
Article:
Castro-Mateos, I., Hua, R., Pozo, J.M. et al. (2016) Intervertebral disc classification by its degree of degeneration from T2-weighted magnetic resonance images. European Spine Journal. ISSN: 0940-6719

https://doi.org/10.1007/s00586-016-4654-6

Please, cite this article in Press: Castro-Mateos, Isaac *et al.*'Intervertebral disc classification by its degree of degeneration from T2-weighted magnetic resonance images" *Eur Spine J (2016)*, doi: 10.1007/s00586-016-4654-6. The final publication is available at Springer via http://dx.doi.org/10.1007/s00586-016-4654-6.

Noname manuscript No.
(will be inserted by the editor)

Intervertebral disc classification by its degree of degeneration from T2-Weighted Magnetic Resonance Images

Isaac Castro-Mateos[1] · Rui Hua[1] · Jose M. Pozo[1] · Aron Lazary[2] · Alejandro F. Frangi[1]

the date of receipt and acceptance should be inserted later

Abstract Purpose The primary goal of this article is to achieve an automatic and objective method to compute the Pfirrmann's degeneration grade of intervertebral discs (IVD) from MRI. This grading system is used in the diagnosis and management of patients with low back pain (LBP). In addition, biomechanical models, which are employed to assess the treatment on patients with LBP, require this grading value to compute proper material properties.

Materials and Method T2-weighted MR images of 48 patients were employed in this work. The 240 lumbar IVDs were divided into a training set (140) and a testing set (100). Three experts manually classified the whole set of IVDs using the Pfirrmann's grading system and the ground truth was selected as the most voted value among them. The developed method employs active contour models to delineate the boundaries of the IVD. Subsequently, the classification is achieved using a trained Neural Network (NN) with eight designed features that contain shape and intensity information of the IVDs.

Results The classification method was evaluated using the testing set, resulting in a mean specificity (95.5%) and sensitivity (87.3%) comparable to those of every expert with respect to the ground truth.

Conclusions Our results show that the automatic method and humans perform equally well in terms of the classification accuracy. However, human annotations have inherent inter- and intra-observer variabilities, which lead to inconsistent assessments. In contrast, the proposed automatic method is objective, being only dependent on the input MRI.

Keywords Classification, IVD, Degeneration, MR, 2D, Automatic, Active Contour Model, Neural Network

I. Castro-Mateos
Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Mechanical Engineering Department, The University of Sheffield, U.K.
E-mail: isaac.casm@sheff.ac.uk

R. Hua
Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Mechanical Engineering Department, The University of Sheffield, U.K.
E-mail: isaac.casm@sheff.ac.uk

J.M. Pozo
Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Mechanical Engineering Department, The University of Sheffield, U.K.
E-mail: j.pozo@sheff.ac.uk

A. Lazary
National Center for Spinal Disorders (NCSD), Budapest, Hungary
E-mail: lazary.aron@gmail.com

A.F. Frangi
Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Mechanical Engineering Department, The University of Sheffield, U.K.
E-mail: a.frangi@sheffield.ac.uk

1 Introduction

Intervertebral disc (IVD) degeneration or Degenerative Disc Disease (DDD) is associated with genetic and environmental factors [1]. In addition, It is believed that many IVD disorders have a mechanical origin, which produce , in many cases, changes in the IVD morphology and histology [2]. Thus, the quantification of the degeneration has been used to estimate the material properties of the IVDs in order to perform biomechanical and mechanobiological simulation of the spinal conditions and treatments [3].

Since the biological process of IVD degeneration and its clinical consequences are still unclear in many de-

tails, there have been some studies to correlate DDD with other disorders, such as low back pain [4,5] or osteoarthritis of the facet joint [6].

These studies are often performed with the recruitment of hundreds or even thousands of subjects. In the everyday clinical practice, the follow-up of disc degeneration, in a given patient, may be important in understanding the progress of the patient's symptoms. In addition, the manual classification of disc degeneration suffers from a relatively high intra- and inter-observer variability. Therefore, a reliable method for the automated classification of IVD degeneration would be a very helpful and time saving tool.

In the literature, different grading system has been proposed [7,8,9,10]. In this article, the Pfirrmann grading system [8] was selected because it is widely used in clinical practice, and most of the methods performing automatic classification of IVD degeneration use this grading system. This criterion divides the degeneration into five degrees as depicted in Fig. 1.

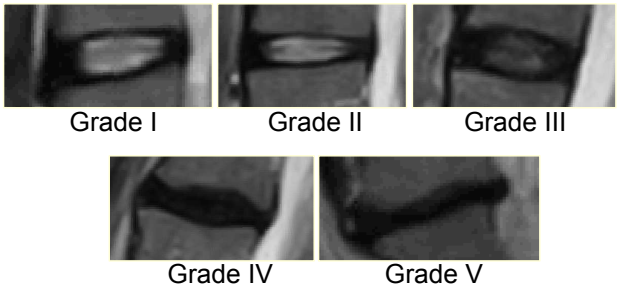

Fig. 1 Example of the five degrees of degeneration.

In the literature, most of the methods dealing with the automatic classification of IVDs according to their degree of degeneration, only consider healthy and unhealthy discs [11,12,13,14,15,16]. This separation corresponds to the difference between degrees I, II and III, IV, V in the Pfirrmann classification.

In contrast, Lootus *et al.*[17], proposed a method to perform a three level classification of the IVD by using a ± 1 precision. Thus, neighbour grades are considered to be the same, reducing the clinical applicability. Their approach extracts the IVD region using a vertebral body segmentation approach. Subsequently, a group of features is extracted from the IVD region and the degeneration grade is estimated using a regression model. A recent work, developed by Ruiz-España *et al.*[18], produces a full classification of the IVDs using the intensity across the height of the IVD. Although, these intensities are well-suited as features, they are not decisive on their own.

In addition, previous methods do not provide a detailed evaluation of the performance, hindering the comparisons with their approaches. In contrast, we present a comprehensive evaluation, specifying the results by different IVD positions and degeneration degrees.

The proposed method employs our previous work on IVD segmentation and classification [19]. The algorithm employed a novel extension of Active Contour Models to perform an automatic segmentation of IVDs allowing the extraction of intensity and shape-based features. However, this preliminary set of features did not encode the Pfirrmann's rationale about the degeneration.

The contributions of the current article are:

- The addition of a new set of features to better explain the intensity changes in the nucleus following Pfirrmann's rationale, providing more accurate results.
- The selection of a optimal classifier by performing a comparison of different techniques and evaluating their performance.
- A in-depth evaluation of the performance of the method.

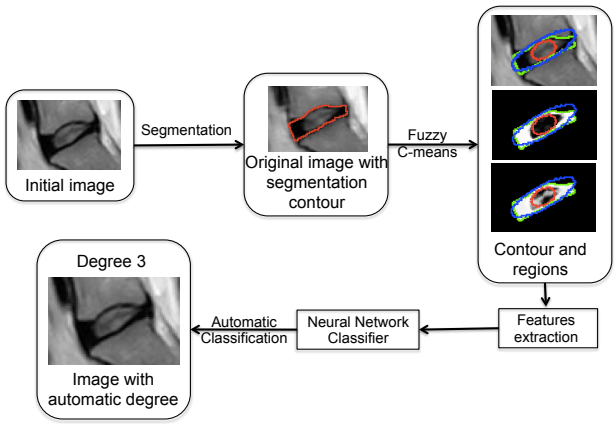

Fig. 2 Workflow of the proposed method.

2 Materials and Method

2.1 Data

The database comprises of 48 T2-weighted MR images with $0.68 \times 0.68\ mm^2$ in-plane resolution and an inter-slice space of 4.5mm. The entire cohort was collected at the National Center for Spinal Disorders (Budapest, Hungary) using an open MR machine of 0.4 T MRI system (Hitachi, Twinsburg, OH). Patients were 25 males and 23 females with a mean age of 44 (age interval: 27-62 years). The images were selected from the database

of the European MySpine project, which had received approval of the institutional review board. In addition, a written informed consent from each subject was retrieved.

The IVDs were grouped according to their degree of degeneration, given by the Pfirrmann's classification [8] (grades I to V), and corresponding lumbar level (Table 1). This classification was performed by three independent experts and the most voted values were selected as the gold standard (Table 1). In the case, that none of the experts coincide, the median value was selected.

Degeneration Grade	**I**	**II**	**III**	**IV**	**V**
Discs in L1-L2 postions	0	14	30	3	1
Discs in L2-L3 postions	0	21	21	6	0
Discs in L3-L4 postions	5	17	17	7	2
Discs in L4-L5 postions	4	10	13	17	4
Discs in L5-S1 postions	2	6	19	9	12
Discs per degeneration	**11**	**68**	**100**	**42**	**19**

Table 1 Degeneration Grade in the Dataset

2.2 Intervertebral Disc Segmentation

We employ our previous method, presented in [19]. This method requires the IVD centre position as input to automatically compute the IVD region (ROI). The intensity in this ROI is scaled between 0 and 255. This step normalizes the intensity respect to the patients, since the ligaments will get values close to 0 value and the spinal cord close to 255. Subsequently, an ellipse is initialized at the centre of the ROI and is warped by a modified active contour model (ACM) with a new geometrical energy in order to find the contour of the IVD. Finally, a post-processing based on fuzzy C-means [20] is applied to improve the details of the IVD contour.

2.3 Classification of the degree of degeneration

The study addressed in this article aims to classify the IVDs into healthy (Grade I and II), low-level degeneration (Grade III), high-level degeneration (Grade IV) and collapsed space (Grade V). There exist two different components that are to be optimized: features and classifier.

Grades I and II were merged due to the large intervariability found among the three experts. Furthermore, our database contains a small number of cases hindering the classification between these two grades. Nevertheless, these grades are considered to be healthy, and in general, it is rare to find IVDs of grade I in adults, since IVDs undergo the process degeneration with age.

2.3.1 Features

In his article [8], Pfirrmann classified the degree of degeneration into five categories employing the following information (Pfirrmann's features):

P1: Signal intensity in the nucleus.
P2: Homogeneity of the intensity. Is there any cracks or dark spots in the nucleus?
P3: Distinction of the nucleus and annulus.
P4: Height of the IVD.

In this article, a set of eight features was employed to classify the IVDs in an attempt to emulate the idea behind Pfirrmann's rationale.

Three images were employed to compute the proposed set of features.

- **Image A:** Original T2-weighted image.
- **Image B:** Image resulting from fuzzy clustering applied to the segmented region of the studied IVD.
- **Image C:** Image as a result of a fuzzy clustering, considering the whole ROI for the studied IVD as explained in section 2.2.

In both derived images B and C, the fuzzy clustering method created an image with the probability that a pixel belonged to the darkest cluster within the segmentation area. The number of clusters was selected as 2, since the classification should separate between annulus and nucleus. Considering only the segmented region, Image B (Figure 3), allows a large intensity separation between the nucleus and annulus, providing information between grade 3 and 4. On the other hand, the use of the whole ROI , Image C (Figure 3), provides large intensity separation between annulus and nucleus when the grade are 1 or 2, and permits the integration of spots and cracks information that improve the separation between grade 2 and 3.

In addition, it was defined a region A within the IVD to roughly represent the nucleus. This region was described as an ellipse with a short axis of $0.9 \times b$ and a long axis $1.5 \times b$ (red ellipse in the images in Figure 3).

The eight features and their relationship with the Pfirrmann's features are depicted in Table 2.

2.3.2 The Classifier

Although the selection of the set of features is the most important information to perform proper classifications, in this work we compared the specificity that different classifiers achieved on the training set.

Features	Explanation	Pfirrmann's features
Feature 1 (F_1)	MIA of image B	P1
Feature 2 (F_2)	MIA of image C	P1
Feature 3 (F_3)	SDA of image B	P2
Feature 4 (F_4)	SDA of image C	P2
Feature 5 (F_5)	RIA of image A	P3
Feature 6 (F_6)	RIA of image B	P3
Feature 7 (F_7)	RIA of image C	P3
Feature 8 (F_8)	Flatness of the IVD	P4

MIA: Mean intensity in region A.
SDA: Standard deviation of the intensities in the region A.
RIA: Ratio between the mean intensity in region A and the mean intensity in the whole IVD except region A.

Table 2 Description of the eight features used for the classification of the IVD according to their degeneration grade. The third column shows their correspondence with the Pffirmann features. The flatness of the IVD was computed as the ratio between the short axis (b) and long axis (a) of the fitted ellipse to the IVD segmentation.

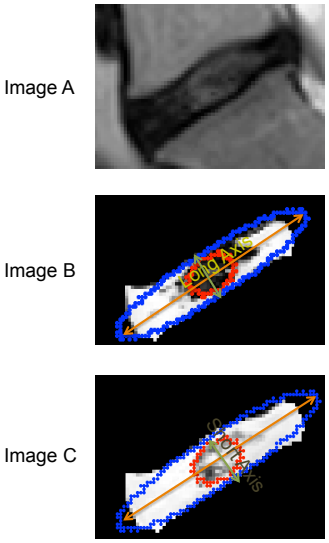

Fig. 3 Representation of the eight features for the classification of the IVD. The first and third features are the mean and variance intensity of the pixels within the small ellipse (region A) in Image B and the second and fourth ones are the same values in Image C. The next three features are the ratios between the intensity in the small ellipse and the intensities in the segmented area but the small ellipse, in Images A, B and C. The last feature is the long axis divided by the short axis of the large ellipse

Table 3 shows the specificity of the IVD classification according to their degree of degeneration, employing a group of classifiers and parameters. The results, shown in the table, are organized according to the sensitivity that was achieved in the training set. The classifiers and their parameters are:

- **NN-1:** Neural network (NN) with a random initialization. This classifier contained 1 hidden layer with 12 neurones
- **Ad:** Adaboost with 10 iterations
- **SVM-1:** Support Vector machine with a polynomial of order 3
- **LR:** Logistic Regression with a polynomial of order 3
- **SVM-2:** Support Vector machine with Gaussian radial bases
- **NN-2:** Neural network (NN) with Swarm optimization. This classifier contained 1 hidden layer with 12 neurones

Classifiers	NN-1	Ad	SVM-1	LR	SVM-2	NN-2
Sensitivity	0.86	0.89	0.90	0.90	0.90	0.91

Table 3 Sensitivity obtained with different Classifiers.

Although the performance of the classifiers is not very different, we selected the NN-2.

2.4 Evaluation of the method

Three experts manually labelled the cohort of 240 IVDs and the most voted value for each IVD was selected as the ground truth. For the creation of the ground-truth, the experts were allowed to look through the different slices of the same image. However, in the case of the automatic method, the mid-sagittal slice was the only input.

This dataset was divided into a training and validation set (140 IVDs) and testing set (100 IVDs). This division was performed automatically, trying to preserve the same relative quantity of the five different type of degeneration at each lumbar position using the method implemented in [21].

The inter-observer variability was computed as the mean of the specificity and sensitivity of each observer compared to the ground truth. In order to evaluate the performance of the proposed automatic method, its specificity and sensitivity were also computed against the ground-truth and compared with the inter-observer variability,

$$\text{Specificity(Spe)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (1)$$

$$\text{Sensitivity(Sen)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negative, respectively. In addition, we present their 95% confidence interval (CI). This CI was computed assuming a Gaussian approximation of the binomial distribution (mean ± 2×standard deviation), since it is a common approach that can be used for benchmarking.

$$\text{Var(Spe)} = \frac{\text{Spe(1-Spe)}}{\text{TN+FP}}, \quad \text{Var(Sen)} = \frac{\text{Sen(1-Sen)}}{\text{TP+FN}} \quad (3)$$

Observe that when the number of samples is small or the Sensitivity or specificity are close to 0 or 1, this approximation fails. Thus, we also employed a Wilson confidence interval (WCI), which has better coverage than the 'exact' interval (Clopper-Pearson interval) and it does not have any Bayesian influence (Jeffrey interval) [22].

Furthermore, we employed the Rao's score test [23] to measure the similarity between the inter-observer variability and our results, since this test has been proved to be the most powerful test when the tested values are close to each other.

3 Results

Table 4 presents both the Gaussian approximation of the CI (GCI) of the sensitivity (sen) and specificity (spe), as well as the respective WCI obtained from the testing set (section 2.4). Observe that the confidence intervals for the global specificity were not computed since these intervals are only valid for binomial distributions and the global specificity involves a multinomial distribution.

In order to compare the inter-observer variability and the results from the proposed method, the score test was used. The p-values obtained with this test were always above 0.97, meaning that no significant difference is found between the manual and the proposed automatic classification.

The algorithm was tested in a Windows 64 bit computer with Intel(R) Xeon CPU E5620 at 2.40 Ghz with 8 GB of RAM. The code was written in MATLAB achieving an elapsed time of 1.13 seconds per disc, including the segmentation step.

In order to provide the relative influence of each feature, Table 5 presents the sensitivity and specificity of each feature together with the eighth feature. Observe that at least two features are necessary to train the classifier and since the eighth feature only provides meaningful information for the separation between the fifth degree and the rest, it was considered the most appropriate for this test.

4 Discussion

The proposed method aims at classifying IVDs according to their degree of degeneration using the mid-sagittal slice of T2 weighted MR images. The advantage of such method is the avoidance of manual intervention for this classification. This classification plays a key role, together with the IVD segmentation, for automatically detecting spinal diseases, as well as for the patient-specific predictive systems to treat various spinal pathologies [24]. In addition, physic-based simulation of the spine requires the grading values to perform properly.

The sensitivity and specificity of the proposed method was estimated on a 100-IVD testing set, providing also their Wilson confidence intervals (Table 4). In addition, the score test could not prove that the values were not equivalent to those achieved by the experts.

Overall, the proposed method achieved accurate results with no subjectivity. And, although the CI ranges are wide, showing that more cases of highly degenerated discs are required, the lowest value of all the intervals is still above 60%, which proves its reliability. Notice that the observers interval are narrower than the automatic method because they were obtained using the whole dataset instead of the testing set. The robustness of the classifier was also proved by checking that it always kept the error within ±1. This means that if the real degeneration grade was 3, it would never choose 1 or 5 or that it would obtain 100% of sensitivity and specificity with ±1 accuracy. This evaluation strategy was also employed in [17], in which the sensitivity was reported as 85.8% with ±1 accuracy.

An interesting observation is that the method does not seem to perform better when the three experts are in agreement. This was proved by observing that the values obtained in Table 4 were similar when only evaluating the IVDs that the three experts graded with the same degeneration. . A possible interpretation is that the agreement is randomly distributed, without a subjacent factor related to the difficulty or ambiguity of the image. In case the inter-observer agreement depended on this hypothetical subjacent factor, it would indicate that the accuracy of the automatic method did not show any significant correlation with it.

Another observation is that the classification method seems not to be influenced by the standard range of accuracy in automatic segmentations. Although, we did not perform an in-depth evaluation of it, the employed segmentation technique provided results with different levels of accuracy, a dice similarity index from 85% to 95%. This difference in quality did not seem to influence the performance of the classifier. Nevertheless, lower accuracy in the segmentation may influence it.

Degeneration Grade	I & II	III	IV	V	Global
GCI Sen Observers	81.3%±8.0%	85.0%±6.4%	90.0%±7.9%	100%±0%	85.2%±4.6%
GCI Spe Observers	94.8%±3.0%	92.7%±3.9%	94.2%±2.9%	97.8%±1.7%	94.4%
GCI Sen Proposed method	87.5% ± 11.7%	85.7%±10.8%	83.3%±17.6%	100%±0%	87.3%±4.4%
GCI Spe Proposed method	93.1% ± 6.0%	88.3% ±8.3%	97.7%±3.2%	100%±0%	95.5%
WCI Sen Observers	72.3% - 87.8%	77.6% - 90.1%	80.1% - 96.0%	86.2% - 100%	79.3% - 89.6%
WCI Spe Observers	91.1% - 97.1%	88.0% - 95.7%	90.6% - 96.4%	95.5% - 99.0%	-
WCI Sen Proposed method	71.9% - 95.0%	72.2% - 93.3%	60.8% - 94.2%	67.6% - 100%	78.2% - 93.9%
WCI Spe Proposed method	84.8% - 97.0%	77.8% - 94.2%	91.9% - 99.4%	96.2% - 100.0%	-

Table 4 Accuracy of the classification, by measuring the specificity (spe) and sensitivity (sen) ± 2 standard deviations and the Wilson confidence interval (WCI) assuming the prior information as an uniform distribution. The ground truth was selected as the most voted value of each disc among the three experts manual labelling. The accuracy of these experts is shown in the observers rows. The automatic method required a training set of 140 IVDs and the result was assessed with a testing set of 100 IVDs.

Degeneration Grade	I & II	III	IV	V	Global
Sen/Spe F_1 & F_8	68/91%	81/72%	75/93%	75/100%	75/90%
Sen/Spe F_2 & F_8	0/100%	17/92%	25/93%	100/100%	50/75%
Sen/Spe F_3 & F_8	81/87%	77/79%	63/95%	100/100%	78/92%
Sen/Spe F_4 & F_8	0/100%	14/84%	25/80%%	50/96%	43/69%
Sen/Spe F_5 & F_8	36/73%	65/51%	75/90%	50/100%	55/78%
Sen/Spe F_6 & F_8	0/100%	84/100%	38/71%	0/100%	41/68%
Sen/Spe F_7 & F_8	72/81%	70/73%	50/95%	75/100%	70/88%

Table 5 Comparison of the classification power of each individual feature together with the eight feature.

Furthermore, the proposed features are robust against different type of MRI machines and resolutions. T2 MR images provide information about the water content in the structures. Thus, in general the relative difference between the intensities in the nucleus and annulus will be similar. The proposed method extracts the region of an IVD and normalizes the intensities to remove the difference in contrast. In addition, the features are statistics and ratios that are robust against local intensity changes and resolutions produced by different machines. Since, we did not have images from different machines, we only evaluated the robustness against the image resolution. The in-plane pixel size of the testing set was reduced to $1.5mm \times 1.5mm$ and all the experiments were performed again. Observe that the training set images were not changed, thus, the classifier was not re-trained. Although the segmentation slightly changed (Fig. 4), the classification results remained exactly the same. This difference in the segmentation results also proves the reliability of the proposed features with respect to the segmentation accuracy.

A valid and reliable method to automatically classify the disc degeneration may help in large-scale population studies and also in the automation of patient-specific techniques used in personalized medicine. Moreover, the proposed automatic IVD segmentation method provides numerical information about the biological and biomechanical features of the tissue even when these correlations with the MR signals are not fully understood. Furthermore, The use of fuzzy clustering to extract the features allows the method to be straightforwardly extended to different MRI sequences.

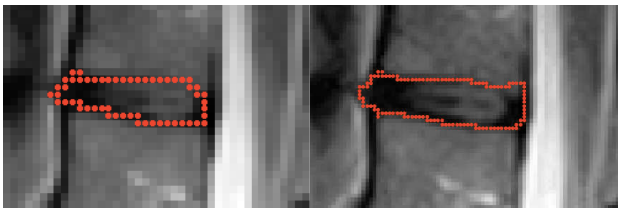

Fig. 4 Comparison of the Snake segmentation with Low (a) and normal resolution (b).

4.1 Limitations

The proposed method requires from the user to select the centre of the IVD that is going to be analysed. Although this step is rather simple, it may be automated using the algorithms from the literature [25,26,27,28].

In addition, the current method has only been tested with images from one type of scanner. Thus, there exists the possibility that the method has to be retrained, or requires an extra factor to accommodate the variation given by different imaging equipments. However,

the nature of the features should be robust enough to palliate or completely mitigate this issue.

Another limitation is the low number of cases, which implied large confidence intervals. Although, this issue hinders the evaluation of the method, the minimum values of the confidence intervals were not extremely low (above 60%). As a future work, we aim at increasing the number of cases.

Another limitation of the method is the type of grading system that was selected. Pfirrmann degeneration grading has been extensively used in research and requires the standard T2-weighted MR image. The main drawback is the selection of a gold standard for evaluating the classification methods, which are commonly subjective because of the vague definition of the different degrees of degeneration. In order to solve this drawback, we collected a manual classification from 3 experts and selected the gold standard as the most voted value. However, there exist new grading systems that define the degrees in a different manner in an attempt to be more objective and they could be a choice for future research work [10,9].

The last limitation is the image modality. Although, there exist correlation between other type of image modalities, such as T1-ρ or T2*, and the Pfirrman's grading system [29,30], we decided to employ the standard T2-weighted because of its direct relationship since the Pfirrmann's grading system was inspired from this modality. In addition, T2-weighted is still the standard modality in many hospitals.

References

1 Sharan A.D., Tang Simon Y., Vaccaro A.R.: Basic Science of Spinal Diseases. JP Medical Ltd2013

2 Freemont A., Watkins A., Le Maitre C., Jeziorska M., Hoyland J.: Current understanding of cellular and molecular events in intervertebral disc degeneration: implications for therapy. The Journal of pathology 196(4), 374–3792002

3 Malandrino A., Pozo J.M., Castro-Mateos I., Frangi A.F., van Rijsbergen M.M., Ito K., Wilke H.J., Dao T.T., Tho M.C.H.B., Noailly J.: On the relative relevance of subject-specific geometries and degeneration-specific mechanical properties for the study of cell death in human intervertebral disk models. Frontiers in bioengineering and biotechnology 32015

4 Luoma K., Riihimäki H., Luukkonen R., Raininko R., Viikari-Juntura E., Lamminen A.: Low back pain in relation to lumbar disc degeneration. Spine 25(4), 487–922000

5 Livshits G., Popham M., Malkin I., Sambrook P.N., Macgregor A.J., Spector T., Williams F.M.K.: Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK twin spine study. Ann Rheum Dis 70(10), 1740–17452011

6 Fujiwara A., Tamai K., Yamato M., An H.S., Yoshida H., Saotome K., Kurihashi A.: The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. European Spine Journal 8(5), 396–4011999

7 Thompson J., Pearce R., Schechter M., Adams M., Tsang I., Bishop P.: Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. Spine 15(5), 411–4151990

8 Pfirrmann C., Metzdorf A., Zanetti M., Hodler J., Boos N.: Magnetic resonance classification of lumbar intervertebral disc degeneration. In: Spine. vol. 26, pp. 1873–18782001

9 Jarman J.P., Arpinar V.E., Baruah D., Klein A.P., Maiman D.J., Muftuler L.T.: Intervertebral disc height loss demonstrates the threshold of major pathological changes during degeneration. European Spine Journal pp. 1–72014

10 Riesenburger R.I., Safain M.G., Ogbuji R., Hayes J., Hwang S.W.: A novel classification system of lumbar disc degeneration. Journal of Clinical Neuroscience 22(2), 346–3512015

11 Chwialkowski M., Shile P., Peshock R., Pfeifer D., Parkey R.: Automated detection and evaluation of lumbar discs in MR images. In: Engin. in Med. Biol. Society, 1989. pp. 571–5721989nov

12 Michopoulou S., Boniatis I., Costaridou L., Cavouras D., Panagiotopoulos E., Panayiotakis G.: Computer assisted characterization of cervical intervertebral disc degeneration in MRI. Journal of Instrumentation 4(05), P050222009

13 Ghosh S., Alomari R., Chaudhary V., Dhillon G.: Composite features for automatic diagnosis of intervertebral disc herniation from lumbar MRI. In: Engin. in Med. Biol. Society,EMBC, 2011. pp. 5068 –50712011Sept

14 Una l.Y., Kocer H., Akkurt H.: A comparison of feature extraction techniques for diagnosis of lumbar intervertebral degenerative disc disease. In: Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on. pp. 490–4942011june

15 Alomari R.S., Corso J.J., Chaudhary V., Dhillon G.: Automatic diagnosis of lumbar disc herniation with shape and appearance features from MRI. In: SPIE Medical Imaging. p. 76241A. International Society for Optics and Photonics2010

16 Neubert A., Fripp J., Engstrom C., Walker D., Weber M., Schwarz R., Crozier S.: Three-dimensional morphological and signal intensity features for detection of intervertebral disc degeneration from magnetic resonance images. Journal of the American Medical Informatics Association 20(6), 1082–10902013

17 Lootus M., Kadir T., Zisserman A.: Automated radiological grading of spinal mri. In: Recent Advances in Computational Methods and Clinical Applications for Spine Imaging, pp. 119–130. Springer2015

18 Ruiz-España S., Arana E., Moratal D.: Semiautomatic computer-aided classification of degenerative lumbar spine disease in magnetic resonance imaging. Computers in biology and medicine 62, 196–2052015

19 Castro-Mateos I., Pozo J.M., Lazary A., Frangi A.F.: 2D segmentation of intervertebral discs and its degree of degeneration from T2-weighted magnetic resonance images. In: SPIE Medical Imaging. pp. 903517–903517. International Society for Optics and Photonics2014

20 Cannon R.L., Dave J.V., Bezdek J.C.: Efficient implementation of the fuzzy c-means clustering algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on (2), 248–2551986

21 Castro-Mateos I., Pozo J.M., Eltes P.E., Del Rio L., Lazary A., Frangi A.F.: 3D segmentation of annulus fibrosus and nucleus pulposus from T2-weighted magnetic

resonance images. Physics in medicine and biology 59(24), 7847–78642014
22 Newcombe R.G.: Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in medicine 17(8), 857–8721998
23 Bera A.K., Bilias Y.: Rao's score, Neyman's C(α) and Silvey's LM tests: an essay on historical developments and some new results. Journal of Statistical Planning and Inference 97(1), 9–442001
24 Malandrino A., Noailly J., Lacroix D., Beard D.A.: The effect of sustained compression on oxygen metabolic transport in the intervertebral disc decreases with degenerative changes. PLOS Computational Biology 7(8)2011Aug
25 Schmidt S., Bergtholdt M., Dries S., Schnorr C.: Spine detection and labeling using a parts-based graphical model. In: Lect Notes Comput Sc (IPMI). pp. 122–1332007
26 Stern D., Vrtovec T., Pernus F., Likar B.: Automated determination of the centers of vertebral bodies and intervertebral discs in CT and MR lumbar spine images. In: Proceedings of SPIE. vol. 7623, p. 7623502010
27 Peng Z., Zhong J., Wee W., Lee J.h.: Automated vertebra detection and segmentation from the whole spine MR images. In: Eng Med Biol Soc Ann. IEEE-EMBS. pp. 2527–25302005jan
28 Corso J.J., Alomari R.S., Chaudhary V.: Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Lect Notes Comput Sc (MICCAI). pp. 202–210. Springer-Verlag, Berlin, Heidelberg2008
29 Ellingson A.M., Mehta H., Polly D.W., Ellermann J., Nuckley D.J.: Disc degeneration assessed by quantitative T2*(T2 star) correlated with functional lumbar mechanics. Spine 38(24), E1533–402013
30 Hon J.Y., Bahri S., Gardner V., Muftuler L.T.: In vivo quantification of lumbar disc degeneration: assessment of adc value using a degenerative scoring system based on pfirrmann framework. European Spine Journal pp. 1–72014