

# Classifying patients with lumbar disc herniation and exploring the most effective risk factors for this disease

This article was published in the following Dove Press journal:  
*Journal of Pain Research*

Samira Jafari<sup>1</sup>  
Tania Dehesh<sup>2</sup>  
Farhad Iranmanesh<sup>3</sup>

<sup>1</sup>Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran; <sup>2</sup>Department of Biostatistics and Epidemiology, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran; <sup>3</sup>Department of Neurology, Kerman University of Medical Sciences, Kerman, Iran

**Objectives:** To classify patients suffering from low back pain (LBP) into two different groups – patients with lumbar disc herniation (LDH) and patients without this disease based on simple questions and without magnetic resonance imaging (MRI) procedure – and to diagnose the most effective risk factors of LDH.

**Methods:** Four hundred patients aged over 18 years suffering from LBP for over 6 months were randomized into two groups in this cross-sectional study. The data were gathered at Besat clinic, in Kerman, southeast of Iran. Twelve dichotomous questions from the main LDH risk factors were asked. Three statistical classification methods – K-nearest neighbors (KNN), support vector machine (SVM), and logistic regression (LR) – were performed. LR was used in order to diagnose the most important risk factors of LDH.

**Results:** SVM method was more efficient among the small sample sizes, while KNN method showed the best classification relative to other methods when the sample size increased. LR model had the least efficiency of all. The drug use increased the chance of LDH more than 7 times (OR=7.249), and the chance of having LDH among people who had associated illness was 4.847 times more compared with people who did not have. Using hookah increased the chance of having LDH more than twice (OR=2.401), and the chance of smokers for LDH was near four times higher than nonsmokers (OR=3.877).

**Conclusion:** The statistical classification methods had acceptable precisions for diagnosis of LDH patients. It is suggested that neurologists become more familiar with these methods and use them before MRI prescription to decrease the unnecessary burden on health services. Addiction to drugs, cigarettes, and hookah is the main factor in the creation of a lumbar disc herniation.

**Keywords:** classification, K nearest neighbors, support vector machine, logistic regression, lumbar disc herniation

## Introduction

One of the common chronic disorders that have increased in recent decades is low back pain (LBP).<sup>1</sup> There is no exact agreement about the definition of LBP in most literature.<sup>2</sup> This disease can vary from a permanent ache to a sudden severe pain in the patients' back.<sup>3</sup> This disorder can be classified depending on the duration of the pain. The chronic LBP lasts between 6 and 12 months.<sup>4</sup> The reported complications of LBP are excessive. In some cases, it causes severe pain that brings motionlessness and weakness that lowers the quality of life.<sup>5</sup> This disease disrupts the daily activities of more than 80% of the people around the world.<sup>6</sup> It is reported that

Correspondence: Tania Dehesh  
Department of Biostatistics and  
Epidemiology, School of Public Health,  
Kerman University of Medical Sciences,  
Haft Bagh Alavi, Kerman, 7617647633,  
Iran  
Tel +98 343 132 5069  
Email tania\_dehesh@yahoo.com

annual prevalence of this disease is between 15% and 45%, and it involves an average of 30% of the people.<sup>7</sup>

Often, all people have experienced LBP between the ages of 30 and 50.<sup>8</sup> One of the most prevalent reasons of LBP is lumbar disc herniation (LDH) which is a musculoskeletal disease.<sup>4</sup> The most important and effective causes of LDH are genetic, biochemical, traumatic, and psychosocial. In most cases, this disease is inherited from parents to children.<sup>9</sup> The usual procedure for diagnosing LDH is labeling the hernia discs by magnetic resonance imaging (MRI).<sup>6</sup> MRI is a medical procedure that is often used in the assessment of people with spinal problems. This procedure has huge budget burden on patients and health-care services.<sup>10</sup>

The increased prevalence of LBP creates an economic burden on society and health services because usually all the patients with backache are referred to an MRI center by the physicians.

Despite incredible benefits of MRI in diagnosing the main cause of LBP such as spinal discs, ligaments, vertebral bodies, vascular structures, muscle tissues, disc degeneration, and lumbar spinal canal stenosis,<sup>11</sup> it may be possible to diagnose the back pain patients due to LDH in the early stages with other procedures rather than MRI.

Many of the patients with LBP need conservative treatments first, and MRI diagnostic procedure could be the next step if the pain worsens. The classification methods have old history in statistics which are used in classifying the observations in similar classes. LR classic classification method has simple interpretations but not in complicated data.<sup>12</sup>

Recently, machine learning methods, which diagnose the existing classes automatically, were introduced to the classification fields.<sup>13</sup> The most frequently used methods in machine learning that classify observations are KNN and SVM.<sup>14</sup>

Despite traditional classification method, LR, the machine learning methods are relaxed from linearity combination of predictor variables in order to find the best decision boundary. Logistic regression (LR) focuses on maximizing the likelihood ratio function in order to have the best coefficients and also prediction. SVM tries to find the separating hyperplane which has the biggest distance to the closest points (the support vectors). If the cases could not be separated by linear function, this procedure could project the cases to higher dimensions with kernel functions and separate them perfectly. KNN method also classifies cases based on distances, not

likelihood function. In fact, this procedure is free from probability density function for outcome.<sup>15</sup> SVM and KNN are called the newest classification methods and LR is the most famous one.<sup>16</sup> Many of these classification methods can classify patients into two groups with high precision. The performances of these classification methods are based on simple information provided by asking simple questions from the patients without incurring any cost.

Using classification methods in statistics, physicians can avoid sending people who have backache for an MRI unnecessarily, and for those patients who do not belong to the LDH group, remedies can be prescribed first. Therefore, the patients can be protected from the stress of MRI, as well as the unnecessary financial burden. This important goal needs the cooperation of biostatistics experts and physicians. Actually, different scientific fields must help each other in order to have better services in the society.

A study conducted by Pedro J. García-Laencina et al in 2015 assessed and predicted breast cancer in women. In this study, KNN and LR were used according to the specificity (SP) and concluded that the KNN model was more efficient than LR.<sup>17</sup> In another study conducted by Eun-Suk Yang et al, KNN and LR have been used to assess the best combination for ovarian cancer. In the present study, according to the values of receiver operating characteristic (ROC) curve, LR was recognized as the best model for the classification of individuals.<sup>18</sup> Chia-Hsun Hsieh et al assessed cancer screening in an asymptomatic population using multiple tumor markers. His results also indicated superiority of SVM over KNN and LR methods.<sup>19</sup>

The present study aims to compare classification methods (LR, KNN, and SVM) to classify patients with backache into two groups – patients with LDH and patients without LDH with high precision – and to explore and diagnose the most effective risk factors of LDH based on LR.

## Methods

### Data collection and preparation

The data of this cross-sectional study were collected from June to September 2017 at Besat clinic which is the main clinic for lumbar disorders in the center of Kerman, southeast of Iran. Eligible patients were 252 women and 148 men aged over 18 years suffering from LBP for at

least 6 months. Also, they complained of severe pain in their low back that disabled them in routine activities such as walking, sitting, and standing. These patients were referred to the MRI center of Besat clinic by the neurologists to investigate the main cause of their severe back pain. Exclusion criteria were lack of understanding Persian language and any kind of back surgery within the last 4 months.

The data were divided into two parts, which are called train and test sets. Three models were built with train set and then their classification efficiency was checked on test set. In fact, the test set did not have any role in building the models.

Each of the train and test tests was composed of 200 samples. The models were made based on training data, and then were implemented on the testing data for checking their efficiency better. In this study, the main variables, which were the most effective risk factors for LDH, were collected through a checklist. This checklist, which was developed under the guidance of a neurologist, consisted of 12 questions. The questions are indicated in Table 1. The history of associated illnesses such as diabetes, kidney stones, hypertension, and underlying diseases such as the elongation of the ligaments or muscles of the lower back musculoskeletal system and low back arthritis<sup>20,21</sup> were also recorded as Yes or No.

The interviewing team was trained by the neurologist about the concept of checklist's questions before dispatching them to the clinic. The participants completed the written informed consent form prior to enrollment in the study. The study was approved by the ethical committee of our institution Kerman University of Medical Sciences (reference code: IR.KMU.REC.1397.078). The study was conducted in compliance with the Helsinki Declaration. The result of the outcome variable was being diagnosed after observing the result of MRI by the neurologist in the form of having lumbar disc/not having lumbar disc.

## KNN method

K-nearest neighbors (KNN) method represents one of the simplest algorithms of the learning machine where we have no model and estimation of parameters. In the KNN method,

the distances between all observations (patients, in our study) are calculated and the nearest observations construct a group. Therefore, observations are classified into K similar classes based on the distances between them. After building the classes based on distances, the new observation (new patient), which does not have any role in previous classification, is classified to the nearest class.

## SVMs method

The SVMs were introduced by Vapnik et al (1990) for the first time as a family of machine-learning methods. The aim of SVMs is to separate observations by a line (two dimensions) or by a plan (more than two dimensions) in a complicated problem.<sup>22</sup>

Essentially, SVMs are looking for the best separating line between the classes.<sup>23</sup>

## LR method

The LR is a traditional statistical technique for classification. The same as the previous two methods, its outcome is a dichotomous variable. The method is a kind of regression method which needs a link function.<sup>24</sup> The explanatory (independent) variables may be continuous, discrete, or combined.<sup>24</sup>

Classifying the training data with three classification methods was done with R software version (3.4.4), packages caret, and e1071. R software is free and also the most used, especially when program writing is needed. The models were examined with the testing data, and the best models are introduced based on four classification criteria. These diagnosis criteria are sensitivity (SE), SP, correct classification rate (CCR), and kappa coefficient (KC).

SE evaluates the validity of the test in detecting the real patient. SP estimates the number of healthy patients that must be excluded. CCR measures how correctly a diagnostic test identifies and rules out a certain condition. CCR of a diagnostic test can be determined from SE and SP with the prevalence.<sup>25</sup> The model with the highest value of SE, CCR, and KC and the lowest value of SP is identified as the best classification model.

SE, SP, and CCR are described in terms of TP, TN, FN, and FP.

---


$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) = (\text{Number of true positive assessment}) / (\text{Number of all positive assessments})$$


---

**Table 1** The list of 12 simple questions (Q)

Checklist's items	With LDH, n (%) or mean $\pm$ SD	Without LDH, n (%) or mean $\pm$ SD
<b>Q1: Sex</b> Female Male	89 (35.2) 48 (32.7)	164 (64.8) 99 (67.3)
<b>Q2: Any waist injury</b> Yes No	16 (38.1) 121 (33.8)	26 (61.9) 237 (66.2)
<b>Q3: Any sensory disorder in the body</b> Yes No	92 (33.7) 45 (35.4)	181 (66.3) 82 (64.6)
<b>Q4: Any movement disorder in the body</b> Yes No	45 (34.1) 92 (34.3)	87 (65.9) 176 (65.7)
<b>Q5: Any associated illness, such as diabetes, kidney stones, hypertension</b> Yes No	59 (38.1) 78 (31.8)	96 (61.9) 167 (68.2)
<b>Q7: Daily smoking</b> Yes No	14 (56.0) 123 (32.8)	11 (44.0) 252 (67.2)
<b>Q8: Daily drug use</b> Yes No	16 (48.5) 121 (33.0)	17 (51.5) 246 (67.0)
<b>Q9: Daily Hookah use</b> Yes No	10 (41.7) 127 (33.8)	14 (58.3) 249 (66.2)
<b>Q10: How long have you experienced pain in the back (in months)</b>	66.83 $\pm$ 89.593	60.86 $\pm$ 80.886
<b>Q11: How long have you experienced severe pain that disrupts daily activity (in months)</b>	12.41 $\pm$ 25.570	12.68 $\pm$ 29.099
<b>Q12: Age</b> Height (m) Weight (kg)	46.70 $\pm$ 14.287 1.62 $\pm$ 1.109 71.64 $\pm$ 15.100	44.94 $\pm$ 14.276 1.62 $\pm$ 0.108 70.66 $\pm$ 15.456

**Abbreviation:** LDH, lumbar disc herniation.

The homogeneity of training and testing groups based on demographic variable were tested by Chi-square test and independent samples *T* test. To explore the effect of sample size on the efficiency of three models, the above classification models were performed in different sample sizes (50, 100, 150, and 200). LR model was used for diagnosing the most effective risk factors of LDH. The odds ratio (OR) index was used to reflect the rate of risk factor effects.

## Results

Table 2 describes the demographic characteristics of all the patients in the present study. As depicted, 36.8% were male and 63.2% female. Also, 137 out of 400 (34.2%) patients that were referred to the neurologist had herniated discs according to MRI results. The mean ( $\pm$ SD) age of the patients was 45 years. It showed that people who suffered from LBP were middle-aged and their mean value of BMI was in the normal range.

**Table 2** Subject characteristics based on sex

Variables	Total (n=400)	Women (n=253)	Men (n=147)	P-value
Age (mean $\pm$ SD)	45.47 $\pm$ 14.77	47.31 $\pm$ 14.12	43.64 $\pm$ 15.26	0.091
BMI (mean $\pm$ SD), kg/m <sup>2</sup>	27.06 $\pm$ 5.02	27.90 $\pm$ 4.76	26.18 $\pm$ 5.18	0.235
Duration of smoking (mean $\pm$ SD) (days)	344.30 $\pm$ 1754.38	100 $\pm$ 1186.52	764.76 $\pm$ 2387.47	0.002
Duration of drug use (mean $\pm$ SD) (days)	388.89 $\pm$ 1841.24	157.25 $\pm$ 1009.60	787.55 $\pm$ 2693.05	0.007
Duration of hookah use (mean $\pm$ SD)(days)	58.09 $\pm$ 310.51	32.67 $\pm$ 183.79	101.84 $\pm$ 449.61	0.077
Duration of LBP (mean $\pm$ SD) (days)	1887.27 $\pm$ 2517.18	1962.35 $\pm$ 2376.65	1758.06 $\pm$ 2745.93	0.435
Duration of severe LBP (mean $\pm$ SD) (days)	377.61 $\pm$ 837.27	433.02 $\pm$ 941.46	282.24 $\pm$ 609.48	0.053
LDH, n (%)				
Has LDH	137 (34.20)	90 (35.60)	49 (33.30)	0.753
Does not have LDH	263 (65.80)	163 (64.60)	98 (66.70)	
Waist injury, (%)				
Yes	42 (10.50)	24 (9.50)	18 (12.20)	0.192
No	358 (89.50)	229 (90.50)	129 (87.80)	
Sensory disorders, number (%)				
Yes	273 (68.20)	181 (71.50)	93 (63.30)	0.282
No	127 (31.80)	72 (28.50)	54 (36.70)	
Movement disorders, number (%)				
Yes	132 (33)	90 (35.60)	43 (29.30)	0.750
No	268 (67)	163 (64.40)	104 (70.70)	
Smoking, number (%)				
Yes	25 (6.20)	4 (1.60)	20 (13.60)	1
No	375 (93.80)	249 (98.40)	127 (86.40)	
History of drug use, number (%)				
Yes	33 (8.20)	11 (4.30)	22 (15)	0.364
No	367 (91.80)	242 (95.70)	125 (85)	
History of hookah use, number (%)				
Yes	24 (6)	11 (4.30)	13 (8.80)	0.068
No	376 (94)	242 (95.70)	134 (91.20)	
Associated illness, number (%)				
Having	70 (17.50)	46 (18.20)	24 (16.30)	0.638
Not having	330 (82.50)	207 (81.80)	123 (83.70)	
Underlying disease, number (%)				
Having	155 (38.80)	121 (47.80)	34 (23.10)	<0.001
Not having	245 (61.20)	132 (52.20)	113 (76.90)	

**Abbreviation:** LDH, lumbar disc herniation.

As observed, the majority of LBP patients were women (n=253, 63.2%) in proportion to men (n=147, 36.8%), but the number of herniated discs between them was not statistically significant ( $P=0.753$ ). The main cause of LBP in patients was not a hit to the back (89.5%). Most patients had a sensory disorder (89.5%), but not any

movement disorder (33%). Against the neurologist's expectation, the prevalence of cigarettes (6.2%) and drug addiction among patients was low (8.2%). These results exhibited that the main reason of LBP in most patients was not the usual reasons that were categorized in the checklist as the main factors by neurologist. Comparing the two

**Table 3** The results of comparing KNN, SVM, and LR in different sample sizes

N	Data set	KNN				SVM				LR			
		SE	SP	KC	CCR	SE	SP	KC	CCR	SE	SP	KC	CCR
50	Train	0.23	0.94	0.21	0.70	0.18	1	0.22	<b>0.72</b>	<b>0.29</b>	<b>0.90</b>	<b>0.23</b>	0.70
	Test	0.47	0.81	0.29	0.68	0.47	1	<b>0.53</b>	<b>0.80</b>	0.47	0.90	0.42	0.74
100	Train	0.03	0.98	0.02	0.66	<b>0.85</b>	1	<b>0.88</b>	<b>0.95</b>	0.17	0.90	0.10	0.66
	Test	0.36	0.92	0.32	0.72	<b>0.89</b>	<b>0.98</b>	<b>0.89</b>	<b>0.95</b>	0.11	0.95	0.08	0.65
150	Train	0.02	0.98	0.14	0.65	<b>0.45</b>	<b>0.99</b>	<b>0.50</b>	<b>0.81</b>	0.04	0.95	0.001	0.64
	Test	<b>0.32</b>	0.92	<b>0.27</b>	<b>0.71</b>	0.06	1	0.07	0.67	0.07	0.96	0.042	0.65
200	Train	0.01	<b>0.98</b>	0.01	0.65	<b>0.66</b>	<b>0.98</b>	<b>0.69</b>	<b>0.87</b>	0.29	0.90	0.23	0.70
	Test	<b>0.31</b>	0.93	<b>0.27</b>	<b>0.71</b>	0.15	<b>0.99</b>	0.181	0.69	0.12	0.96	0.10	0.66

**Notes:** The highest value of all four criteria (SP, SE, CCR and KC) shows the best model in each sample size, therefore these values are shown in bold.

**Abbreviations:** N, sample size; SE, sensitivity; SP, specificity; KC, kappa coefficient; CCR, correct classification rate; KNN, K-nearest neighbors; LR, logistic regression; SVM, support vector machine.

patient groups (train and test), there were no statistically significant differences in all demographic variables. This result demonstrates the homogeneity of two groups according to demographic variables.

The result of this table also shows that there is no statistically significant difference between women and men in terms of demographic and diagnostic LDH variables, except for the duration of drug, cigarettes, and hookah use. Men used all of them significantly longer than women. Women had significantly higher levels of LDH than men ( $P<0.001$ ).

Table 3 shows the results of comparing four criteria (SE, SP, KC, and CCR) in different sample sizes (50, 100, 150, and 200) between three classification methods. As explained in the method section, for best classification, SE, KC and CCR values must be high and SP value must be low. So, in general, the SVM model is useful in classifying low sample size and as the sample size increases, the efficiency of KNN model increases. Also, according to the four criteria, LR model had the least efficiency in all the sample sizes. The comparison of three models with four criteria in different sample sizes is shown for more clarification (Figure 1).

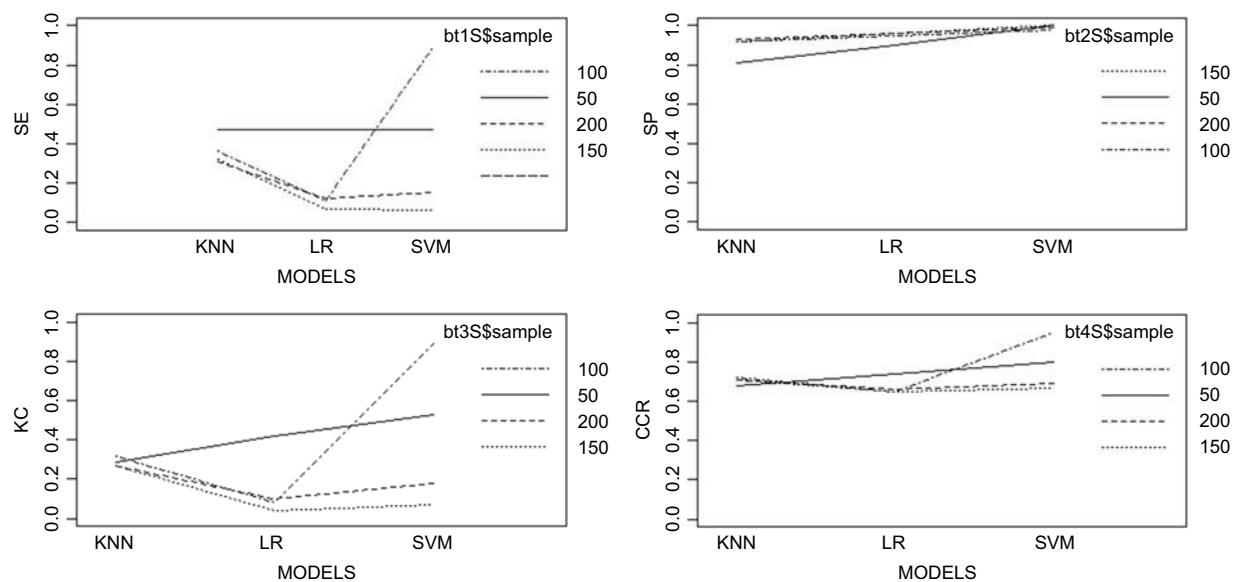
Table 4 shows the results of LR model. According to the values of OR, the most effective risk factors of LDH are the use of drug, associated illness of LDH, use of hookah, and cigarettes, respectively. The use of drug increased the chance of LDH more than 7 times ( $OR=7.249$ ) and the chance of having LDH in people who have associated illness 4.847 times comparing with people who do not have. Using hookah increased the chance of having LDH more

than twice ( $OR=2.401$ ) and the chance of smoker for LDH is near four times higher than nonsmokers ( $OR=3.877$ ). The chance of women was near 1.5 times more than men for LDH ( $OR=1.38$ ) and the sensory disorder in the back increased the chance of LDH ( $OR=1.5$ ). Other risks factors did not significantly affect the development of LDH.

## Discussion

The results of the present study provide a perspective for researchers who want to use simpler statistical diagnostic methods before clinical diagnostic. The present results revealed that basic sciences could have an important role in clinical diagnostic which has been rarely mentioned in the previous studies. In this study, the results of three statistical classification methods KNN, SVM, and LR were compared with the results of MRI in diagnosing patients with herniated discs to determine whether they could be used as the former step of clinical methods to avoid the burden of stress and expense on the patients. For this comparison, simple questions have been asked from the patients in the first step of diagnosis. Since this is the first study appraising these three classification methods with dichotomous questions to diagnose patients with herniated disks, there were no directly comparable studies in this context. Most of the previous studies compared classification methods with qualitative questions. Four criteria which are SE, SP, KC, and CCR in four different sample sizes were used to compare. For small sample sizes (50, 100), the SVM models had better classification based on the four criteria, and the KNN model displayed the best classification for larger sample sizes (150, 200). These





**Figure 1** The comparison of three models with four criteria in different sample sizes.

**Abbreviations:** SE, sensitivity; SP, specificity; KC, kappa coefficient; CCR, correct classification rate; KNN, K-nearest neighbors; LR, logistic regression; SVM, support vector machine.

**Table 4** The effects of risk factors on LDH

Variables	$\beta$	P-value	95% CI	OR
Intercept	-4.567	0.002	-7.417 to 1.716	0.010
Age in years	0.003	0.761	-0.015 to -0.02	1.003
BMI	-0.013	0.595	-0.06 to -0.034	0.987
Sex	-0.326	0.227	-0.856 to -0.203	1.388
History of hit to the lumber	0.392	0.303	0.353 to 1.137	1.480
Sensory disorders	-0.397	0.122	-0.900 to -0.107	1.488
Movement disorders	-0.026	0.920	-0.54 to -0.487	0.974
Smoking	1.355	0.049	-0.004 to 2.706	3.877
History of drug use	1.981	0.005	-0.611 to 3.351	7.249
Duration of smoking	-3.051	0.730	0.000	1.000
Duration of drug use	0.000	0.015	8.072 to 0.001	1.000
Duration of hookah use	0.001	0.366	0.000 to 0.002	1.001
Duration of LBP	-1.669	0.731	0.000 to 7.860	1.000
Duration of severe LBP	9.914	0.508	0.000	1.000
History of hookah use	0.876	0.036	-0.573 to 2.324	0.036
Associated illness	1.578	0.000	-0.997 to 2.160	4.847
Underlying disease	0.193	0.453	-0.311 to 0.697	1.213

results depict that SVM model classifies observations better than other methods, especially in small sample sizes. This may be due to the fact that in small sample sizes with dichotomous questions the information gained is little. Since the SVM model does not pay attention to the distribution of data and works based on constructing a high-dimensional plan between observations, these results are in accordance with actual expectations. In

contrast, the KNN model is made based on distances between observations, and calculating the distances with quantitative values is simpler relative to dichotomous data. The present results are also in accordance with those of the previous works, such as those noted in the study by Wu Yunfeng in 2016 about knee vibration quantum. In the present study, the comparison is between SVM and LR, and SVM model was chosen in competition with LR.<sup>26</sup> LR

method is a completely parametric method that displays the best result in follow-up studies, but KNN method is not a parametric method, and it is conducted based on distances between observations. Therefore, when the number of predictors becomes more, KNN procedure shows better results.

The present researchers' findings reveal that sample size has the most influence on the efficiency of KNN compared with the other two methods. Another important finding is that addiction increases the chance of LDH, especially use of drugs, hookah, and cigarettes. These findings are completely in accordance with the previous study.<sup>27</sup> The chance of having LDH for people who have sensory disorder and also women is higher than in others. They should be more careful about their back. This result also approves previous results.<sup>28</sup>

An important strength of the present study is that the predictor variables are made up of the simplest possible questions and the physician collects the information without any measurements. The researchers tried to depict the precision of statistical models in proportion to the clinical method (MRI) in the correct diagnosis of patients. Hence, if there is only one computer system with statistical programs in the doctor's office s/he can ask patients suffering from LBP the simplest questions and enter their information into the system. If the patient is classified in the group of LDH patients, then the neurologist could refer the patients to MRI for definitive diagnosis. Finally, most studies agree with the present results. Small differences exist because of the differences between type and number of predictors. Therefore, offering methods which could diagnose patients without any measurement is valuable. An important strength of the present study is that today, basic sciences like biostatistics and medicine must cooperate to avoid the burden of many emotional and financial costs on patients. Separating different scientific branches is completely wrong when they could be complementary. The present study had some limitations that merit attention when interpreting the findings. First of all, most of the 12 questions in the checklist were based on one neurologist experiences and also neurology text. In fact, there are other valuable questions that could be taken into account in this study. Maybe other neurologists have useful experiences. Furthermore, diagnosis of LDH was done by one neurologist who could diagnose incorrectly. It could be better to have other neurologists' cooperation in order to prevent incorrect diagnosis.

Planning to compare these three models with a variety of questions and cooperation of more than one neurologist in order to demonstrate the efficiency of models is necessary.

## Conclusion

To the best of the present researchers' knowledge, this study was one of the first ones that compared classification methods when the majority of predictors are binomial. This study confirms the precision of statistical methods in classifying patients which is near MRI results. Hence using them could prevent unnecessary MRI, especially in the first stage of the disease. Since the mean age of our patients' population is not high, we could emphasize that different types of addiction were the main cause of LDH in young people.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Vivas ELA, Oliva FJC, Aguilar RQF, González AG, Cruz JG, Ávila MJM. Application of a semiautomatic classifier for modic and disk hernia changes in magnetic resonance. *Columa/Columna*. 2015;14(1):18–22. doi:10.1590/S1808-1851201514010R121
2. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet*. 1999;354(9178):581–585. doi:10.1016/S0140-6736(99)01312-4
3. Sheet. National Institute of Neurological Disorders and Stroke, National Institutes of Health; updated February 2011. Available from: <https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets>. Accessed March 21, 2019.
4. Mio F, Chiba K, Hirose Y, et al. A functional polymorphism in COL11A1, which encodes the  $\alpha 1$  chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet*. 2007;81(6):1271–1277. doi:10.1086/522377
5. Buchmann N, Preuß A, Gempt J, et al. Outcome after surgical treatment for late recurrent lumbar disc herniations in standard open microsurgery. *World Neurosurg*. 2016;89:382–386. doi:10.1016/j.wneu.2016.02.028
6. Ghosh S, Chaudhary V. Supervised methods for detection and segmentation of tissues in clinical lumbar MRI. *Comput Med Imaging Graph*. 2014;38(7):639–649. doi:10.1016/j.compmedimag.2014.03.005
7. Takahashi M, Haro H, Wakabayashi Y, Kawa-Uchi T, Komori H, Shinomiya K. The association of degeneration of the intervertebral disc with 5a/6a polymorphism in the promoter of the human matrix metalloproteinase-3 gene. *J Bone Joint Surg*. 2001;83(4):491–495. British volume.
8. Owlia M, Salimzadeh A, Alishiri G, Haghighi A. Comparison of two doses of corticosteroid in epidural steroid injection for lumbar radicular pain. *Singapore Med J*. 2007;48(3):241.
9. Kimura T, Nakata K, Tsumaki N, et al. Progressive degeneration of articular cartilage and intervertebral discs. *Int Orthop*. 1996;20(3):177–181.
10. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of magnetic resonance imaging readings for lumbar disc herniation in the Spine Patient Outcomes Research Trial (SPORT). *Spine*. 2008;33(9):991. doi:10.1097/BRS.0b013e31816c8379



11. Elliott J, Flynn T, Al-Najjar A, Press J, Nguyen B, Noteboom JT. The pearls and pitfalls of magnetic resonance imaging for the spine. *J Orthop Sports Physl Ther*. 2011;41(11):848–860. doi:10.2519/jospt.2011.3636
12. Rasmussen CE. Gaussian processes in machine learning. In: *Advanced Lectures on Machine Learning*. Springer; 2004:63–71.
13. Robert C. *Machine Learning, a Probabilistic Perspective*. 2nd ed. Abingdon-on-Thames: Taylor & Francis; 2014.
14. Rani P, Liu C, Sarkar N, Vanman E. An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Anal Appl*. 2006;9(1):58–69. doi:10.1007/s10044-006-0025-y
15. Lesmeister C. *Mastering Machine Learning with R*. Packt Publishing Ltd; 2015.
16. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Vol. 112. Springer; 2013.
17. García-Laencina PJ, Abreu PH, Abreu MH, Afonso N. Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Comput Biol Med*. 2015;59:125–133. doi:10.1016/j.compbimed.2015.02.006
18. Yang E-S, Kim J-D, Park C-Y, Song H-J, Kim Y-S. Best combination looking for ovarian cancer. 2015.
19. Wang H-Y, Hsieh C-H, Wen C-N, Wen Y-H, Chen C-H, Lu J-J. Cancers screening in an asymptomatic population by using multiple tumour markers. *PLoS One*. 2016;11(6):e0158285. doi:10.1371/journal.pone.0158285
20. Heliövaara M, Mäkelä M, Knekt P, Impivaara O, Aromaa A. Determinants of sciatica and low-back pain. *Spine*. 1991;16(6):608–614.
21. Kuisma M, Karppinen J, Niinimäki J, et al. Modic changes in endplates of lumbar vertebral bodies: prevalence and association with low back and sciatic pain among middle-aged male workers. *Spine*. 2007;32(10):1116–1122. doi:10.1097/01.brs.0000261561.12944.ff
22. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*. 2000;16(10):906–914.
23. Meyer D, Wien FT. Support vector machines. *R News*. 2001;1(3):23–26.
24. Subasi A, Ercelebi E. Classification of EEG signals using neural network and logistic regression. *Comput Methods Programs Biomed*. 2005;78(2):87–99. doi:10.1016/j.cmpb.2004.10.009
25. Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG Proc*. 2010;19:67.
26. Wu Y, Chen P, Luo X, et al. Quantification of knee vibroarthrographic signal irregularity associated with patellofemoral joint cartilage pathology based on entropy and envelope amplitude measures. *Comput Methods Programs Biomed*. 2016;130:1–12. doi:10.1016/j.cmpb.2016.03.021
27. Feldman DE, Shrier I, Rossignol M, Abenhaim L. Risk factors for the development of low back pain in adolescence. *Am J Epidemiol*. 2001;154(1):30–36.
28. Schneider S, Randoll D, Buchner M. Why do women have back pain more than men?: a representative prevalence study in the Federal Republic of Germany. *Clin J Pain*. 2006;22(8):738–747. doi:10.1097/01.ajp.0000210920.03289.93

## Journal of Pain Research

### Publish your work in this journal

The Journal of Pain Research is an international, peer reviewed, open access, online journal that welcomes laboratory and clinical findings in the fields of pain research and the prevention and management of pain. Original research, reviews, symposium reports, hypothesis formation and commentaries are all considered for publication. The manuscript

Submit your manuscript here: <https://www.dovepress.com/journal-of-pain-research-journal>

management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress