub>2</sub>-weighted MRI to progressive degenerative changes.<sup>17</sup> Associations between LBP and degenerative changes seen in T<sub>2</sub>-weighted MRI have been observed.<sup>18-22</sup> Disc herniation revealed on MRI has been related to LBP in sciatica patients.<sup>23</sup> Multifidus fat infiltrations visible in T<sub>2</sub>-weighted MRI have been strongly associated with ever having LBP and leg pain.<sup>24,25</sup> The so-called Modic changes categorize degenerative changes in vertebral endplate and bone marrow into three types,<sup>26</sup> and a significant association has been found between them and pain as well.<sup>27,28</sup> Furthermore, correlation between Modic changes and Pfirrmann grades of disc degeneration has been observed.<sup>29</sup>

Radiomics stands for the extraction of quantitative features from radiographic medical images. This process consists of image acquisition, reconstruction, segmentation, extraction of features, and building a data analysis pipeline for the given task. Radiomics studies are often conducted by means of texture analysis (TA), which refers to the characterization of images by their texture content. TA encodes images into feature vectors that characterize image properties such as roughness or smoothness by analyzing spatial variation in pixel intensities. Usually, a large amount of features is collected, and afterward, the most important ones are selected by statistical

**TABLE 1** Descriptive statistics from the used data subset

	Symptomatic (N = 110, 21%) N (%)	Asymptomatic (N = 408, 79%) N (%)
Male*	43 (39%)	210 (51%)
Female*	67 (61%)	198 (49%)
Modic type 1*	24 (22%)	50 (12%)
Modic type 2*	41 (37%)	99 (24%)
Modic type 3*	19 (17%)	31 (7.6%)
Pfirrmann 1	1 (0.9%)	1 (0.2%)
Pfirrmann 2	103 (94%)	393 (96%)
Pfirrmann 3	80 (73%)	289 (71%)
Pfirrmann 4*	60 (55%)	164 (40%)
Pfirrmann 5*	30 (27%)	54 (13%)
<i>mean ± STD</i>		
Weight	80.2 ± 16.1	77.6 ± 15.1
Height	170.2 ± 9.2	171.9 ± 9.2
BMI†	27.6 ± 4.6	26.2 ± 4.4

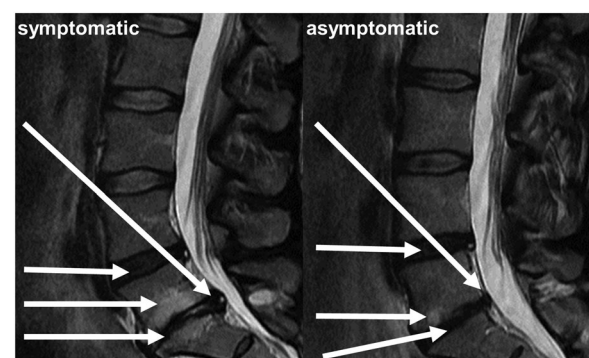
Note: Modic changes and Pfirrmann grades are listed by incidence of each grade across all lumbar levels.

Abbreviations: BMI, body mass index; STD, standard deviation.

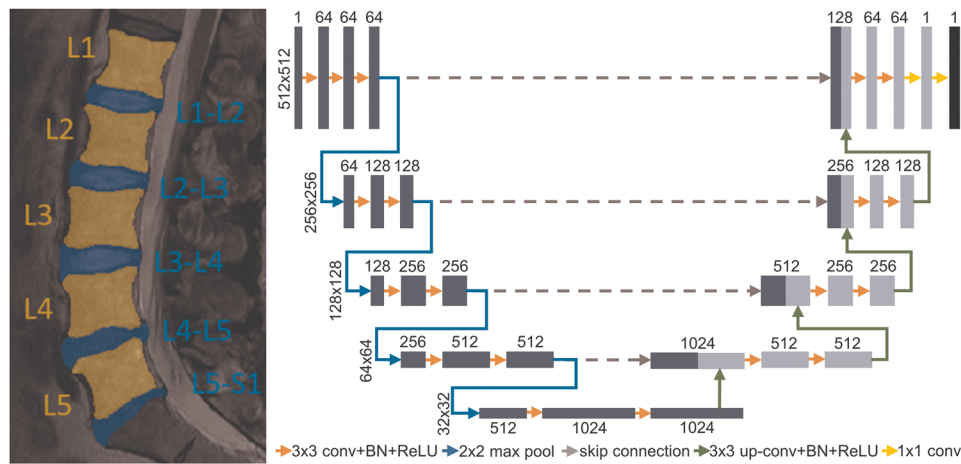
\*Variable significantly different between the symptomatic and asymptomatic groups ( $p < .05$ ),  $\chi^2$  test with contingency tables.

†Variable significantly different between the two groups ( $p < .05$ ), independent t-test.

methods, or alternatively, data are transformed into a fewer-dimensional space by dimensionality reduction techniques such as principal component analysis, to avoid the so-called "curse of dimensionality." Statistical measures or machine learning methods can then be applied to classify these feature data. Recent scientific contributions to using this methodology in spinal MRI include, but are not limited to, assessment of fatty infiltrations in paraspinal musculature,<sup>30,31</sup> classification of spinal metastases of different



**FIGURE 1** Examples from the data showing both a symptomatic and an asymptomatic subject with typical signs of spinal degeneration. Arrows correspond to, from top to bottom: herniation, disc degeneration, endplate changes, and collapsed disc



**FIGURE 2** Deep learning pipeline for segmentation of lumbar MRI data. Left: Example image showing computed segmentations for vertebrae L1...L5 and intervertebral discs L1-L2...L5-S1. Right: U-net architecture used in the segmentation task; conv, convolution; BN, batch normalization; ReLU, rectified linear unit [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

origins,<sup>32-34</sup> automatic categorization of Modic changes,<sup>35</sup> quantitative analysis of neural foramina,<sup>36</sup> quantification of vertebral bone marrow alterations due to aging,<sup>37</sup> and distinguishing bone metastases from spinal hyperplastic hematopoietic bone marrow.<sup>38</sup>

Automation or semiautomation of image processing steps, such as reconstruction, noise and artifact removal, or segmentation makes the imaging pipeline more fluent and time-efficient, and artificial intelligence (AI)-enhanced analysis of imaging findings can greatly

relieve the workload of often-overburdened medical professionals.<sup>39-41</sup> If AI could be used to perform routine tasks or process clear negative findings, more human resources could be used for cases that are more challenging to diagnose.

To our best knowledge, the association between TA of lumbar MRI and the level of LBP individuals are experiencing has not been previously studied. Therefore, the aim of this study is to investigate whether TA could yield more specific MRI predictors of LBP that

**TABLE 2** Description of the computed textural features ( $N = 603$ ) and other features ( $N = 8-9$  depending on the ROI)

Texture type	Feature names	Notes
Grayscale histogram	Maximum value, minimum value, mean, variance, skewness, kurtosis, percentiles (1, 10, 25, 50, 90, and 99)	$N = 12$
Gray-level co-occurrence matrix	Angular second moment, contrast, correlation, sum of squares variance, inverse difference moment, sum average, sum entropy, sum variance, entropy, difference variance, difference entropy, information measures of correlation, maximal correlation coefficient	Directions: $0^\circ$ , $45^\circ$ , $90^\circ$ , and $135^\circ$ ; radii: 1-5 pixels; discretization: 8-bit; $N = 280$
Run-length matrix	Long and short run emphasis, run-length and grayscale nonuniformity, run percentage	Directions: $0^\circ$ , $45^\circ$ , $90^\circ$ , and $135^\circ$ ; max run-length: 8; discretization: 8-bit; $N = 20$
Absolute gradient	Non-zero values, mean, variance, skewness, kurtosis	$N = 5$
Gradient angle	Mean, variance, skewness, kurtosis	$N = 4$
Autoregressive model	phi1-4, sigma	$N = 5$
Wavelet transform	Energy in low-frequency and high-frequency (horizontal, vertical, and diagonal) sub-bands	5 decomposition levels; $N = 20$
Local binary patterns	Local binary patterns histogram, mean	Radius: 1 pixel; $N = 257$
Other		
Imaging phenotypes	Modic changes for individual vertebrae, Pfirman grading for individual intervertebral discs, mean and maximum Modic and Pfirman grades in the ROI	Graded by two independent physicians. $N = 4-5$
Demographic variables	Gender, weight, height, BMI	$N = 4$

Note: The  $N$  in the right-hand column refers to the number of individual extracted features from each feature type.

Abbreviations: BMI, body mass index; ROI, region-of-interest.

could classify lumbar MRI data to symptomatic and asymptomatic cases. Furthermore, our aim is to compare predictive ability between (1) discs and vertebral bodies and (2) upper and lower lumbar levels. Finally, we aim to project our results to a subset of data exhibiting nonspecific LBP symptoms. Such AI-enhanced analysis could be beneficial in the ever-increasing flow of radiological images in terms of time and resource savings.

## 2 | METHODS

### 2.1 | Level of evidence

The level of evidence of this prospective cohort study is 2.

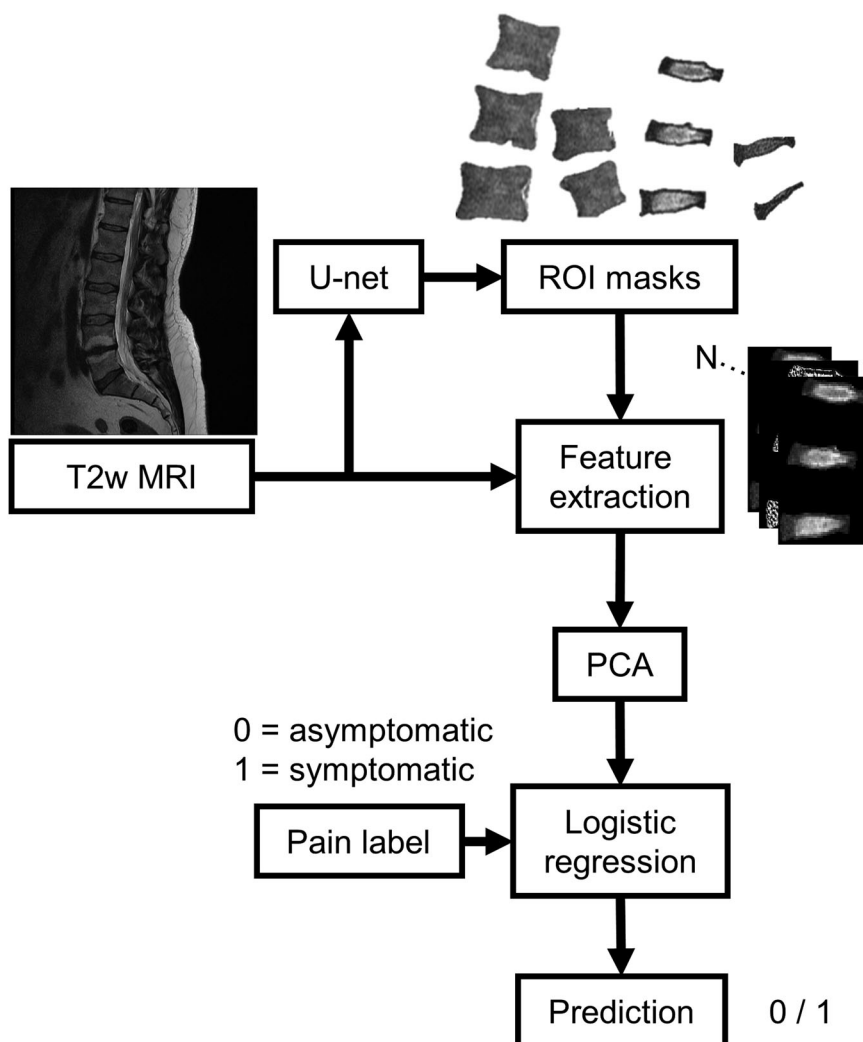
### 2.2 | Data

The Northern Finland Birth Cohort 1966 (NFBC1966, <http://www.oulu.fi/nfbc/>) data were used in this study. The NFBC1966 is a prospective population-based birth cohort. The collection of

NFBC1966 data started in 1965 in Northern Finland. Pregnant women living in Oulu and Lapland were asked during their maternity clinic appointment to take part in the NFBC1966. The inclusion criterion was that the child's expected date of birth was between January 1st and December 31st, 1966. Health and lifestyle data on the mothers ( $N = 12\,068$ ) and children ( $N = 12\,231$ ) have been collected ever since, via postal questionnaires and clinical examinations.<sup>42</sup>

Postal questionnaires on health status and lifestyle were sent to the cohort members whose addresses were known ( $N = 10\,321$ ) at the age of 46–48 (2012–2014). Out of the 66% of the recipients ( $N = 6\,825$ ) who responded with the filled questionnaires, those who currently lived in Finland were invited to clinical examinations, which 57% of the recipients ( $N = 5\,861$ ) attended. Among other examinations, height, and weight (i.e. total body mass) were measured by a trained study nurse, and body mass index (BMI,  $\text{kg}/\text{m}^2$ ) was calculated from these measurements.

Prevalence of LBP over a period of 12 months was elicited by a questionnaire. The first question was “Have you had any aches or pains in your low back?” Drawings were used to elucidate the correct anatomical area. The alternatives to respond to this question were (1) no and (2) yes. In the case of a positive answer, the second



**FIGURE 3** Data processing workflow. Regions-of-interest (ROI) were segmented from MRI data with U-net. ROIs were further separated to contain the three uppermost and two lowermost vertebrae and discs. Textural features were then extracted from these data. Principal component analysis (PCA) and logistic regression were used to predict the pain label. MRI, magnetic resonance imaging

**TABLE 3** First three principal components (PCs) in each analysis. Explained variance and top five features contributing to each PC are listed. Numbers in parentheses indicate the direction (in terms of either displacement in vertical and horizontal directions, or degrees). Square brackets indicate the same feature occurred consecutively but with different parameters

Vertebrae	Explained variance	Feature names of top five features contributing to PC
Upper	PC1 28.2%	Difference entropy [(-3, -3); (-3, 3)]; difference variance [(-3, -3); (-4, -4)]; difference entropy (-2, -2)
	PC2 12.3%	Information measure of correlation [(-3, -3); (-4, -4); (-3, 3); (-4, 4); (0, 5)]
	PC3 11.2%	Grayscale nonuniformity (90°, 0°, 135°, 45°); LBP histogram bin 255
Lower	PC1 29.0%	Sum entropy [(-1, 0); (0, 1); (-1, 1); (-1, -1); (0, 2)]
	PC2 12.4%	Information measure of correlation [(-2, 0); (-2, -2); (-3, 0); (-2, 2); (-1, -1)]
	PC3 8.5%	LBP histogram bin 255; grayscale nonuniformity (90°, 0°); non-zero gradient values; run-length nonuniformity (135°)
<b>IVDs</b>		
Upper	PC1 29.8%	Sum entropy [(-2, -2); (-3, -3); (-2, 2); (-2, 0); (-3,0)]
	PC2 12.4%	Non-zero gradient values; LBP histogram bin 112; information measure of correlation (-1, 0); LBP histogram bin 143; wavelet energy (decomposition level 4)
	PC3 9.8%	Run-length nonuniformity (0°); run percentage (0°), information measure of correlation [(0, 1); (0, 2)]; inverse difference moment (0, 1)
Lower	PC1 32.1%	Sum entropy [(0, 5); (-2, 2); (-3, 3); (-3, 0); (-4, 4)]
	PC2 13.7%	Information measure of correlation [(-1, 0); (-2, 0); (-1, 1); (-2, -2); (-3, 0)]
	PC3 6.3%	Grayscale nonuniformity (0°, 45°, 90°, 135°); LBP histogram bin 255

question was “How often have you had aches or pains during the last 12 months?” to which the alternatives were (1) 1–7 days, (2) 8–30 days, (3) more than 30 days, and (4) daily. In addition, the individuals reporting pain during the past 12 months were asked about the intensity of the pain they were experiencing, using a numerical rating scale from 0 (no pain) to 10 (extremely severe or bothersome pain).

An invitation to undergo lumbar MRI examinations was sent to cohort members that participated in the clinical examinations living no more than 100 km from the city of Oulu ( $N = 1\,988$ ). A total of 1 540 cohort members underwent MRI of the spine.

$T_2$ -weighting was chosen as the studied MRI contrast because it allows for clear visualization of both the vertebrae and IVDs. For texture analysis, mid-sagittal slice images were selected from the imaged volumes. Imaging was performed on a Signa HDxt 1.5 T MRI system (GE Healthcare) using a fast spin-echo sequence (TE = 112.7 ms, TR = 3,500 ms, ETL = 27, slice thickness = 4 mm, matrix size = 512 × 512, and FOV = 280 × 280 mm). Based on the pain characteristics in the questionnaire at the time of imaging, two groups were formed for classification: subjects with clinically relevant pain (frequency ≥30 days and intensity ≥6/10), and no pain (frequency ≤7 days, intensity ≤3/10, and no previous pain episodes in the follow-up period). These cut-offs were motivated by the data: pain intensity of greater than or equal to 6 was perceived as clearly symptomatic and, on the other hand, very mild and infrequent pain symptoms are common and likely not a sign of prolonged LBP. These criteria were met by 518 subjects, comprising 110 and 408 subjects in the symptomatic and asymptomatic groups, respectively. Descriptive statistics about demographic variables, as well as Modic changes and Pfirrmann grades that were read by two independent physicians, are shown in Table 1. Statistical methods used on

the data in Table 1 were the independent t test for continuous variables (distribution normality was tested with the Kolmogorov-Smirnov test) and the  $\chi^2$  test (with contingency tables) for binary variables. Figure 1 shows example images of symptomatic and asymptomatic patients with typical spinal degeneration phenotypes seen in MRI.

### 2.3 | Image segmentation with U-net

Before texture analysis, lumbar vertebrae L1..L5 and IVDs L1-L2..L5-S1 were segmented from the MR images (Figure 2). 200 samples were segmented by hand and used to train a U-net<sup>43</sup> deep learning convolutional neural network that is known to perform well in segmentation tasks. A subset of 15% of the training data was used for model validation. The U-net comprised of five encoding and four decoding layers and was trained for 300 epochs using the combination of binary cross-entropy and Jaccard index (with equal weights) as the loss metric (Figure 2). Separate models were trained to segment the vertebrae and IVDs. The trained networks were then used to segment the vertebrae and IVDs from the rest of the MR images.

### 2.4 | Feature extraction

The obtained segmentation masks were split to four regions-of-interest (ROI): vertebrae L1, L2, and L3; vertebrae L4 and L5; IVDs L1-L2, L2-L3, and L3-L4; and IVDs L4-L5 and L5-S1. A custom-made MATLAB (v.9.7, The MathWorks Inc., Natick, MA, 2019) program was used to extract textural features ( $N = 603$ ) from the ROIs. These features consisted of

**TABLE 4** Classification metrics for grading-based classification (i.e. using only Modic and Pfirrmann metrics) and texture analysis based classification (i.e. including textural features)

ROI	Grading based classification			
	Modic changes		Pfirrmann grading	
	L1-L2, L2-L3, and L3-L4	L4-L5 and L5-S1	L1-L2, L2-L3, and L3-L4	L4-L5 and L5-S1
Specificity	0.62	0.61	0.56	0.55
Sensitivity	0.54	0.59	0.64	0.64
NPV	0.84	0.85	0.85	0.85
Precision	0.28	0.29	0.28	0.27
Accuracy	0.60	0.61	0.58	0.57
ROC-AUC	0.57	0.64	0.60	0.62
ROI	Texture analysis based classification			
	Lumbar vertebrae		Lumbar intervertebral discs	
	L1, L2, and L3	L4 and L5	L1-L2, L2-L3, and L3-L4	L4-L5 and L5-S1
Specificity	0.84	0.77	0.79	0.83
Sensitivity	0.55	0.77	0.55	0.82
NPV	0.87	0.93	0.87	0.94
Precision	0.48	0.47	0.41	0.56
Accuracy	0.78	0.77	0.74	0.83
ROC-AUC	0.78	0.84	0.76	0.91

Note: Scoring metrics for classification quality for the different anatomical ROI with test data are reported. Bolded values highlight the best values across the ROIs. All fitting operations were done on training data. Abbreviations: NPV, negative predictive value; ROC-AUC, receiver operating characteristics area-under-curve; ROI, region-of-interest.

histogram features, gradient features, Haralick features from the grayscale co-occurrence matrix, run-length encoding features, wavelet features, and local binary patterns.<sup>44-46</sup> Images were standardized (to zero mean and unit variance) before feature extraction. In addition to the textural features described above, Modic grading was added to the vertebral features and Pfirrmann grading was added to the IVD features. A more detailed description of the features can be found in Table 2.

## 2.5 | Classification

Sklearn (v. 0.21.2) machine learning library was used in Python (v. 3.7.3) to build a machine learning pipeline to analyze the feature data (Figure 3). Data were split into training (80%) and test (20%) sets with equal class distributions (Table 1). Data were standardized and shuffled before analyses. Principal component analysis (PCA) was used for dimensionality reduction using 80% as the variance threshold. A logistic regression classifier was then implemented to predict the presence of LBP in the subjects. A five-fold

cross-validation scheme was used in conjunction with a grid search to tune the amount of  $L_2$  regularization in the classifier. The best-performing parameters on the training set were then used in the final classification task. The analysis was done separately for different ROIs. All fitting procedures were done on the training data.

For the quantification of classification results, specificity, sensitivity, negative predictive value (NPV), precision, and accuracy scores were computed. In addition, receiver operating characteristic (ROC) curves were visualized and the areas-under-curve (ROC-AUC) were determined. Furthermore, to compare our results with basic grading-based classification, we performed similar logistic regression analysis to Modic and Pfirrmann grade data.

## 2.6 | Sensitivity analysis for nonspecific LBP

To investigate sensitivity for nonspecific LBP, the classification pipeline was run in two different scenarios as follows: (1) discarding cases with sciatica symptoms from the symptomatic group. This was done by asking "Have you had aches in your lower back that are associated with radiating pain or numbness below the knee". In addition, protrusions and extrusions were evaluated from the MRI data of the symptomatic group. If a subject had both radiating pain below the knee and a protrusion or extrusion, they were removed from this analysis (referred to as NS1). This group of nonspecific symptomatic subjects included 69 cases (14.5%). (2) Additionally, cases exhibiting Modic 1 or 2 changes (MC) exceeding 25% of the height of the adjacent vertebrae were discarded as the larger MC were more strongly related to clinically relevant LBP in our previous study.<sup>47</sup> This other group of nonspecific symptomatic subjects (referred to as NS2) included 54 cases (11.7%).

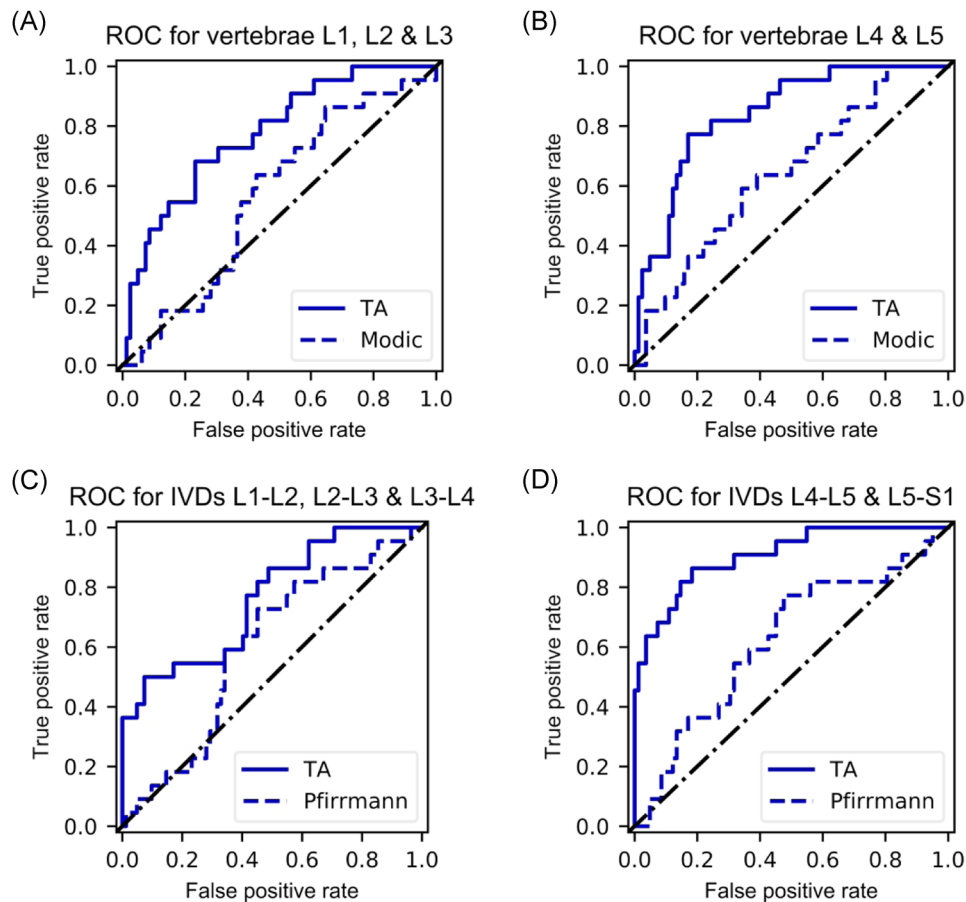
## 3 | RESULTS

### 3.1 | Data preprocessing

An example output from the segmentation network is shown in Figure 2. PCA yielded 44 and 55 principal components for the upper and lower vertebral ROIs, and 53 and 56 principal components for the upper and lower IVD ROIs, respectively. Several gray-level co-occurrence matrix and run-length encoding matrix features were frequently ranked high in the first principal components (Table 3).

### 3.2 | Classification

Using demographic variables with Modic and Pfirrmann grades in the logistic regression analysis resulted in poor classifier performance, and when textural features were added to the analysis, classification results improved greatly (Table 4, Figure 4). When comparing the classification performance between different ROIs, the best results in the test set were obtained with the ROI with the two lowest IVDs with 83% accuracy, 83% specificity, 82% sensitivity, 94% NPV, and 56% precision



**FIGURE 4** (A) Receiver-operating characteristic (ROC) curves for the three uppermost (L1, L2, and L3) vertebrae using Modic grading based (dashed line, area under curve (AUC) = 0.57) and texture analysis (TA) based (AUC = 0.78) classification. (B) ROC curves for the two lowest (L4 and L5) vertebrae using Modic grading based (dashed line, AUC = 0.64) and TA based (AUC = 0.84) classification. (C) ROC curves for the three uppermost (L1-L2, L2-L3, and L3-L4) intervertebral discs (IVDs) using Pfirrmann grading based (dashed line, AUC = 0.60) and TA based (AUC = 0.76) classification. (D) ROC curves for the two lowest (L4-L5 and L5-S1) IVDs using Pfirrmann grading based (dashed line AUC = 0.62) and TA-based (AUC = 0.91) classification. The diagonal lines represent a random classifier (AUC = 0.5) [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(Table 4). Classification with the three uppermost vertebrae yielded a ROC-AUC of 0.78 (Figure 4A), and the two lowest yielded a ROC-AUC of 0.84 (Figure 4B). The ROC-AUC scores for upper (Figure 4C) and lower (Figure 4D) IVDs were 0.76 and 0.91, respectively.

In the sensitivity analysis for nonspecific LBP, in the NS1 group (subjects with sciatica symptoms discarded) all the classification metrics for the lowest two discs improved slightly (0.94 ROC-AUC, 93% specificity, 95% NPV, and 63% precision) apart from sensitivity (71%; Table 5, Figure 5B). The other ROIs exhibited similar improvements (Table 5, Figure 5A-B). In the NS2 group, classification accuracy and ROC-AUC were slightly lower for the lowest discs and slightly higher for the other ROIs (Table 5, Figure 5C, D).

## 4 | DISCUSSION

In this study, the texture of T<sub>2</sub>-weighted MR images was analyzed and machine learning methodology (logistic regression) was used to classify textural features by a binarized pain variable based on a

questionnaire. A subsample of N = 518 subjects from the NFBC1966 data was used in this population-based study. Various classification metrics were computed along with ROC analysis to assess the quality of the classifier.

Best classification accuracy (83%) and ROC-AUC (0.91) in the test set were achieved using the two lowest IVDs (Table 4, Figure 4). The specificity score of 83% suggests that true negatives were relatively well-identified. The sensitivity score of 82% in turn suggests that also true positives were identified by the classification algorithm. NPV refers to the proportion of true negatives in all negative results, and a score of 94% means there were only a few false negatives. Out of all the classification scores, precision was the lowest (56%), suggesting that the classification was not very robust to false positives. False positives were likely represented because up to 87% of asymptomatic subjects are known to display signs of IVD degeneration.<sup>12</sup> However, because asymptomatic subjects account for the majority of the population (79% in our study), classification based on degenerative tissue changes instead of texture analysis would be expected to result in worse precision score than what our model

**TABLE 5** Sensitivity analyses for nonspecific low back pain

ROI	Discarding sciatica symptoms (NS1)			
	Lumbar vertebrae		Lumbar intervertebral discs	
	L1, L2, and L3	L1, L2, and L3	L1-L2, L2-L3, and L3-L4	L4-L5 and L5-S1
Specificity	0.88	0.87	0.77	0.93
Sensitivity	0.64	0.50	0.79	0.71
NPV	0.94	0.91	0.95	0.95
Precision	0.47	0.39	0.37	0.63
Accuracy	0.84	0.81	0.77	0.90
ROC-AUC	0.83	0.90	0.91	0.94
ROI	Additionally discarding large MC1/2 (NS2)			
	Lumbar vertebrae		Lumbar intervertebral discs	
	L1, L2, and L3	L4 and L5	L1-L2, L2-L3, and L3-L4	L4-L5 and L5-S1
Specificity	0.85	0.93	0.93	0.83
Sensitivity	0.73	0.45	0.55	0.82
NPV	0.96	0.93	0.94	0.97
Precision	0.40	0.45	0.50	0.40
Accuracy	0.84	0.87	0.88	0.83
ROC-AUC	0.88	0.85	0.80	0.90

Note: NS1 refers to the case where cases with notable protrusions/extrusions and sciatica symptoms were discarded, and NS2 refers to the case where additionally Modic 1/2 changes (MC1/2) exceeding 25% of adjacent vertebrae height were discarded. Scoring metrics for classification quality for the different anatomical regions-of-interest (ROI) are reported. Bolded values highlight the best values across the ROIs. All fitting operations were done on training data.

Abbreviations: NPV, negative predictive value; ROC-AUC, receiver operating characteristics area-under-curve.

exhibited. Indeed, the comparative grading-based classification using only Modic and Pfirrmann grades yielded inaccurate classification (Table 4, Figure 4) when compared to TA-based classification. Furthermore, LBP is a very complex phenomenon and cannot be attributed to one factor, such as imaging findings, solely.<sup>1</sup> Pain is a highly subjective variable and although the used classification division had strict rules (frequency  $\geq 30$  days and intensity  $\geq 6/10$  in the symptomatic group, and frequency  $\leq 7$  days and intensity  $\leq 3/10$  in the asymptomatic group), it is possible that subjects with mild symptoms were present in both groups, thus affecting the results with regard to precision. Results do, however, suggest a proportionally lower number of false negatives, which is desirable in medical studies.

Rest of the ROIs exhibited similar phenomenon albeit the scores were lower. Interestingly, the texture features of IVDs in the upper lumbar levels could not classify as well as the texture features obtained from the two lowest IVDs or the vertebral ROIs. Classification results between the two vertebral ROIs were similar, with the lower

vertebrae showing higher scores in terms of sensitivity, NPV, and ROC-AUC and the upper vertebrae outperforming in terms of specificity, precision, and accuracy.

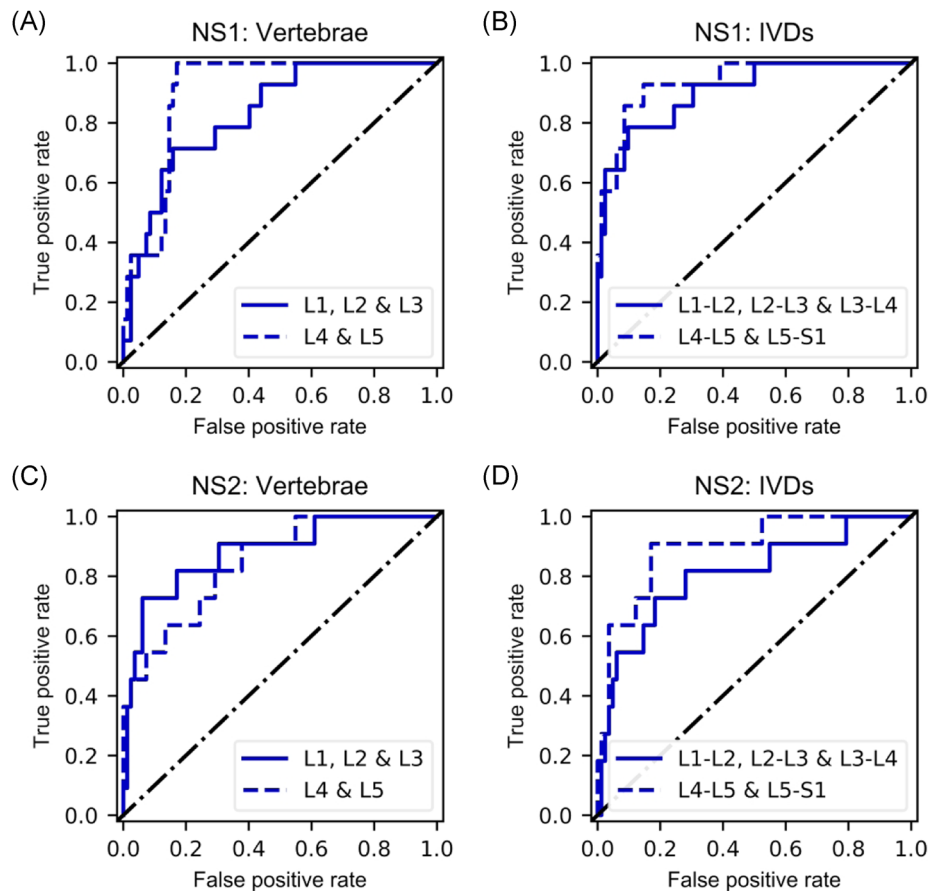
Our results suggest that texture in IVDs L4-L5 and L5-S1 in T<sub>2</sub>-weighted MRI play a role in the manifestation of LBP in the data we used. This is also supported by a recent genetic study that showed a strong significant genetic correlation between IVD problems and back pain.<sup>48</sup> Furthermore, two phenotypes of disc degeneration in the upper and lower lumbar levels have been identified, originating from different injuries to the annulus and endplate.<sup>49,50</sup> While our study does not address the origin of these phenotypes, it shows that T<sub>2</sub>-weighted MRI texture in the lower lumbar levels is more predictive of LBP. Texture features contributing to classification success were more prevalent in the lower discs, which could be indicative of such a distinction of two different IVD phenotypes existing.

Interestingly, our results improved slightly in the NS1 group of sensitivity analysis for nonspecific LBP. In this analysis, subjects with sciatica symptoms were removed from the symptomatic group. The observed improvement could be because protrusions and extrusions are not always causing pain symptoms, and thus they exist in the asymptomatic group as well. In the NS2 group, large Modic 1 and 2 changes were discarded as well. Compared to the original division, this resulted in slightly worse performance for the IVDs, especially in terms of precision, but improved the results for vertebrae. Modic changes are closely related to vertebral signal intensity and discarding the extreme cases seems to result in improvement of texture-based classification of the vertebrae. It should be noted that in both of these cases the initially small group of symptomatic patients became roughly halved. Therefore, the proportion of positives in the group is considerably smaller. However, we believe this analysis gives more insight into how the presented methodology would work on nonspecific cases, and will certainly delve into this further in future studies.

Principal component analysis performs a change of basis on feature data, computing linear combinations of the original features as new features along which the variance is maximized. As most of the variance in the data is explained by the first principal components, we studied the three first principal components in the different classification studies we performed (Table 3). Entropy features from the gray-scale co-occurrence matrix and non-uniformity features from the run-length matrix were present in these principal components in all studies, and information measures of correlation were present in these principal components in all but one study. As similar features were among the highest-contributing features to principal components, those features may have higher predictive power than others. A more thorough investigation comparing classification outcomes using different feature types would be welcome in the future as it could reveal fundamental connections between MRI texture and symptoms.

Our study includes several limitations. We defined a binary pain outcome and a conservative definition of asymptomatic subjects, excluding those with mild symptoms. All subjects were of similar age (46–48 years old) due to the nature of the study population





**FIGURE 5** Sensitivity analyses for nonspecific low back pain. NS1 refers to the case where cases with notable protrusions/extrusions and sciatica symptoms were discarded, and NS2 refers to the case where additionally Modic 1/2 changes (MC1/2) exceeding 25% of adjacent vertebrae height were discarded. (A) Receiver-operating characteristic (ROC) curves for three uppermost (L1, L2, and L3) vertebrae (dashed line, area-under-curve (AUC) = 0.83) and two lowest (L4 and L5) vertebrae (AUC = 0.90) in NS1. (B) ROC curves for three uppermost (L1-L2, L2-L3, and L3-L4) intervertebral discs (IVDs) (dashed line, AUC = 0.91) and two lowest (L4-L5 and L5-S1) IVDs (AUC = 0.94) in NS1. (C) ROC curves for three uppermost vertebrae (AUC = 0.88) and two lowest vertebrae (AUC = 0.85) in NS2. (D) ROC curves for three uppermost IVDs (AUC = 0.80) and two lowest IVDs (AUC = 0.90) in NS2. The diagonal lines represent a random classifier (AUC = 0.5) [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(birth cohort). Because of these statistical concerns, the data we analyzed represents a subset of a population and thus should not be considered as representative of the population at large. Our study concerns only vertebral bodies and IVDs seen from two-dimensional images of the mid-sagittal plane of imaging. In this plane, for example, posterolateral fissures in the disc annulus or pathologies in the facet joints cannot be seen. Including more slices, imaging planes, or data from an isotropic three-dimensional imaging sequence, although computationally far more challenging, would allow for larger regions and more anatomy and tissues to be analyzed. Furthermore, we analyzed IVDs as whole, while further segmentation to the high-intensity nucleus pulposus and low-intensity annulus fibrosus would allow for comparison within the IVD.

While our results show promise in using texture analysis and machine learning in predicting pain from T<sub>2</sub>-weighted images, further investigations are required. In future, we aim to apply deep learning approaches for pain classification from MRI. Deep convolutional neural networks would allow for, for example,

localization of “hot-spots” or attention maps pinpointing the regions contributing to the classification outcome.<sup>51,52</sup> By doing this, each lumbar level could be analyzed without specific ROI delineation for each tissue type, and textures within the IVD or at the vertebral endplate could be analyzed. In addition, using more MRI contrasts, such as T<sub>1</sub> or short-TI inversion recovery (fat suppression), and incorporating 3D data containing more anatomy, is planned. Additional clinical imaging features and specific lesions (such as annulus fissures, high-intensity zones, and endplate defects) will be analyzed in the future as well. Furthermore, while this study suggests that MRI can reveal quantitative image features indicative of the presence of LBP, it does not address which biological processes are behind the alterations to these features. This warrants further investigation in the future, at least for different phenotypes of Modic 1 changes, as it seems that size<sup>47</sup> and location<sup>53</sup> of Modic 1 changes may be related to LBP. Regardless, to our best knowledge, this is the first work applying texture analysis and machine learning in an attempt to predict the presence of LBP from MRI.

To conclude, texture analysis of lumbar MRI data shows promise as a diagnostic tool in the assessment of LBP. This methodology could be used, for example, to identify which tissues and anatomical regions account for the presence of LBP, or to process clearly negative cases to lighten the diagnostic workflow of medical professionals in routine imaging tasks.

## ACKNOWLEDGEMENTS

We gratefully acknowledge support from the Technology Industries of Finland Centennial Foundation and Jane & Aatos Erkko Foundation funds (the Future Makers -program), as well as personal grants from the Tauno Tönnig Foundation. We thank Dr. Juhani Määttä for assistance in processing the clinical MRI data.

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## AUTHOR CONTRIBUTIONS

J.H.J.K. performed programming, analyzed the data, interpreted the results, wrote the manuscript. S.I.I. designed the study, analyzed the data, interpreted the results, and edited the manuscript. J.K. acquired the data, designed the study, interpreted the results, supervised and edited the manuscript. J.N. and O.T. acquired the data and edited the manuscript. M.T.N. designed the study and supervised and edited the manuscript. All authors have read and approved the final submitted manuscript.

## ORCID

Juuso H. J. Ketola  <http://orcid.org/0000-0002-7760-6241>

Satu I. Inkinen  <http://orcid.org/0000-0002-9774-8925>

Jaro Karppinen  <http://orcid.org/0000-0002-2158-6042>

Jaakko Niinimäki  <http://orcid.org/0000-0002-5591-3726>

Miika T. Nieminen  <http://orcid.org/0000-0002-2300-2848>

## REFERENCES

- Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet*. 2018;391:2356-2367.
- Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J*. 2008;8(1):8-20.
- Videman T, Battié MC, Gibbons LE, Maravilla K, Manninen H, Kaprio J. Associations between back pain history and lumbar MRI findings. *Spine (Phila Pa 1976)*. 2003;28(6):582-588.
- Videman T, Nurminen M. The occurrence of annular tears and their relation to lifetime back pain history: A cadaveric study using barium sulfate discography. *Spine (Phila Pa 1976)*. 2004;29(23):2668-2676.
- Boswell MV, Singh V, Staats PS, Hirsch JA. Accuracy of precision diagnostic blocks in the diagnosis of chronic spinal pain of facet or zygapophysial joint origin: A systematic review. *Pain Physician*. 2003; 6(4):449-456.
- Wang Y, Videman T, Battié MC. ISSLS prize winner: Lumbar vertebral endplate lesions: Associations with disc degeneration and back pain history. *Spine (Phila Pa 1976)*. 2012;37(17):1490-1496.
- Sheehan NJ. Magnetic resonance imaging for low back pain: indications and limitations. *Ann Rheum Dis*. 2010;69(1):7-11.
- Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet*. 2009;373(9662): 463-472.
- Balagué F, Mannion AF, Pellisé F, Cedraschi C. Non-specific low back pain. *Lancet*. 2012;379(9814):481-491.
- Carragee EJ. Persistent Low Back Pain. *N Engl J Med*. 2005;352(18): 1891-1898.
- Toyone T, Takahashi K, Kitahara H, Yamagata M, Murakami M, Moriya H. Vertebral bone-marrow changes in degenerative lumbar disc disease. *J Bone Jt Surg*. 1994;76(5):757-764.
- Jarvik JJ, Hollingworth W, Heagerty P, Haynor DR, Deyo RA. The longitudinal assessment of imaging and disability of the back (LAIDBack) study. *Spine (Phila Pa 1976)*. 2001;26(10):1158-1166.
- Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW. Abnormal lumbar magnetic-resonance spine scans of the in asymptomatic subjects: a prospective investigation. *J Bone Jt Surg*. 1990;72(3):403-408.
- Boos N, Semmer N, Elfering A, et al. Natural history of individuals with asymptomatic disc abnormalities in magnetic resonance imaging: predictors of low back pain-related medical consultation and work incapacity. *Spine (Phila Pa 1976)*. 2000; 25(12):1484-1492.
- Chou D, Samartzis D, Bellabarba C, et al. Degenerative magnetic resonance imaging changes in patients with chronic low back pain. *Spine (Phila Pa 1976)*. 2011;36(21):43-53.
- Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without pain. *N Engl J Med*. 1994;331(2):69-73.
- Pfirschmann CWA, Metzendorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa 1976)*. 2001;26(17):1873-1878.
- Luoma K, Riihimäki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. *Spine (Phila Pa 1976)*. 2000;25(4):487-492.
- Kim SY, Lee IS, Kim BR, et al. Magnetic resonance findings of acute severe lower back pain. *Ann Rehabil Med*. 2012;36(1):47-54.
- Takatalo J, Karppinen J, Niinimäki J, Taimela S. Does lumbar disc degeneration on magnetic resonance imaging associate with low back. *Spine (Phila Pa 1976)*. 2011;36(25):2180-2189.
- Takatalo J, Karppinen J, Niinimäki J, et al. Association of Modic changes, Schmorl's nodes, spondylolytic defects, high-intensity zone lesions, disc herniations, and radial tears with low back symptom severity among young Finnish adults. *Spine (Phila Pa 1976)*. 2012; 37(14):1231-1239.
- Cheung KMC, Karppinen J, Chan D, et al. Prevalence and pattern of lumbar magnetic resonance imaging changes in a population study of one thousand forty-three individuals. *Spine (Phila Pa 1976)*. 2009;34(9): 934-940.
- el Barzouhi A, Vleggeert-Lankamp CLAM, Lycklama à Nijeholt GJ, et al. Influence of low back pain and prognostic value of MRI in Sciatica patients in relation to back pain. *PLoS One*. 2014;9(3):1-8.
- Kjaer P, Bendix T, Sorensen JS, Korsholm L, Leboeuf-Yde C. Are MRI-defined fat infiltrations in the multifidus muscles associated with low back pain? *BMC Med*. 2007;5(1):2.
- Kader DF, Wardlaw D, Smith FW. Correlation between the MRI changes in the lumbar multifidus muscles and leg pain. *Clin Radiol*. 2000;55(2):145-149.
- Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. *Radiology*. 1988;166(1):193-199.
- Määttä JH, Wadge S, MacGregor A, Karppinen J, Williams FMK. ISSLS Prize 2015. Vertebral endplate (Modic) change is an independent risk factor for episodes of severe and disabling low back pain. *Spine (Phila Pa 1976)*. 2015;40:1187-1193.
- Mok FPS, Samartzis D, Karppinen J, Fong DYT, Luk KDK, Cheung KMC. Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. *Spine J*. 2016;16(1):32-41.

29. Yu LP, Qian WW, Yin GY, Ren YX, Hu ZY. MRI assessment of lumbar intervertebral disc degeneration with lumbar degenerative disease using the Pfirrmann grading systems. *PLoS One*. 2012;7(12):e48074.
30. Mannil M, Burgstaller JM, Thanabalasingam A, et al. Texture analysis of paraspinal musculature in MRI of the lumbar spine: analysis of the lumbar stenosis outcome study (LSOS) data. *Skeletal Radiol*. 2018;47(7):947-954.
31. Mannil M, Burgstaller JM, Held U, Farshad M, Guggenberger R. Correlation of texture analysis of paraspinal musculature on MRI with different clinical endpoints: Lumbar Stenosis Outcome Study (LSOS). *Eur Radiol*. 2019;29(1):22-30.
32. Lang N, Zhang Y, Zhang E, et al. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magn Reson Imaging*. 2019; 64:4-12.
33. Filograna L, Lenkowicz J, Cellini F, et al. Identification of the most significant magnetic resonance imaging (MRI) radiomic features in oncological patients with vertebral bone marrow metastatic disease: a feasibility study. *Radiol Medica*. 2019;124(1):50-57.
34. Larhman MA, Mahmoudi S, Drisis S, Benjelloun M. A texture analysis approach for spine metastasis classification in T1 and T2 MRI. *International Conference on Bioinformatics and Biomedical Engineering*. Springer; 2018:198-211.
35. Jamaludin A, Kadir T, Zisserman A. Automatic Modic changes classification in spinal MRI. *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*. Springer; 2015:14-26.
36. Gaonkar B, Xia Y, Villaroman DS, et al. Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images. *IEEE J Transl Eng Heal Med*. 2017;5:1-12.
37. Burian E, Subburaj K, Mookiah MRK, et al. Texture analysis of vertebral bone marrow using chemical shift encoding - based water-fat MRI: a feasibility study. *Osteoporos Int*. 2019;30(6):1265-1274.
38. Shigematsu Y, Hirai T, Kawanaka K, et al. Distinguishing imaging features between spinal hyperplastic hematopoietic bone marrow and bone metastasis. *Am J Neuroradiol*. 2014;35(10):2013-2020.
39. Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging*. 2016;35(5):1153-1159.
40. Wang G, Ye JC, Mueller K, Fessler JA. Image reconstruction is a new frontier of machine learning. *IEEE Trans Med Imaging*. 2018;37(6): 1289-1296.
41. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61(13):R150-R166.
42. Rantakallio P. The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr Perinat Epidemiol*. 1988;2(1):59-88.
43. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2015:234-241.
44. Haralick RM. Statistical and structural approaches to texture. *Proc IEEE*. 1979;67(5):786-804.
45. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol*. 2004;59(12):1061-1069.
46. Nanni L, Lumini A, Brahnam S. Local binary patterns variants as texture descriptors for medical image analysis. *Artif Intell Med*. 2010; 49:117-125.
47. Saukkonen J, Määttä J, Oura P, et al. Association between Modic changes and low back pain in middle age: a northern Finland birth cohort study. *Spine (Phila Pa 1976)*. 2020;45(19):1360-1367.
48. Freidin MB, Tsepilov YA, Palmer M, et al. Insight into the genetic architecture of back pain and its risk factors from a study of 590,000 individuals. *Pain*. 2019;160(6):1361-1373.
49. Adams MA, Dolan P. Intervertebral disc degeneration: evidence for two distinct phenotypes. *J Anat*. 2012;221(6):497-506.
50. Li Y, Samartzis D, Campbell DD, et al. Two subtypes of intervertebral disc degeneration distinguished by large-scale population-based study. *Spine J*. 2016;16(9):1079-1089.
51. Jamaludin A, Kadir T, Zisserman A. SpineNet: automated classification and evidence visualization in spinal MRIs. *Med Image Anal*. 2017;41:63-73.
52. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep*. 2018;8(1):1-10.
53. Määttä JH, Karppinen J, Paananen M, et al. Refined phenotyping of Modic changes: imaging biomarkers of prolonged severe low back pain and disability. *Medicine*. 2016;95(22):e3495.

**How to cite this article:** Ketola JHJ, Inkinen SI, Karppinen J, Niinimäki J, Tervonen O, Nieminen MT. T<sub>2</sub>-weighted magnetic resonance imaging texture as predictor of low back pain: A texture analysis-based classification pipeline to symptomatic and asymptomatic cases. *J Orthop Res*. 2021;39:2428-2438. <https://doi.org/10.1002/jor.24973>