Diagnosis of Disc Herniation Based on Classifiers and Features Generated from Spine MR Images *

Jaehan Koh, Vipin Chaudhary[a] and Gurmeet Dhillon, MD[b]

[a]Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA;
[b]Proscan Imaging of Buffalo, Williamsville, NY, USA

ABSTRACT

In recent years the demand for an automated method for diagnosis of disc abnormalities has grown as more patients suffer from lumbar disorders and radiologists have to treat more patients reliably in a limited amount of time. In this paper, we propose and compare several classifiers that diagnose disc herniation, one of the common problems of the lumbar spine, based on lumbar MR images. Experimental results on a limited data set of 68 clinical cases with 340 lumbar discs show that our classifiers can diagnose disc herniation with 97% accuracy.

Keywords: Computer-aided diagnosis; Classifier design and evaluation; MRI

1. INTRODUCTION

Lower back pain is one of the major public health problems in industrialized countries. Besides, it has an enormous economic impact on suffering patients and their families.[1] According to Ambulatory Health Care Data,[2] more than 20 million MRI exams are conducted annually in the United States and 50% of them are related to spine. In recent years, it is reported that there is a concern about a shortage of diagnostic radiologists.[3] Accordingly, the demand for computer-assisted image processing and analysis has grown in the diagnosis of lower back pain problems. Though it is often considered that an anatomic diagnosis is not feasible, the cause of the lower lumbar disorders may be found by using non-invasive imaging such as computerized tomography (CT), and magnetic resonance imaging (MRI). As imaging technologies advance, we can acquire the spinal information more reliably from ultrasound, X-rays, CT, MRI, and a combination of them.

Among many others, disc herniation is regarded as one of the common spinal disorders,[4] so many studies have been reported to develop new methods of diagnosing disc herniation based on MRI and/or CT.[5] Kim *et al.*[6] first classified the disc herniation types using MRI based on the extent of derangement of the constituents of the disc. According to them, 242 herniated discs were predicted, achieving an overall accuracy of 85%. Krämer[7] combined the size of the herniation with the direction of migration of the extruded disc material for disc-related classification. Chwialkowski *et al.*[8] presented a method to detect lumbar pathologies in MR images. This algorithm localizes candidate vertebrae with an estimated vertebrae model, finds the line of bisection for the discs from the centers of gravity of two neighboring vertebrae, and verifies the findings by analyzing the trends of changes for subsequent images in the study. They suggested that this method could be applied to the analysis of disc abnormalities in lumbar MR images. Milette[9] attempted to supplement CT, discography or CT discography with MRI in the diagnosis of disc herniation and the determination of its exact location. Pfirrmann *et al.*[10] developed a classification system for lumbar degeneration disc disease (DDD) using MRI. In the system, the lumbar disc degeneration is assessed by way of disc structure, distinction of nucleus and annulus, signal intensity, and the height of invertebral disc. Lumbar intervertebral discs on the T2-weighted sagittal images are classified as Grade I, II, III, IV, and V, where Grade V represents discs in which the disc structure is inhomogeneous, the distinction between nucleus and annulus is lost, signal intensity is hypointense, and the disc space is collapsed. Fardon and Milette[11] proposed a nomenclature and classification system for lumbar disc pathology, describing the terms and pathologic conditions of lumbar discs. Tsai *et al.*[12] described a method of diagnosing lumbar herniated inter-vertebral disc (HIVD) using transverse sections of CT. Shape feature and position of disc herniation are used in the analysis and diagnosis. Lumbar HIVD is classified

* This research was supported in part by a grant for NYSTAR.
Send correspondence to jkoh@buffalo.edu.

Medical Imaging 2010: Computer-Aided Diagnosis, edited by Nico Karssemeijer, Ronald M. Summers,
Proc. of SPIE Vol. 7624, 76243O · © 2010 SPIE · CCC code: 1605-7422/10/$18 · doi: 10.1117/12.844386

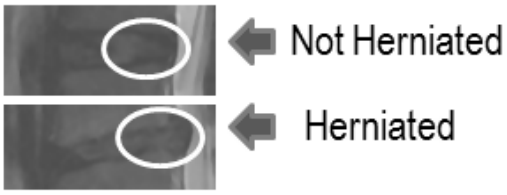

Figure 1. A normal disc and a herniated disc. The top figure shows a normal disc and the bottom figure is a herniated disc.

according to the degree to which the disc substances herniate the ligaments: bulging, protrusion, extrusion, and separation. Thalgott *et al.*[13] presented a new classification system for DDD of the lumbar spine from MRI, provocative discography and plain anteroposterior plus lateral radiographs. On the basis of the three modalities, they tried to assess lumbar discs simply in terms of MRI image appearance, lordotic angle of the intervertebral segment, the shape and condition of the end plates, and so on. Griffith *et al.*[14] modified the Pfirrmann grading system for lumbar DDD and tested on 260 lumbar intervertebral discs of elderly subjects. The original 5-level Pfirrmann grading system that was not discriminatory when applied to senior subjects was extended to a 8-level grading system based on MR images. Sagittal T2-weighted images were employed for classification where Grade 1 denotes no disc degeneration and Grade 8 corresponds to severe disc degeneration. They also claimed that the modified system gave good inter-reader and excellent intra-reader agreement. Alomari *et al.*[15] proposed a probabilistic classifier for the detection of abnormality of intervertebral discs. For labeling abnormal discs, they used appearance, location, and context features. In an experiment with 80 clinical cases, they achieved over 91% abnormality detection accuracy.

The method proposed in this paper also focuses on the shapes of discs that neighbor the spinal cord like other approaches[10, 12–15] . Rather than using multiple modalities, our method diagnoses a herniated disc from classifiers and a simple feature generated based on one modality, MRI that is increasingly used for primary investigation of lumbar disc disorders. In this paper, we present several classifiers that can accurately diagnose herniated discs from MR Images. In the preprocessing stage, the segmentation of the lumbar vertebrae as well as the generation of a feature vector is performed. In the classification stage, three classifiers are used to diagnose disc herniation and the results are compared. The rest of this paper is organized as follows. Our problem and method are discussed in Section 2. The usefulness of our method is evaluated in Section 3, followed by conclusion in Section 4.

2. METHOD

2.1 Problem Description

Our goal is to develop classifiers to accurately detect herniated discs based on full protocol MRI of patients as in Fig. 1. By full protocol MRI, we mean T1-weighted sagittal, T2-weighted sagittal, and T2-weighted axial scans of each patient. Formally, it is represented as

$$\{(X_i, D, R) : X_i \in \mathcal{X}, D \in \{L1{-}L2, L2{-}L3, L3{-}L4, L4{-}L5, L5{-}S1\}, R \in \{herniated, non{-}herniated\}, 1 \leq i \leq k\} \quad (1)$$

where $\mathcal{X}$ represents pattern set, D refers to region of interest associated with lumbar disc level, R the diagnostic results, and k the number of test patterns.

Our method of herniation detection consists of two stages: the preprocessing stage and the classification stage. In the preprocessing stage, ROIs are cropped and class labels are assigned to disc, vertebra and spinal cord regions based on manually marked boundary points. Then the feature vector is generated by computing the percentage of each label within each sub-window. In the classification stage, three classifiers are used for training and testing. The classification results are evaluated by comparing the diagnostic decisions from the classifiers against those of radiologists'.

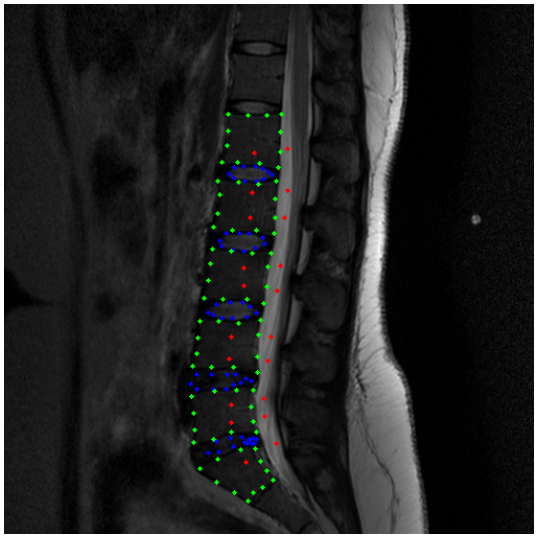

Figure 2. Boundaries of lumbar disc area. Boundaries are identified by dots. Green dots represent vertebrae, blue dots discs, and red dots disc-vertebra-spinal regions.

2.2 Preprocessing

To diagnose herniated discs, we only need to look at the regions where lumbar vertebra, intervertebral discs and the spinal cord meet. To this end, vertebra and disc boundaries are roughly marked manually as in Fig. 2. Green dots mark vertebra boundaries whereas blue dots represent nucleus pulposus. Red dots form the boundaries of disc-vertebra-spinal cord region. Disc boundaries are generated from adjacent bounding boxes of two vertebrae as follows. Assume that twelve points in the ith vertebra are numbered counter-clockwise from top left in the order of $1_i, 2_i, \ldots, 12_i$ and $1_{i+1}, 2_{i+1}, \ldots, 12_{i+1}$ in the $(i+1)$th vertebra, where $(i+1)$th vertebra is one level lower than ith vertebra. A disc bounding box is formed from these two neighboring vertebrae and the relationship between them is as follows: $\{X(1), X(2), X(3), X(4), X(5), X(6), X(7), X(8)\} = \{4_i, 1_{i+1}, 12_{i+1}, 11_{i+1}, 10_{i+1}, 7_i, 6_i, 5_i\}$. This process is required since the disc boundaries should include annulus fibrosus along with nucleus pulposus as shown in Fig. 3(a). Then, the bounding box for segmenting an ROI that contains disc-vertebra-spinal cord is obtained from red dots. Suppose again that four red points are assigned clockwise from top left as $\{(y_1, x_1), (y_2, x_2), (y_3, x_3), (y_4, x_4)\}$. Then the rectangular bounding box is formed by using the following information: the starting point of $\{\frac{y_1+y_2}{2}, x_1\}$, the width of $(x_3 - x_1)$, the height of $(y_4 - y_2)$. This is briefly explained in Fig. 3(b). Fig. 4 shows some of ROIs based on these rectangular boundaries. Note that we do not need $z-$coordinate information since we crop ROIs from sagittal views. These boundaries of vertebrae and discs are used to assign class labels to ROIs and the feature vectors are generated from the ROIs and the labels.

Afterwards, sub-windows are obtained from the ROI (refer to Fig. 5(a) by subdividing it into 9 pieces. The sub-windows are labeled from 1 to 9 in the top-to-bottom, and left-to-right order as in Fig. 5(b). A feature vector for diagnosing disc herniation is generated per sub-window. The feature vector based on the percentage of each class (i.e., disc, vertebra or spinal cord) in each sub-window is represented by

$$X_i = \{(D_j, W_k, P_l) : D_j \in \{L1-L2, L2-L3, L3-L4, L4-L5, L5-S1\}, 1 \leq W_k \leq 9, P_l \in \{(P_d, P_v, P_s)\}\} \quad (2)$$

where D_j, W_k, and P_l denotes disc index, sub-window index, and percentage index, respectively. Since we have 9 sub-windows, W_k goes from 1 to 9. P_d, P_v, and P_s refer to the percentage of disc, the percentage of the vertebra, and the percentage of the spinal cord in the sub-window, respectively. To be specific, the percentage of each class within each sub-window is computed by

$$\frac{\text{the number of pixels in class } c_i \text{ within } W_k}{\text{the total number of pixels in sub-window } W_k} \quad (3)$$

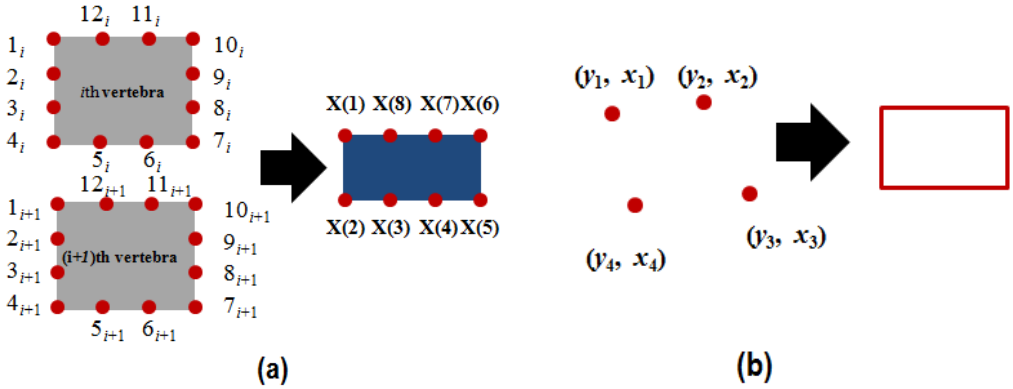

Figure 3. Formation of bounding boxes. Part (a) shows how to form disc boundary from vertebrae boundaries information. Part (b) shows how to form rectangular bounding boxes for ROI segmentation from four red dots.

Figure 4. Cropped ROIs. Based on the bounding boxes, ROIs are obtained.

We pick the percentage as our feature vector because it is observed from our clinical image data that when there is disc herniation, the average intensity of the corresponding window becomes darker. It is reported that this approach is widely adopted in many researchers and medical doctors already[10, 12–15] .

2.3 Classification

Three classifiers are used for the diagnosis of herniated discs: a support vector machine (SVM) classifier, a perceptron classifier, and a least mean square (LMS) classifier. Classification results (i.e., a herniated disc or a normal disc) are compared against gold standards based on reports written by radiologists.

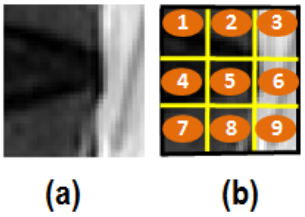

Figure 5. Partitioning into sub-windows. Part (a) shows one cropped ROI. Part (b) shows how the ROI is divided into sub-windows for feature generation.

Table 1. Ground truth based on doctor's reports

PatientID	$L1-L2$	$L2-L3$	$L3-L4$	$L4-L5$	$L5-S1$
b2a3p01	N	N	N	N	Y
b2a3p02	N	N	N	Y	N
b2a3p03	N	Y	N	N	N
b2a3p04	N	N	N	Y	Y
b2a3p05	N	N	N	N	Y

3. EXPERIMENTS

3.1 Image Set and Hardware

A total of 68 clinical MR image sets from several protocols are used for our experiments. Among them, 16 MR patient image sets are from normal patients (i.e., having no herniated discs) and the remaining 52 image sets are from abnormal patients (i.e., having one or more herniated discs). The images are obtained by 3 a Tesla Philip's Medical Systems MRI scanner and are used in a clinical environment. The dimension of each image slice is 512 $\times$ 512 pixels. The experiments are carried out on a machine with an Intel(R) Xeon CPU at 2GHz speed, and 3GB physical memory.

3.2 Gold Standards

Reports generated by radiologists in a clinical environment are used as gold standards for our classifiers. These reports contain basic patient's data, type of examination, the findings and conclusions from the diagnosis as in Table 1. By comparing the classification results of the system with the gold standards, we evaluated our classifiers.

3.3 Classifiers

Three classifiers used for the experiments are: a perceptron classifier, a least-mean-square (LMS) classifier, and a support vector machine (SVM) classifier. The cross-validation experiments were conducted with 48 randomly-chosen training patterns out of 68 image data sets and with 14 testing patterns picked randomly from the remaining feature patterns. This process has been repeated seven times. Training is done per intervertebral disc, since each disc has different geometric shapes and the classifier can train more effectively. Moreover, it allows the classifiers to diagnosis more than two herniated discs at the same time.

After training is done, testing is performed. The perceptron and the LMS classifiers use the inner product of weight W and input pattern X_k in determining the class. The SVM classification works as follows: given any test input X, set the class $y(X)$ by

$$sign\left(\sum_{i=0}^{n_s} d_i \hat{\lambda_i} K(X, X_i) + \hat{w_0}\right) \tag{4}$$

Based on that *sign*, the class of each input pattern is predicted.

A **perceptron classifier** classifies the feature vector in the training set $\mathcal{T} = \{(X_k, d_k) \mid X_k \in \mathbb{R}^{n+1}, d_k \in (0,1)\}_{k=1}^{Q}$ where each pattern X_k is labeled as one of two classes c_0 or c_1 (i.e., a normal disc or a herniated disc), and the desired output d_k is denoted by 0 or 1. Given a training set $\mathcal{T}$, consisting of augmented vectors $X_k \in \mathbb{R}$ (a bias term is added), the weight vector $W \in \mathbb{R}^{n+1}$ is initialized with 0. Then the following processes are repeated until convergence: (i) $X_k \in \mathcal{T}$ is selected; (ii) The inner product $X_k^T W_k$ of X_k and W is calculated; (iii) W_{k+1} is updated where

$$W_{k+1} = \begin{cases} W_k + \eta_k X_k & \text{if } X_k \in c_1 \text{ and } W_k^T X_k \leq 0 \\ W_k - \eta_k X_k & \text{if } X_k \in c_0 \text{ and } W_k^T X_k \geq 0 \end{cases} \tag{5}$$

η_k is the learning rate that scales X_k before adding or subtracting it from W_k. In the experiments, we set η_k to be 0.1 and the training is performed for 10000 iterations.

Table 2. Diagnostic results of disc herniation in terms of detection rates

Classifier Type	$L1-L2$	$L2-L3$	$L3-L4$	$L4-L5$	$L5-S1$
SVM	100%	27%	82%	42%	53%
Perceptron	100%	94%	97%	72%	64%
LMS	98%	100%	100%	95%	91%

A **least-mean-square classifier** tries to classify the feature vector in the training set $\mathcal{T} = \{(X_k, d_k) \mid X_k \in \mathbb{R}^{n+1}, d_k \in \mathbb{R}\}$ by incorporating a linear error measure into the weight update procedure. Similar to perceptron, given the training set $\mathcal{T}$, the weight vector, $W \in \mathbb{R}^{n+1}$, is initialized with random numbers. Then the following are repeated until convergence: (i) $X_k \in \mathcal{T}$ is selected; (ii) The inner product $X_k^T W_k = s_k$ of X_k and W is calculated; (iii) The error measure, e_k, is computed by $e_k = d_k - s_k = d_k - X_k^T W_k$ and then minimized over all patterns in the training set; the weights, W_{k+1}, are updated according to

$$W_{k+1} = W_k + \eta e_k \frac{X_k}{\|X_k\|^2}. \tag{6}$$

Here the learning rate η is set to 0.1 and the training is performed for 5000 epochs.

A **support vector machine classifier** attempts to classify the feature vector in the training set $\mathcal{T} = \{(X_k, d_k) \mid X_k \in \mathbb{R}^n, d_k \in \{-1, +1\}\}_{k=1}^{Q}$ using the smallest norm of weights. d_k takes on either $+1$ or -1, indicating the class (normal disc or herniated disc) to which the feature X_k belongs. This boils down to the following optimization problem:

$$\text{Minimize } \frac{1}{2}\|W\|^2 \tag{7}$$

subject to the constraints,

$$d_k(X_k \cdot W + w_0) - 1 \geq 0, k = 1, \ldots, Q.$$

The objective function tries to maximize the margin $\frac{1}{\|W\|}$. Specifically, given a training set $\mathcal{T}$ and the desired output $d_k \in \{-1, +1\}$, the Hessian matrix H_{ij} is set to $d_i d_j K(X_i, X_j)$ and the control parameter C for determining the upper bound of the Lagrange multipliers is set to 10. By maximizing the objective function, Lagrange multipliers are optimized.

3.4 Results and Discussion

The recognition rates of the each classifier per disc level are shown in Table 2. The average recognition rates for each classifier are: 61% for the SVM classifier, 85% for the perceptron classifier, and 97% for the LMS classifier. In case of the SVM classifier, the recognition rates are relatively low and fluctuate across disc level because the SVM classifier is sensitive to the proper selection of kernel function. The right selection of the kernel function for our problem is not easy since it should account for the variability among patients of different ages, gender, and physical conditions. Furthermore, the quadratic programming introduces errors into the system, thereby degrading the performance of the SVM classifier. However, for the case of $L1-L2$, the SVM classifier gives excellent diagnostic results. Different kernel functions for each disc level might increase the rates. In the case of the perceptron classifier and the LMS classifier, the classification rates fall as the classifiers diagnose lower lumbar regions, especially in the case of $L5-S1$. The occurrence of a herniate disc in the lower disc level is higher compared to other disc levels and therefore the features should accurately express the abnormalities associated with each disc level. The LMS classifier works better than the perceptron classfier across almost all disc regions. It also works better than the SVM classifier. This is, in part, because the feature vector generated in the preprocessing stage has a linear characteristic. Fig. 6 shows one snapshot of signal values versus feature values when training of the LMS classifier is finished. Patterns of our feature vector are linearly separable across all feature values (i.e., percentage of each label). Therefore, the linear classifiers (i.e., the perceptron and the LMS classifiers) works better than the non-linear classifier (i.e., the SVM classifier).

The performance of each classifier is also compared in terms of sensitivity and specificity. Sensitivity is defined as

$$\text{Sensitivity} = \frac{\text{the number of true positives}}{\text{the number of true positives} + \text{the number of false negatives}} \tag{8}$$

Table 3. Sensitivity and specificity of the classifiers

Classifier Type	Sensitivity	Specificity
SVM	51%	63%
Perceptron	60%	90%
LMS	99%	96%

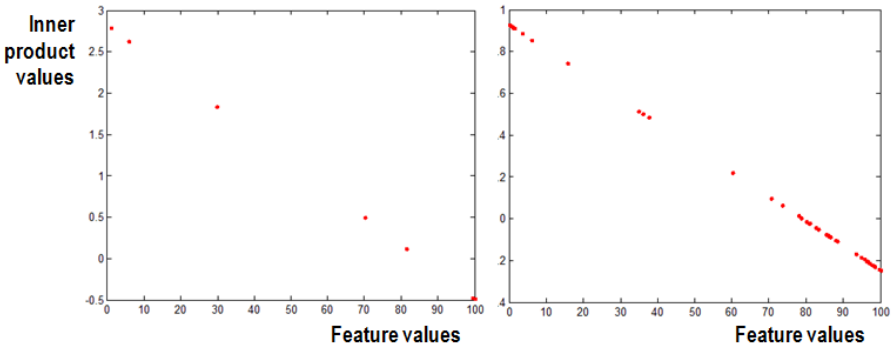

Figure 6. Distribution of features. The plot of signal values versus feature values after training is done show that they are linearly separable.

Similarly, specificity is calculated as

$$\text{Specificity} = \frac{\text{the number of true negatives}}{\text{the number of true negatives} + \text{the number of false positives}} \tag{9}$$

The sensitivity and specificity for each classifier is shown in Table 3. According to the metrics, the LMS classifier performs best among others. The LMS classifier has a higher probability of correctly diagnosing a herniated patient as herniated one (i.e., 99%) than correctly identifying a non-herniated patient as normal one (i.e., 96%). The opposite is true for the SVM classifier and the perceptron classifier.

There is no single perfect classifier for diagnosing disc herniation of the whole intervertebral discs. Instead, some classifiers performs better than others for certain disc levels. Specifically, all three classifiers can be used to diagnose disc level $L1-L2$. For disc regions $L2-L3, L3-L4$, the perceptron classifier and the LMS classifier are good choices but not the SVM classifier. The LMS classifier can detect disc herniation accurately in disc levels $L4-L5$ and $L5-S1$. By comparing recognition results from other classifiers we can verify the diagnostic results of the classification system, making precise diagnostic decisions.

Since we performed experiments with a limited number of patient data, we plan to run our classifiers with more patient data containing abnormalities in disc level $L1-L2$. By increasing the number of clinical image data and analyzing the characteristics of the feature vector across the image sets, we will be able to understand the feature more closely and to improve the performance of the classifiers in herniation diagnosis.

4. CONCLUSIONS AND FUTURE WORK

This paper presents three classifiers to diagnose herniated discs in lumbar MR images. On the basis of the feature vector generated in the preprocessing stage, classification is performed. Experimental results on 68 clinical cases with 340 lumbar discs show that our classifiers diagnose disc herniation with 97% accuracy. In the future, we plan to increase the size of patient data so that our classifiers work well with patients of different ages and genders.

REFERENCES

[1] Frymoyer, J. W. and Cats-Baril, W. L., "An overview of the incidences and costs of low back pain," *Orthop. Clin. North Am.* **22**, 263–271 (1991).

[2] Cherry, D. K., Hing, E., Woodwell, D. A., and Rechtsteiner, E. A., "National ambulatory medical care survey: 2006 summary," *National Health Statistics Reports* (3), 1–39 (2008).

[3] Bhargavan, M., Sunshine, J. H., and Schepps, B., "Too few radiologists?," *American Journal of Roentgenology* **178**(5), 1075–1082 (2002).

[4] Deyo, R. A., Mirza, S., and Martin, B. I., "Back pain prevalence and visit rates: Estimates from u.s. national surveys," *Spine* **31**, 2724–2727 (2002).

[5] Haughton, V., "Medical imaging of intervertebral disc degeneration: Current status of imaging," *Spine* **29**(23), 2751–2756 (2004).

[6] Kim, K. Y., Kim, Y. T., Lee, C. S., Kang, J. S., and Kim, Y. J., "Magnetic resonance imaging in the evaluation of the lumbar herniated intervertebral disc," *International Orthopaedics* **17**(4), 241–244 (1993).

[7] Krämer, J., "A new classification of lumbar motion segments for microdiscotomy," *European Spine Journal* **4**(6), 327–334 (1995).

[8] Chwialkowski, M. P., Shile, P. E., Peshock, R. M., Pfeifer, D., and Parkey, R. W., "Automated detection and evaluation of lumbar discs in mr images," *IEEE Engineering in Medicine and Biology* , 2527–2530 (1989).

[9] Milette, P. C., "Classification, diagnositc imaging, and imaging characterization of a lumbar herniated disk," *Radiologic Clinics of North America* **38**, 1267–1292 (2000).

[10] Pfirrmann, C. W. A., Metzdorf, A., M. Zanetti, J. H., and Boos, N., "Magnetic resonance classification of lumbar intervertebral disc degeneration," *Spine* **26**(17), 1873–1878 (2001).

[11] Fardon, D. F. and Milette, P. C., "Nomenclature and classification of lumbar disc pathology," *Spine* **26**(5), 93–113 (2001).

[12] Tsai, M., Jou, S., and Hsieh, M., "A new method for lumbar herniated inter-vertebral disc diagnosis based on image analysis of transverse sections," *Computerized Medical Imaging and Graphics* **26**(6), 369–380 (2002).

[13] Thalgott, J. S., Albert, T. J., Vaccaro, A. R., Aprill, C. N., Giuffre, J. M., Drake, J. S., and Henke, J. P., "A new classification system for degenerative disc disease of the lumbar spine based on magnetic resonance imaging, provocative discography, plain radiographs and anatomic considerations," *Spine* **4**, 167S–172S (2004).

[14] Griffith, J. F., Wang, Y.-X. J., Antonio, G. E., Choi, K. C., Yu, A., Jhuja, A. T., and Leung, P. C., "Modified pfirrmann grading system for lumbar intervertebral disc degeneration," *Spine* **32**(24), E708–E712 (2007).

[15] R. S. Alomari, J. J. Corso, V. C. and Dhillon, G., "Computer-aided diagnosis of lumbar disc pathology from clinical lower spine MRI," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)* (2009).