MedLife Open Access

A machine learning-based method in order to diagnose lumbar disc herniation disease by MR image processing

Elias Ebrahimzadeh*,†,1, Farahnaz Fayaz†, Fereshte Ahmadi† and Mehran Nikravan‡

*School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran, †Department of Biomedical Engineering, School of Electrical Engineering, Payame Noor University of North Tehran, Tehran, Iran, ‡Department of Biomedical Engineering, Shahed University, Tehran, Iran

ABSTRACT

Lumbar disc herniation is one of the most common diseases that results in lower back pain. A computer-aided diagnosis system can be helpful to generate diagnostic results within a short time. In addition, it can be useful to increase precision of diagnosis and eliminate the human errors caused by tiredness and inevitable visual errors of radiologists. Therefore, if there was an automatic intelligent diagnosis of lumbar disc herniation, the need of tiresome analysis of the images by the physicians would be unnecessary. In this article, a new method for automatic intelligent diagnosis of lumbar disc herniation is proposed, which is based on the clinical magnetic resonance (MR) imaging data. We used T2-W sagittal and myelographic images. We used Otsu thresholding method to extract the spinal cord from MR images of lumbar disc. Then a third-order polynomial is aligned on the extracted spinal cords, and in the end of the preprocessing step, all the T2-W sagittal images are prepared for extracting disc boundary and labeling. After labeling and extracting an region of interest (ROI) for each disc, intensity and shape features are used for classification. This method was applied on 30 clinical cases, each containing 7 discs (210 lumbar discs) for the herniation diagnosis. The extracted features have been selected by local subset feature selection. Eventually, the results revealed 96.38% and 97.80% accuracy for artificial neural network and support vector machine classifiers, respectively. The results indicate the superiority of the proposed method to those mentioned in similar studies.

KEYWORDS

lumbar disc diseases, LBP, automatic diagnosis, herniation, MRI, myelography, spinal cord

INTRODUCTION

Lumbar disc diseases are the most widespread complaints for lower back pain (LBP).[1] Herniation in the lumbar area is one of the most common diseases that results in LBP.[2] Herniation is a disease in which disc substance compresses the spinal cord and roots. The degree of lesions depends on the nucleus pulposus state. The first stage is called bulging, in which the annulus fibrosus is intact and the disc has a mild herniation with a little compression on ligaments and spinal cord. In the second stage (called protrusion), the herniated disc breaks the inner ligaments and compresses the spinal cord or roots. In the third stage, extrusion, the disc substance breaks all ligaments except outer ones, and the compression on the spinal cord and roots is considerable. In this situation, herniated disc has a free fragment. In another type of herniation, the free fragment migrates superiorly. In the fourth stage, there is a herniated disc with a free fragment that has ruptured through the posterior ligaments and is against the thecal sac.[3] A computer-aided diagnosis (CAD) system can be helpful to generate diagnostic results within a short time. In addition, it can be useful to increase precision of diagnosis and eliminate the effects

Publisher Name: MedText Publications
Manuscript compiled: Friday 16th March, 2018
[1]Corresponding author: School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran.
E-mail: e_ebrahimzadeh@ut.ac.ir

of human errors caused by tiredness and inevitable visual errors of radiologists. Therefore, the need of tiresome analysis of the images by the physicians is unnecessary.[4] The spinal cords have wide variability including turning angle, size, shape, etc. Moreover, abnormal pertaining conditions such as vertebral fusions, disc diseases, and spinal infections add to these varieties.[5] The labeling and detection of lumbar disc are necessary to design a CAD system for automatic diagnosis of lumbar area diseases such as herniation on clinical magnetic resonance (MR) images. Some radiologists make use of computed tomography (CT) to detect back disorders. Clinically speaking, an MR imaging provides images with higher resolution. Furthermore, due to the higher safety of the MRI in comparison to the radiative emission of CT, the MRI is a more suitable method for this purpose. CT is usually used in the situations where the disease is located inside bone structures such as injuries to back bones. By the way, this issue is not within the scope of this study. Figure 1 depicts the MRI, which is extracted from the data set.

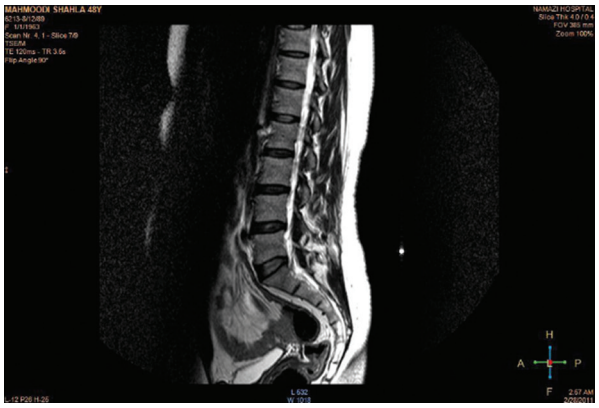

Figure 1 T2-weighted magnetic resonance sagittal view of lumbar region.

In this study, we present a new method for disc labeling by means of fitting a third-order polynomial on extracted spinal cord region from MRI and magnetic resonance myelographic (MRM) images. After labeling and region of interest (ROI) selection in the discs limited area, our method generates the features and diagnoses lumbar disc herniation (LDH) by means of artificial neural network and support vector machine (SVM) classifiers. Automatic lumbar disc diseases diagnosing has been developed in many of the previous works. These studies investigated a computer-based method for the diagnosis of disc degeneration.[6–8] Michopoulou et al.[9] developed a fuzzy-c means classifier to perform normal and degenerated segmentation. This method was semiautomatic and achieved 86%–88% accuracy (AC) on 34 cases. Roberts et al.[10] employed watershed techniques for automatic detection of intervertebral discs from a combination of proton density (PD) and T2-waighted MR images. Chevrefils et al.[11] utilized a combination of watershed and morphological operations to detect the discs from MR images. Recently, Alomari et al. used GVF-snake and designed a Gibbs-based classifier to classify each disc as either normal or herniated.[12] In the second section, we discuss methods and materials, which consist of data set, previous related works, feature extraction, our method; in the third section, we discuss experimental results of our method; in the fourth section, we discuss about the study results, and we finally concludes in the last section in detail. Based on the aforementioned explanations, the block diagram of our approach for diagnosis of lumbar disc herniation disease is shown in Figure 2.

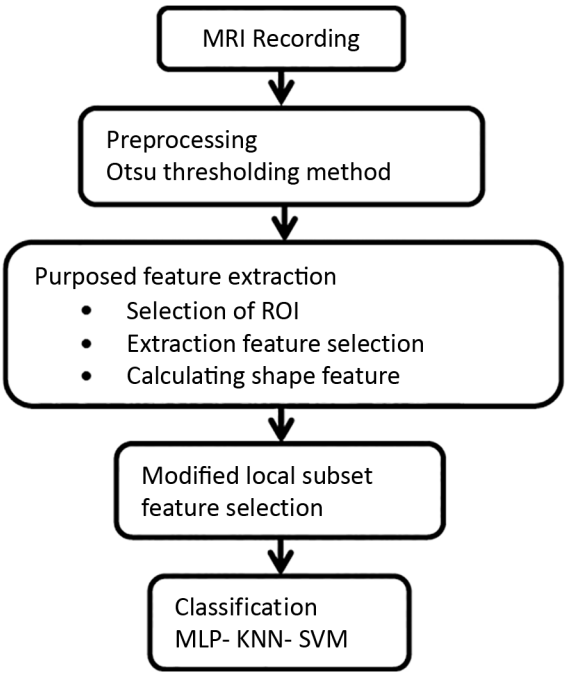

Figure 2 Block diagram of the proposed approach for the diagnosis of lumbar disc herniation disease. KNN, k-nearest neighbor; MLP, multilayer perceptron; MRI, magnetic resonance imaging; SVM, support vector machine.

METHODS AND MATERIALS

Data set

Data were obtained using a 1.5-T MR imaging system (Siemens) from subjects aged between 21 and 73 years. Each case comprisesd seven slices. In this study, we used just T2-W sagittal (30 images) and MRM series (30 images) for diagnosis. In the set of T2-W sagittal images (nine images), the best one is the fifth image. The parameters of imaging were as follows: flip angle of 150° for both T2-W sagittal and MRM images, echo time of 121 ms, repetition time of 3,100 ms, field of view (FOV) of 32 cm for T2-W sagittal images, echo time of 1,300 ms, repetition time of 10,000 ms, and field of view of 35 cm for MRM images. The images, labeled by the radiologists, contain 12 normal subjects and 18 patients. Each person has 2.66 disc herniation on average (eight persons with three disc herniation, six with two disc herniation, and four with one disc herniation).

Preprocessing

In this article, according to our previous work, a new method for boundary detection has been used in CAD for diagnosis of LDH based on the MRI.[13] We discriminate and label the intervertebral discs automatically using a combination of different algorithms including Otsu thresholding, morphological operations to improve the images, and a third-order polynomial function to fit the spinal canal.

Otsu thresholding method

The Otsu automatic thresholding approach has been applied to extract spinal canal.[14–16] It is essential to select an appropriate threshold of gray level to extract a specific object through image processing. A lot of techniques for object discrimination have been developed

and used until now. Suppose that the established image has N pixels and the gray level of pixels is between 1 and L. If the number of pixels with gray level i is depicted by n_i, probability distribution function for each gray level i will be equal to:

$$p_i = \frac{n_i}{N}, \quad p_i \geq 0, \sum_{i=1}^{L} P_i = 1. \tag{1}$$

Let us suppose that pixels of the image fall into two subgroups called C_0 and C_1 with the threshold of K for gray level. The gray level of pixels in C_0 and C_1 subgroups is within the intervals of [1, 2,…, K] and [K+1,…, L], respectively. Now, it is possible to identify a probability distribution function and a mean for each subgroup. ω_0 and ω_1 are the probability distribution functions of C_0 and C_1 subgroups, respectively.

$$\omega_0 = P_r(C_0) = \sum_{i=1}^{K} P_i \tag{2}$$

$$\omega_1 = P_r(C_1) = \sum_{i=K+1}^{L} P_i = 1 - \omega_0 \tag{3}$$

and

$$\mu_0 = \sum_{i=1}^{k} iP_r(i \mid C_0) = \sum_{i=1}^{K} iP_i / \omega_0 = \frac{\mu(K)}{\omega(K)}, \tag{4}$$

$$\mu_1 = \sum_{i=K+1}^{L} iP_r(i \mid C_1) = \sum_{i=K+1}^{L} iP_i / \omega_1 = \frac{\mu_T - \mu(K)}{1 - \omega(K)}, \tag{5}$$

where

$$\omega(K) = \omega_0 = \sum_{i=1}^{K} P_i \tag{6}$$

$$\omega(K) = \sum_{i=1}^{K} iP_i \tag{7}$$

and

$$\mu_T = \mu(L) = \sum_{i=1}^{L} iP_i. \tag{8}$$

μ_T is the average of the image and the following equations hold.

$$\omega_0\mu_0 + \omega_1\mu_1 = \mu_T, \qquad \omega_0 + \omega_1 = 1. \tag{9}$$

The variances of C_0 and C_1 will be given by the following relations, respectively.

$$\delta_0^2 = \sum_{i=1}^{K} (i - \mu_0)^2 P_r(i \mid C_0) = \sum_{i=1}^{K} (i - \mu_0)^2 P_i / \omega_0. \tag{10}$$

$$\delta_1^2 = \sum_{i=K+1}^{L} (i - \mu_1)^2 P_r(i \mid C_1) = \sum_{i=K+1}^{L} (i - \mu_1)^2 P_i / \omega_1. \tag{11}$$

By using below equation, Otsu method calculates the variances between the two subgroups to find a proper threshold:

$$\delta_B^2 = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2. \tag{12}$$

The optimized threshold depicted by t^* will be calculated for $1 \leq t < L$:

$$t^* = \arg\text{Max} \left\{\delta_B^2(t)\right\}. \tag{13}$$

Equation (12) can be generalized to multilevel thresholding. If there are $M-1$ thresholds, shown by $\{t_1, t_2, \ldots, t_{M-1}\}$, the main image will fall into M subgroups. So that subgroups C_0, C_1, and C_{M-1} will have the following thresholds intervals $\{1, \ldots, t_1\}$, $\{t_1 + 1, \ldots, t_2\}$, and $\{t_{M-1} + 1, \ldots, L\}$, respectively, and the optimized thresholds $\{t_1^*, t_2^*, \ldots, t_{M-1}^*\}$ will be selected by maximizing δ_B^2.

$$\delta_B^2 = \sum_{K=0}^{M-1} \omega_K (\mu_K - \mu_T)^2 \tag{14}$$

$$\omega_K = \sum_{i \in C_K} P_K \tag{15}$$

and

$$\mu_K = \sum_{i \in C_K} i(P_i / \omega(K)). \tag{16}$$

$$\{t_1^*, t_2^*, \ldots, t_{M-1}^*\} = \arg\text{Max}\{\delta_B^2(t_1, t_2, \ldots, t_{M-1})\}. \tag{17}$$

We follow the below instruction to extract spinal cord and detect a proper boundary for intervertebral discs. Sagittal images of T2-weighted MRI are longitudinal slices. After studying all slices belonging to 30 patients, it was determined that the spinal canal was depicted with higher resolution in slice number 5 (Figure 3).

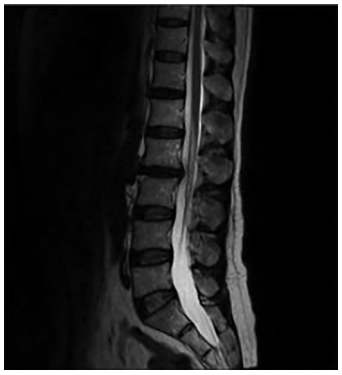

Figure 3 The fifth cut is the most proper one among nine slices.

After investigating the image number 5 by means of trial and error, it is identified that only middle parts of the image are included in our ROI because the discs and spinal canal are located in this area of the image. Therefore, we removed right and left redundant parts to speed up the process and simplify the analyses. Figure 4 shows the main image after the removal of the useless areas.

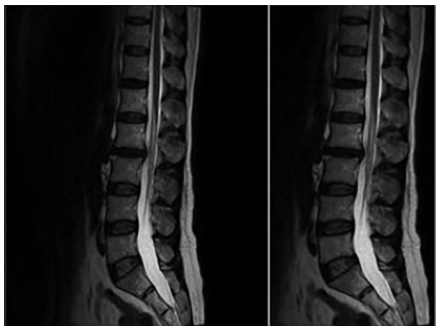

Figure 4 Left, The original image. Right, Image after removal and ROI selection.

As discussed in section "Otsu thresholding method", we use automatic Otsu thresholding to extract spinal canal. In this method, discrimination and extraction of spinal canal is done by finding a distinct threshold automatically and dividing the gray level of pixels in a specific image into two classes. Figure 5 shows the images after applying the Otsu method. It is essential to refine the image of extracted spinal canal and prepare it for later processing. We did this primary processing by means of morphological operations and filtering methods. Figure 6 demonstrates results of such preprocessing operations. Besides T2-weighted sagittal MRI, another type of image called MRM image is available for physicians to diagnose disc herniation. Spinal cord is identifiable in MRM images. Figure 7 demonstrates a sample of MRM images.

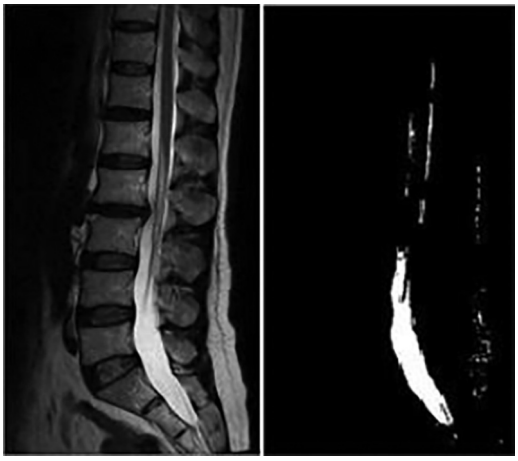

Figure 5 Left, The original image after ROI selection. Right, Image after extracting spinal canal based on one specific threshold.

Figure 6 Spinal canal after preprocessing operations and improvement.

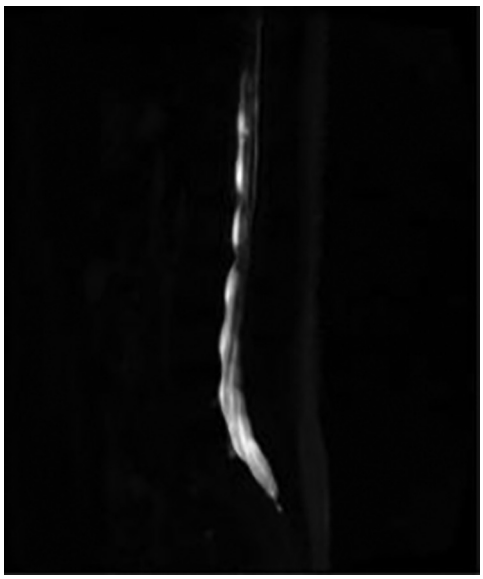

Figure 7 Magnetic resonance myelographic image.

Regarding the fact that the spinal cord images obtained from the preprocessed T2-weighted sagittal images have lower quality than MRM images, to be able to align a more suitable third-order polynomial on it, we try to achieve better image in this process through matching these two images with each other in a way that the spinal cord in these images would fit well on each other. Figure 8 shows the results of this matching used for fitting a third-order polynomial. To extract the boundary of discs, we first rotated Figure 8 to 90° and then fit a third-order polynomial on the spinal cord. Figure 9 shows the results of this alignment.

Figure 8 Result of matching Figure 5 with magnetic resonance myelographic binary image.

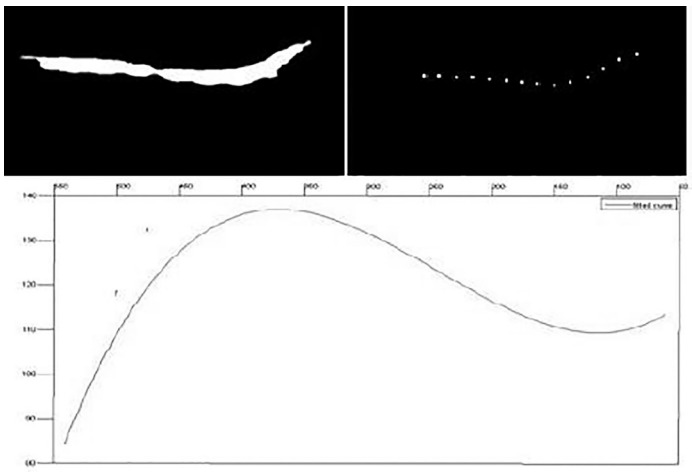

Figure 9 Top left, Spinal cord after 90° rotation. Top right, Tracking point on spinal cord. Below, Fitting third-order polynomial.

In the first stage, we tried to obtain a suitable binary image to reveal the discs. For this purpose, we converted the image to black and white using a threshold to achieve a suitable binary image. Then by using morphological operations, we made this image smoother. Figure 10 depicts the resulting binary image and the image after conducting morphological operations. To extract the location of the discs, a 50 × 40-pixel rectangle has been slid longitudinally along the third-order polynomial perpendicular to the curve. The sliding steps were chosen empirically as 1/120 of the total curve length. The procedure of the discs' location extraction is as follows.

The aforementioned rectangle has been used as a mask on the binary image developed in the previous step (Figure 11). At each 120 positions of the rectangle, we calculated the total number of the white pixels inside the masked area. This number is plotted against the step number in Figure 12.

Figure 10 Left, Binary image after applying threshold. Right, Smooth image by means of morphological operators.

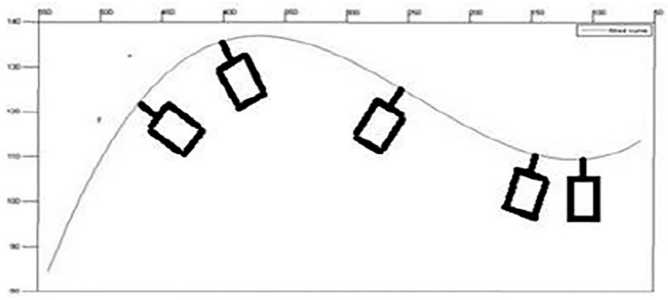

Figure 11 Sliding rectangle on third-order polynomial.

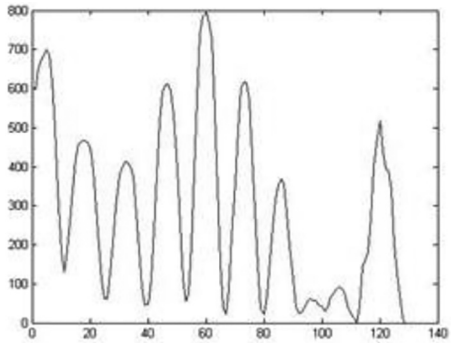

Figure 12 The peaks show the maximum overlap of the rectangle with discs in 120 steps.

As illustrated in Figure 12, it can be concluded that where the maximum overlapping between the rectangular area and the binary image's white pixels occurs, the location of one disc is found.[13]

As mentioned earlier, each peak in the above diagram depicts a disc's place. It should be mentioned that the maximum points in Figure 12 have been extracted by Pan-Tompkins algorithm.[17–20][I] In the second step, by approaching each disc's location, another rectangular area with the size of 140 × 70 pixels is placed on the image in a way that it covers the whole disc. The size of this rectangle has been determined through trial and error. These observations have been shown in Figure 13. For disc labeling, we considered the diagram of illumination intensity. Figure 14 shows the intensity of the image's pixels through the length of the image. At each point, the intensity is calculated as the summation of the intensity of the pixels in each column of the image. With regards to the fact that the spinal cord area in disc image has higher resolution, the maximum X occurs at the spinal cord area. This characteristic is shown at the right side of Figure 14. First, we calculate this maximum value (X_1) and then begin to find a peak at the left side of Figure 14 (X_2), which is at the left side of disc and is certainly out of the disc range. After achieving these two values for X, X_1 and X_2, we begin to obtain the mid value called X_m.

Then, at the dissection X_m, in direction to Y-axis has been shown in Figure 14, with the use of darkness of the lower and upper bound having the least intensity, we obtained the points related to the edge, named A and B in Figure 14. In fact, by drawing this diagram in direction to dissection related to X_m, these two points are specified as a minimum, as shown in Figure 15. As mentioned earlier, the lower and upper edges of disc were cut by vertical cutting X_m in point A and B, and with respect to the modification of resolution of pixels at the direction to mid cutting, when cross from vertebra and enter to disc (point A) and as well when exit the disc (point B) in resolution diagram, we had two minimum that were detectable easily and in such a way the points A and B were obtained. Then, for tagging at appropriate point in disc, the mid transverse point between A and B (point C in Figure 16) is obtained, and thus, we could tag the specified discs by using this method.

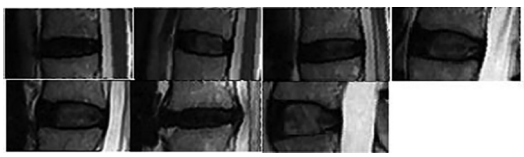

Figure 13 The extracted seven lumbar discs.

[I]Ref. 19: Article in Persian with an abstract in English

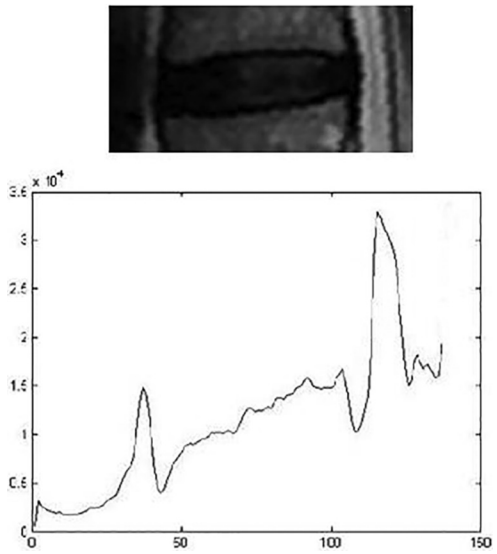

Figure 14 Disc and its intensity diagram.

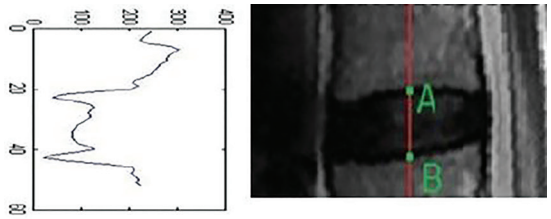

Figure 15 Disc image and its intensity diagram in direction of red line.

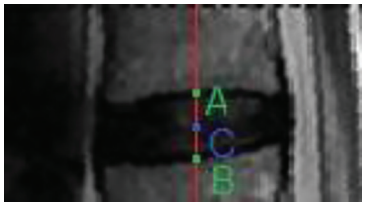

Figure 16 The mid-point C between A and B.

Feature extraction

Selection of ROI from disc for intensity feature extraction

To obtain a suitable ROI in disc range, which can be used for feature extraction to classify normal and herniated discs, we act as follows.

First, we considered a cutting in parallel to vertical cutting X_m on disc, and the cross points of the lower and upper edges of the disc were called q and p, respectively, and next, the mid-point between q and p was obtained similar to that of mid-point C and we called it "r". Figure 17 shows this parallel cutting and the mid-point r. With points r and C, the linear equation, crossing these two points, was obtained, and for each of these related images, we achieve a suitable binary image through extracted discs as shown in Figure 16 (right). The line d_1, crossing the points C and r, has crossed from the last white pixel at the right side at point l_1, and we considered this point as a most straight disc point. After obtaining the right side point, which is actually the right side peak of intensity diagram, we obtained a point in the left side and considered it.

Figure 17 Left, Binary image of disc by means of Otsu method. Right, Two parallel slices inside boundary of disc.

Then by means of four points around the disc, we obtain ROI of disc and we draw a rectangle around the disc, as shown in Figure 18. We can extract intensity feature of discs from this image. The herniated discs have dark pixels in ROI region. One of the main features that distinguish normal and herniated discs is the intensity of discs boundary in T2-weighted MR images. Thus, the average intensity within the disk drive can be used as a criterion for diagnosis of herniated discs.

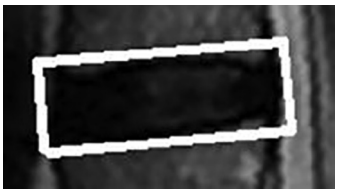

Figure 18 Disc ROI obtained by means of four points in top, bottom, left, and right.

Feature selection

The most appropriate measure of the radiologists for the diagnosis of LDH is the severely herniated disc. In the abnormal disc, substances compress the spinal cord and roots. The degree of lesions depends on the nucleus pulpous state. For the system to automatically distinguish between normal and herniated disc, in terms of this characteristic, we used binary images obtained by the Otsu method (Figure 19).

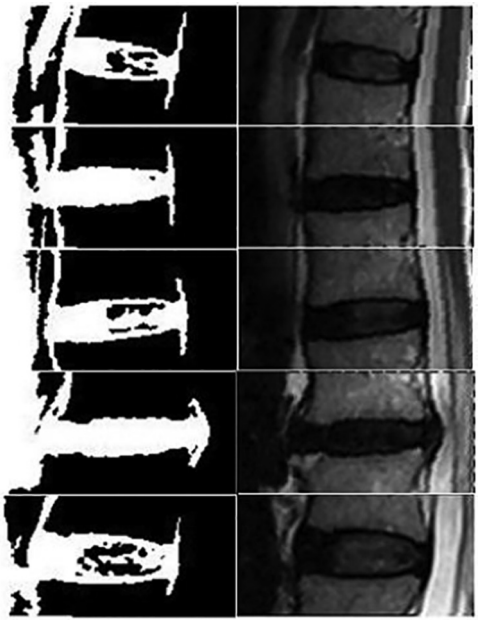

Figure 19 Extracted discs from magnetic resonance images and binarization of them by means of Otsu method for herniation feature extraction.

As shown in Figure 18, features of a herniated disc are visible at the right side of binary images. Important points that can be used as a suitable measure of binary image of each disc has been shown in Figure 20 (points "t", "l_1", and "b").

To obtain the coordinates of the point "t" in Figure 18, we started from top right corner of the image, pixel by pixel in the specified range as shown in Figure 18, and we store the coordinates of the first pixel with white intensity value as point "t". Similarly, the coordinates of the point "b" are obtained by storing the first white pixel in the lower right. In the next step, having three points "t", "b", and "l_1", we modeled herniated features with an angle (Figure 21).

According to the investigation of cases and controls identified in consultation with the radiologist, there are almost significant differences in abnormal and normal angle of eigenvectors $\vec{bl_1}$, $\vec{tl_1}$. So get this angle, a suitable feature for classification.

Figure 20 The binary image and points "t", "b", and "l_1".

Figure 21 Eigenvectors $\vec{bl_1}$ and $\vec{tl_1}$.

Calculating shape feature

We use an important, but simple shape feature R given by $R = a/b \approx w/h$, where a and b are the lengths of the major axis and the minor axis of the disc, respectively, whereas w and h are the width and height of the disc ROI bounding box, respectively (Figure 22).

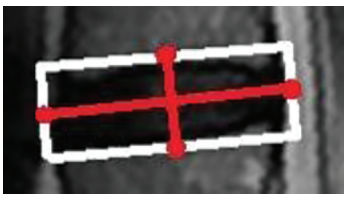

Figure 22 Lines of the major axis and the minor axis of the disc.

Local subset feature selection

Several methods have been proposed in the literature for feature selection, most of them tend to select global features, i.e., they select a single subset of features for classification in the whole sample space. Nevertheless, there might be cases in which different subsets of features are the most informative ones for classification in different parts of the sample space.[21,22] The proposed method is premised on the idea of formulating the problem of feature subset selection as a sequential decision-making problem through the agency of feature trees. As mentioned, we are interested in partitioning the sample space into a number of localities and selecting features for each of them. In this regard, to profit from axis-aligned localities, it is assumed that the partitioning can be represented using a univariate binary decision tree whose benefit is to ensure that the representation of the localities will depend only on a limited number of features.

A decision tree is typically composed of two types of nodes: splitting nodes and leaf nodes. The former signifies a split in the sample space and thus have two children and the latter, on the other hand, is attributed to a single locality. It is worthwhile noting that the decision boundary in splitting nodes is determined based on the mean value of the corresponding feature over the training samples corresponding to that subtree. In pursuance of local feature selection, the notion of decision tree has been developed to obtain a unified model, known as the feature tree, to represent both localities and corresponding selected features. In this regard, another type of node is put forward by the name of feature node, which signifies a feature that is assigned to all of its descendant localities, and may have at most one child. In addition, the concept of a compound locality is used to refer to each subtree that corresponds to a set of neighbor localities. This representation facilitates the selection of similar features for neighbor localities as they are known to be more likely to share a number of features.

A feature node is represented by a circle with a single feature inside, and a splitting node is depicted by a rectangle containing a feature and a threshold. Localities are shown by leaves. The subtree r_i corresponds to the compound locality cl_i, which consists of two single localities. It should be noted that the mutual features of neighbor localities are factored together in the parent feature node.

Applying a feature tree enables us to assign a training/test sample to a unique leaf. This will be achieved through attributing the sample to one of the descendants in the root repeatedly until it is assigned to a leaf. Therefore, in each locality, we accumulate a subset of training samples and a subset of features, i.e., the set of feature nodes from the leaf to the root. To classify a test sample, we first assign it to a locality based on the feature tree and subsequently classify it in the locality using the corresponding features and training samples.

Selecting suitable local features is, in this context, a matter of paramount importance, which requires to adopt a criterion to compare different feature trees. We expect that the selected features make the classification more convenient. In this respect, the samples of different classes should be separable in the new space, which is constructed by the selected features. One realization of this requirement is a space, where each sample and its neighbors are likely to belong to the same class. Taking this into consideration, we assume S and ft to be the training set and the feature tree, respectively. Given ft and an arbitrary sample x, we can find the subset of S that belongs to the same locality as x. Let $s \subset L(x, s, \text{ft}, k)$ be the k-nearest neighbors (KNNs) of x among the members of this subset. The score, i.e., fitness, of ft with respect to the training set S is calculated as follows:

$$\text{SCORE}(\text{ft}) = \frac{1}{K.|S|}\sum_{x \in s}\sum_{y \in L(x,s,\text{ft},k)} \begin{cases} 1 \text{label}(y) = \text{label}(x) \\ 0 \text{otherwise} \end{cases}, \qquad (18)$$

where label (.) gives the label, i.e., the class, of a sample.

It is noteworthy that, in this setting, each node of the tree is equivalent to a state for the reinforcement learning (RL) machine. This state consists of the sequence of nodes from the root to the current node in the tree. The RL agent selects an action at every state it arrives at. The decision expands the feature tree at its current node. Here, selecting an action for each node means choosing the node type (a feature node, a

splitting node, or a leaf node) and the corresponding feature index. Therefore, the set of all possible actions in each state is Actions = $\{f_1, f_2, \ldots, F_f, S_1, S_2, \ldots, S_F, T\}$, in which F is the number of features, f_i and s_i correspond to a feature node and a splitting node, respectively, and T is the terminating action. The termination action finishes feature selection in the current node, leaving it as a leaf.

Classification

In this study, we used three classifiers, the multilayer perceptron (MLP), KNN, and SVM, to differentiate between healthy subjects and patients. To evaluate the performance of classifiers, a 10-fold cross-validation method is used. MR images are divided into 10 parts, and the numbers of images in parts are equal unless in two or three groups. One part was used to test the classifier, and nine parts were used to train the classifier. This process was repeated 10 times for each different test set, and the average performance for AC, sensitivity (SN), and specificity (SP) was calculated. To increase the AC, we repeated each of the 10 processes 15 times.

Multilayer perceptron

The MLP neural network is applied to diagnose LDH disease. The classifier makes use of a three-layer MLP with the error back propagation algorithm and variable learning rate. The input layer has the same number of nodes as the input's vector length for each interval time. The output layer, on the other hand, contains one node accounting for a possibility of only two classes to be classified. In addition, all the possible combinations of the selected numbers of neurons in the hidden layer were selected and trained, leading to the optimized number being 5. It is worth mentioning that the training process has solely been conducted on the training data. Once the training error has dropped to a minimum, we move on to testing the network using the testing data. In fact, never has the data network observed the testing data when selecting the optimal architecture, which brings an improvement on the proper generalization of the network. The output nodes and the hidden layer use a linear transfer function and a sigmoid function, respectively. Network training proceeds until the mean value is less than 0.01, or the number of training iterations reaches 1000.

K-nearest neighbor

This classifier stores labeled feature vectors and calculates the minimum distance between stored and new feature vectors. The basic steps of the KNN algorithm are as follows:

To compute the distances between all samples that have already been classified into clusters.

To find the k samples with the smallest distance values.

To approve new data. A new sample will be added (classified) to the largest cluster out of k selected samples. We tested the values of k from 1 to 25 to make a comparison with,[23] and we found that $k = 5$ and $k = 15$ get the best results with the KNN classifier for 4 and 5 min before sudden cardiac death, respectively. We show three k values (5, 15, and 25) reduce the complexity of tables.

Support vector machine

This classifier is a supervised learning method, which is an extension of nonlinear models of the generalized portrait algorithm. The SVM algorithm is based on the statistical learning theory.[24,25] The goal of regression is to determine the best model from a set of models (named estimating functions) to approximate future values accurately. The generic support vector regression estimating function is shown in the following equation:

$$f(x) = (w \cdot \Phi(x)) + b, \tag{19}$$

where $w \subset R_n$, $b \subset R$, and Φ is a nonlinear function that maps x into a higher dimensional space. The weight vector (w) can be written as:

$$w = \sum_{i=1}^{L} (\alpha i - \alpha_i^*). \tag{20}$$

By substituting Eq. (19) into Eq. (20), the generic equation can be rewritten as:

$$f(x) = \sum_{i=1}^{L} (\alpha_i - \alpha_i^*) k(x_i) \cdot \Phi(x)) + b. \tag{21}$$

$$f(x) = \sum_{i=1}^{L} (\alpha_i - \alpha_i^*) k(x_i . x) + b. \tag{22}$$

In Eq. 22, the function $k(x_i x)$ is replaced with the dot product and known as the kernel function. The choice of kernel functions and kernel parameters usually depends on the application. Some of the useful kernel functions are radial basis functions (RBFs) and polynomial kernel functions. The formulas of these kernel functions are shown below:

$$\left\{ \frac{-|x - x_i|^2}{21^2} \right\}. \tag{23}$$

$$[(x^* x_i) + 1]^d. \tag{24}$$

In this study, RBFs and polynomial kernel functions were used with different sigma values ($\sigma = 0.8$, 1, and 1.2) and orders ($d = 1$, 2, and 3), respectively.

Evaluation

The ability of the proposed method for diagnosis of LDH disease is evaluated using AC, SN, SP, and precision (P). In Eqs. 25–28, TP refers to true positives (correctly diagnosed as LDH), TN refers to true negatives (correctly diagnosed as non-LDH), FN refers to false negatives (incorrectly diagnosed as non-LDH), and FP refers to false positives (incorrectly diagnosed as LDH).

AC is the ratio of correct diagnosis to the total diagnosis.

$$\mathrm{AC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}. \tag{25}$$

SN is the ratio of true positives to the total positives.

$$\mathrm{SN} = \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{TP}}. \tag{26}$$

SP is the ratio of true negatives to the total negatives.

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \tag{27}$$

P is the ratio of predicted positive cases that were correct.

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP}}. \tag{28}$$

RESULTS

The proposed method in this study has been evaluated through differentiating the patients from normal subjects. The results denote a noticeable capacity in classifying the two classes based on SVM, KNN, and MLP classifiers.[17–19,23–25][II] In this regard, using features extracted from the disc ROIs, we build three individual classifiers. The first classifier is a SVM, implemented using a linear kernel. This classifier shows 95.23, 98.78, 82.60, and 95% of accuracy, specificity, SN, and P, respectively. The second classifier is MLP classifier. We train the MLP to classify discs in two classes of normal and herniated. For training the MLP, 90% of the discs has been used as a training set and the rest as a test. The MLP and KNN classifiers show around 91.90 and 92.38% of accuracy, respectively. The statistical measures for validation by SVM, KNN, and MLP classifiers have been shown in Table 1. The results of our classifiers are shown in Table 2.

Table 1 The statistical measures for validation by SVM, KNN, and MLP classifiers.

	TP	FP	FN	TN
SVM	38	2	8	162
KNN	34	6	10	160
MLP	35	5	12	158

FN, false negative; FP, false positive; KNN, k-nearest neighbor; MLP, multilayer perceptron; SVM, support vector machine; TN, true negative; TP, true positive.

Table 2 Classifier performance results in percentage for 10-fold cross-validation

Classifier	Accuracy	Specificity	Sensitivity	Precision
SVM	95.23	98.78	82.60	95.00
KNN	92.38	96.38	77.27	85.00
MLP	91.90	96.93	74.46	87.50

KNN, k-nearest neighbor; MLP, multilayer perceptron; SVM, support vector machine.

DISCUSSION

In this article, a new approach for diagnosis of disc herniation has been proposed. In addition to noticeable classification accuracy, the proposed method is superior to others in terms of its simple implementation. Underpinned by mathematical logics, this method aids to diagnose the disc herniation in a highly accurate, precise, and automated manner, eliminating the need to refer to a doctor. Automated diagnostic systems have long been a significant necessity in medical society. Formerly presented methods have failed to fully satisfy this need on account of the corresponding complicated calculations and implementation, which lead to less time and cost-effectiveness (leading to increased expenses). The suggested method increases such aforementioned problems through offering a simpler implementation. A very important point is the location of disc. In some cases, it is just in the midline (called central) or near the midline (para-central), or lateral, or even far-lateral. In this study, the precise geometrical condition of disks is determined through the proposed mathematical method; it is needless to mention that the extracted information has been a major concern of radiologists.

[II]Ref. 19, 24 & 25: Article in Persian with an abstract in English

CONCLUSION

We proposed a method to detect herniated discs from sagittal and myelo lumbar MR images using robust intensity and texture features. Initially, we labeled the discs from sagittal images by fitting a third-order polynomial on the spinal cord, where we automatically localize a point inside each of the disc. The major advantage of this method is that it does not require precise segmentation of the lumbar intervertebral discs. Also, this approach extracts good features, which are witnessed from the high accuracies of the SVM and artificial neural network classifiers. Another added advantage is that the final majority voting classifier not only shows a high accuracy and SN but also elevates the assurance of the diagnosis. As an extension to this approach, we will be working on axial MR images to further decrease the diagnostic error rates. Moreover, we will also work on automatic classification of different subtypes of herniated discs that are bulged, protruded, extruded, fragmented, and migrated, and it is very important to define the subtypes.

On the other hand, analyzing a more number of features would be useful when attempting to increase the diagnosis accuracy. Making use of the mixture of experts classification, in another sense, ensures a precise decision-making on the output of different areas processes on the grounds that the final outcome is a product of different decision-makings from various perspectives towards this phenomenon. After applying as mentioned above, we pursue improvement of the studies that are conducted in this field.

The segmentation of the normal disc in this approach was very easy, but when a disc dehydrates, it is very difficult to find its margins. The mentioned problems continue to exist in our proposed method to spot discs' boundaries and thus to label the subjects as patients. This could be improved through image preprocessing methods in further studies.

REFERENCES

1. National Institute of Neurological Disorders and Stroke (NINDS), Low back pain fact sheet, NIND Brochure; 2008.
2. Luoma K, Riihimäki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. Spine. 2000;25:487.
3. Snell RS. Clinical anatomy by systems. Philadelphia, Pennsylvania: Lippincott Williams & Wilkins; 2006.
4. Alomari RS, Chaudhary V, Dhillon G. Computer aided diagnosis system for lumbar spine. Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), Barcelona, Spain; 2011.
5. Rijn JCV, Klemetso N, Reitsma JB, Majoie CBLM, Hulsmans FJ, Peul WC, et al. Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR Am J Roentgenol. 2005;184:299.
6. Alomari RS, Corso JJ, Chaudhary V. Abnormality detection in lumbar discs from clinical MR images with a probabilistic model. Proceedings of the 23rd International Congress and Exhibition on Computer Assisted Radiology and Surgery, Switzerland; 2009.
7. Alomari RS, Corso JJ, Chaudhary V, Dhillon G. Desiccation diagnosis in lumbar discs from clinical MRI with a probabilistic model. Proceedings of the 6th IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro (ISBI), Boston, MA; 2009. p. 546–9.
8. Alomari RS, Corso JJ, Chaudhary V, Dhillon G. Computer aided diagnosis of lumbar disc pathology from clinical lower spine MRI. Int J Comput Assist Radiol Surg. 2010;5:287.

9. Michopoulou S, Costaridou L, Panagiotopoulos E, Speller R, Panayiotakis G, Todd- Pokropek A. Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. IEEE Trans Biomed Imaging. 2009;56:2225.
10. Roberts N, Gratin C, Whitehouse GH. MRI analysis of lumbar intervertebral disc height in young and older populations. J Magn Reson Imaging. 1997;7:880.
11. Chevrefils C, Chériet F, Grimard G, Aubin C. Watershed segmentation of intervertebral disk and spinal canal from MRI images. Lect Notes Comput Sci ICIAR. 2007;4633:1017–1027.
12. Alomari RS, Corso JJ, Chaudhary V, Dhillon G. Toward a clinical lumbar cad: herniation diagnosis. Int J Comput Assist Radiol Surg. 2011;6:119.
13. Nikravan M, Ebrahimzadeh E, Izadi MR, Mikaeili M. Toward a computer aided diagnosis system for lumbar disc herniation disease based on MR images analysis. Biom Eng Appl Basis Comm. 2016;28:1650042.
14. Liao PS, Chen TS, Chung PC. A fast algorithm for multilevel thresholding. J Inf Sci Eng. 2001;17:713.
15. Senthilkumaran N, Vaithegi S. Image segmentation by using thresholding techniques for medical images. Comp Sci Eng Int J (CSEIJ). 2016;6:1.
16. Athertya J, Kumar G. Fuzzy clustering based segmentation of vertebrae in T-1 weighted spinal MR images. Int J Fuzzy Logic Sym. (IJFLS). 2016;6:2.
17. Ebrahimzadeh E, Pooyan M, Bijar A. A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. PLoS One. 2014;9:1.
18. Ebrahimzadeh E, Pooyan M. Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals. Biomed Sci Eng. 2011;11:699.
19. Ebrahimzadeh E, Pooyan M. Prediction of sudden cardiac death (SCD) using time-frequency analysis of ECG signals. ISEE. 2013;3:15.
20. Ebrahimzadeh E, Pooyan M, Jahani S, Bijar A, Setaredan SK. ECG signals noise removal: selection and optimization of the best adaptive filtering algorithm based on various algorithms comparison. Biomed Eng Appl Basis Commun. 2015;27:1550038.
21. Ebrahimzadeh E, Manuchehri MS, Amoozegar S, Araabi BN, Soltanian-Zadeh H. A time local subset feature selection for prediction of sudden cardiac death from ECG signal. Med Biol Eng Comput. 2017. [Epub ahead of print].
22. Ebrahimzadeh E, Araabi BN. A novel approach to predict sudden cardiac death using local feature selection and mixture of experts. Comput Intell Electr Eng. 2016;7:15–32.
23. Ebrahimzadeh E, Alavi SM, Bijar A, Pakkhesal AR. A novel approach for detection of deception using smoothed pseudo Wigner-Ville distribution (SPWVD). J Biomed Sci Eng. 2013;6:8.
24. Ebrahimzadeh E, Alavi SM, Samsami Khodadad F. Implementation and designing of lie-detection system based on electroencephalography (EEG). Sci Res J Army Univ Med Sci I.R. Iran. 2013;11:20.
25. Amoozegar S, Pooyan M, Ebrahimzadeh E. Classification of brain signals in normal subjects and patients with epilepsy using mixture of experts. ISEE. 2013;4:1.