

# RELIABILITY OF THE MICHIGAN STATE UNIVERSITY (MSU) CLASSIFICATION OF LUMBAR DISC HERNIATION

## CONFIABILIDADE DA UNIVERSIDADE DO ESTADO DE MICHIGAN (MSU) CLASSIFICAÇÃO DA HERNIAÇÃO DE DISCO LOMBAR

ARUN-KUMAR KALIYA-PERUMAL<sup>1, 2</sup>, CHI-AN LUO<sup>1</sup>, YU-CHENG YEH<sup>1</sup>, YI-FANG TSAI<sup>1</sup>, MICHAEL JIAN-WEN CHEN<sup>1</sup>, TSUNG-TING TSAI<sup>1</sup>

1. Department of Orthopedic Surgery, Spine Division, Bone and Joint Research Center, Chang Gung Memorial Hospital and University College of Medicine, Taoyuan, Taiwan.  
2. Department of Orthopedic Surgery, Melmaruvathur Adhiparasakthi Institute of Medical Sciences and Research, Melmaruvathur, Tamil Nadu, India.

### ABSTRACT

**Objective:** The Michigan State University (MSU) classification of lumbar disc herniation (LDH) is periodically used by various authors to classify disc herniation. We assessed the reliability of this classification system among orthopedic residents at our institute. **Methods:** Fifty T2 axial-cut magnetic resonance images (MRI) corresponding to the level of maximal disc herniation from patients diagnosed with a single LDH were selected and distributed to six orthopedic residents. All six residents gave a specific rating for each image based on the MSU classification; in addition, three residents gave ratings on two different occasions. The degree of agreement among residents was analyzed by calculating inter-observer and intra-observer reliability using the Kappa statistic. **Results:** The inter-observer reliability among the six residents calculated as the Fleiss' Kappa was 0.422, which indicates moderate reliability. The intra-observer reliability of three selected residents calculated by Cohen's Kappa was 0.750, 0.772, and 0.859, which indicates substantial to almost perfect reliability. Variations in ratings were frequent in images portraying a broad-based disc herniation with spinal canal stenosis. **Conclusion:** Our findings demonstrate moderate homogeneity of ratings given by residents; however, test-retest results proved the ratings to be consistent. **Level of Evidence II, Diagnostic studies - investigating a diagnostic examination.**

**Keywords:** Inter-observer variability. Intervertebral disc. Intervertebral Disc Displacement. Reliability. Spondylosis.

### RESUMO

**Objetivo:** A classificação da hérnia de disco lombar (LDH) da Michigan State University (MSU) é usada periodicamente por vários autores para classificar as hérnias discais. Pretendemos avaliar a confiabilidade deste sistema de classificação entre os residentes de ortopedia em nosso instituto. **Métodos:** Cinquenta imagens de RM axial do corte T2 correspondendo ao nível de hérnia discal máxima de pacientes que foram diagnosticados com uma única LDH foram selecionadas e distribuídas para seis residentes ortopédicos. Todos os seis residentes deram uma classificação específica para cada imagem com base na classificação MSU; Além disso, três residentes deram notas em duas ocasiões diferentes. O grau de concordância entre os residentes foi analisado calculando-se a confiabilidade interobservador e intraobservador pela estatística Kappa. **Resultados:** Descobrimos que a confiabilidade interobservador entre seis residentes, calculando o Kappa de Fleiss, foi de 0,422; isso indica confiabilidade moderada. No entanto, a confiabilidade intra-observador de três residentes selecionados mostrou-se substancial (Kappa de Cohen = 0,750, 0,772 e 0,859 em três residentes, respectivamente). Variações na observação foram frequentes se houvesse hérnia discal ampla com estenose do canal vertebral. **Conclusão:** Nossos achados demonstram homogeneidade moderada das avaliações dadas pelos residentes; no entanto, teste-reteste provou que as classificações eram consistentes. **Nível de Evidência II, Estudos diagnósticos - investigação de um exame para diagnóstico.**

**Descritores:** Variações dependentes do observador. Disco Intervertebral. Deslocamento do Disco Intervertebral. Reprodutividade dos testes. Espondilose.

**Citation:** Kaliya-Perumal AK, Luo CA, Yeh YC, Tsai YF, Chen MJW, Tsai TT. Reliability of the michigan state university (MSU) classification of lumbar disc herniation. *Acta Ortop Bras.* [online]. 2018;26(6):411-4. Available from URL: <http://www.scielo.br/aob>.

### INTRODUCTION

Displacement of disc material beyond the limits of the intervertebral disc space is termed as a disc herniation.<sup>1</sup> Such lumbar disc herniations (LDHs) are supposedly classified according to the long-established anatomical classification system.<sup>1</sup> This system incorporates all varieties of herniations and classifies them into protrusion, extrusion and sequestration. The generality of this

classification makes it difficult to imagine or picturise the exact shape of the disc herniation by knowing just the type, without looking at the magnetic resonance (MR) image. This disadvantage can be overcome by using more precise systems as that of Wiltse et al or the Michigan State University (MSU) classification system.<sup>2,3</sup> We believe that the MSU classification is simple and clearly defines the shape, location and extent of the disc herniation particularly in

All authors declare no potential conflict of interest related to this article.

Study conducted at the Department of Orthopedic Surgery, Spine Division, Bone and Joint Research Center, Chang Gung Memorial Hospital and University College of Medicine, Taoyuan, Taiwan.

Correspondence: Tsai TT. Orthopedic Spine Division, Chang Gung Memorial Hospital at Linkou, Taoyuan 33305, Taiwan. E-mail: [tsai1129@gmail.com](mailto:tsai1129@gmail.com)

Article received in 05/30/2018, approved in 10/01/2018.



the lumbar spine. It only requires a single T2 axial cut MRI image that corresponds to the level of maximal herniation, considering the upward or downward migration in case of a sequestered disc.<sup>2</sup> Based on this classification, the size of the disc herniation is described as Grade 1, 2 and 3; the location of the disc herniation is described as Zone 1, 2 and 3 (Figure 1). On combining the size of the disc herniation with its location, ten distinct types can be obtained. Our residents were comfortable with this classification; hence, we decided to quantify the reliability of this objective system, among orthopaedic residents at our institute.

## MATERIALS AND METHODS

Retrospectively, we selected 50 T2 axial cut MR images at the level of maximal herniation that belonged to patients who were diagnosed with a single LDH that required intervention. This selection included patients with varied severities of disc induced lumbar radiculopathy, who underwent conservative management, selective nerve root block and/or mini open discectomy as a definite management. Our selection included patients with degenerative spondylosis or ligamentum flavum thickening at the chosen level; however, none of the patients had a concomitant inflammation, infection or neoplasia affecting the disc level.

A single appropriate T2 axial cut MR image corresponding to the level of maximal herniation in each patient was chosen by a single experienced surgeon. These images were given to six orthopaedic residents for categorising the disc herniation based on the MSU classification system. The residents were previously aware of this classification system; however, they did not use it as a routine. They were initially briefed about the system in a calibrating teaching session using the original work published by Mysliwiec LW et al.<sup>2</sup> All queries were addressed, following which a copy of the original work and the 50 selected MR images were provided to the residents. Residents were advised to take adequate time to analyse each image before giving a response. They were not put under an obligation to time, as to when their responses need to be submitted. This was not a part of their routine work, but to be carried out at their will, during their free time without any stress. By this way, instances of fatigue affecting the judgement of the residents were avoided; also, the precision and consistency of their judgment were maintained. All residents returned their response with a classification for each MR image within a week.

As the classification system includes 10 types, each type was given a number from 1 to 10. Therefore, six sets of nominal variables were

obtained from the responses received from the residents. This data was used to determine the inter-observer reliability by calculating the Fleiss' Kappa (statistical measure for assessing the reliability of agreement between multiple raters). The same MRI images were shuffled and provided to three of the residents (Resident 1, 2 and 6) for reassessment after a month. Their response was collected and was compared with their previous ratings. This data was used to determine the intra-observer (test-retest) reliability by calculating the Cohen's Kappa (statistical measure for assessing the reliability of agreement between two raters) for each resident. The inferred results were tabulated. Statistical analyses were done using IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp. Implied consent was obtained from the study participants when they agreed to participate in this research. This study was approved by the institutional review board of Chang Gung Memorial Hospital (IRB No – 201700227B0) and was performed in compliance with the 1964 declaration of Helsinki, its later amendments or comparable ethical standards.

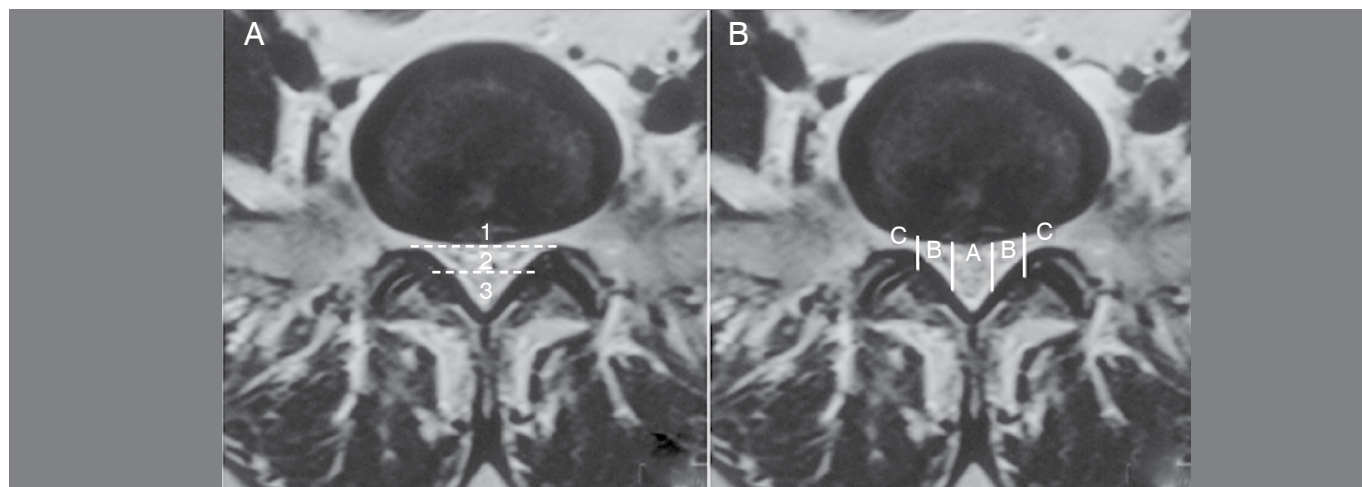
## RESULTS

The selected MRI images (n = 50) included all types of LDHs described in the MSU classification taken from 50 different patients (Age = 46.9 ± 8.7; Male = 31; Female = 29). The residents who rated the MRI images were in their third year of residency training and they took approximately a week to classify all 50 axial cut MRI images. Data received from the residents were in the form of classification types described in the MSU system. The most appropriate classification type for an MRI image was considered as the one which majority of the residents had an agreement. Accordingly, the total number of images belonging to each classification type was tabulated; this represents the range and severity of disc herniations among the selected images (Table 1). The classifications provided by the Residents were later numerically rated from 1 to 10 for computation purposes.

There was agreement among three or more raters for 48 (96%) of the selected MRI images which reduced to 37 (74%) when calculated

**Table 1.** Cases in each classification type and their agreement percentages.

Type	1A	1B	1C	2A	2B	2AB	2C	3A	3B	3AB
No. of Cases	5	5	1	6	10	8	3	5	5	2
Percentage of agreement	70.02	73.34	100	80.57	65	60.44	55.57	70	63.34	75



**Figure 1.** Grading and Zoning as per the MSU classification system. A) Lines representing grading of disc prolapse are drawn in the horizontal axis. B) Lines representing zoning of disc prolapse are drawn in the vertical axis.

for agreement among four or more raters. Only six (12%) of the MRI images had 100% agreement among raters; these images were of types 1A, 1B, 1C, 3B and two of 2A. However, this data does not depict the reliability of the classification system.

The agreement percentage for each MRI image was calculated, based on which the mean agreement percentage for each classification type was calculated to check if there was a relation between herniation severity and resident agreement (Table 1). We noticed the types 1A, 1B, 1C, 2A, 3A and 3AB to have a mean agreement percentage of 70 or above. However, 2B, 2AB, 2C and 3B had mean agreement percentages ranging between 55 and 65, with 2C having the least mean agreement percentage of 55.57. These relatively low mean agreement percentages among Residents could be due to the herniations being broad based in an already stenosed canal (Figure 2).

The tabulated ratings of all residents were used to calculate the pair wise Cohen's Kappa and a matrix was generated (Table 2). The inter-rater or inter-observer reliability was determined by calculating the Fleiss' Kappa which was found to be 0.422 (Table 3). According to Cohen, our measure of Kappa falls under moderate agreement (0.41-0.60).<sup>4</sup> This can be accepted considering that reliability is expected to be low when multiple data collectors are required to make finer discriminations as in MSU classification; however, a measure above 0.60 could have been adequate.<sup>4</sup>

After a month's interval, the MRI images were shuffled and provided to three Residents for reassessment, independent of their previous measures. Their previous and latest rating for each MRI image was tabulated. We found that 39 (78 %), 40 (80 %) and 44 (88%) of the recent ratings by Resident 1, 2 and 6 respectively, were consistent with their previous measures. This data was used to determine the Intra-rater or intra-observer (test-retest) reliability by calculating the Cohen's Kappa for each resident. A Kappa measure of 0.750 (Substantial agreement), 0.772 (Substantial agreement) and 0.859 (almost perfect agreement) was obtained for resident 1, 2 and 6 respectively. Hence, the intra-observer reliability can be interpreted as substantial to almost perfect.

## DISCUSSION

Classifying lumbar disc herniation can provide vital assistance for clinical management of the condition. MRI is considered the ideal tool for analysis of such lumbar disc herniations.<sup>5-9</sup> Both sagittal and axial cut images can provide valuable information of the underlying pathology. However, axial cut image at the pathological level is given sole priority by the MSU classification system which is periodically

**Table 2.** Pair wise Kappa matrix.

	Resident 1	Resident 2	Resident 3	Resident 4	Resident 5	Resident 6
Resident 1	1.000	0.461	0.365	0.357	0.502	0.656
Resident 2	0.461	1.000	0.394	0.330	0.546	0.508
Resident 3	0.365	0.394	1.000	0.292	0.363	0.451
Resident 4	0.357	0.330	0.292	1.000	0.235	0.458
Resident 5	0.502	0.546	0.363	0.235	1.000	0.456
Resident 6	0.656	0.508	0.451	0.458	0.456	1.000

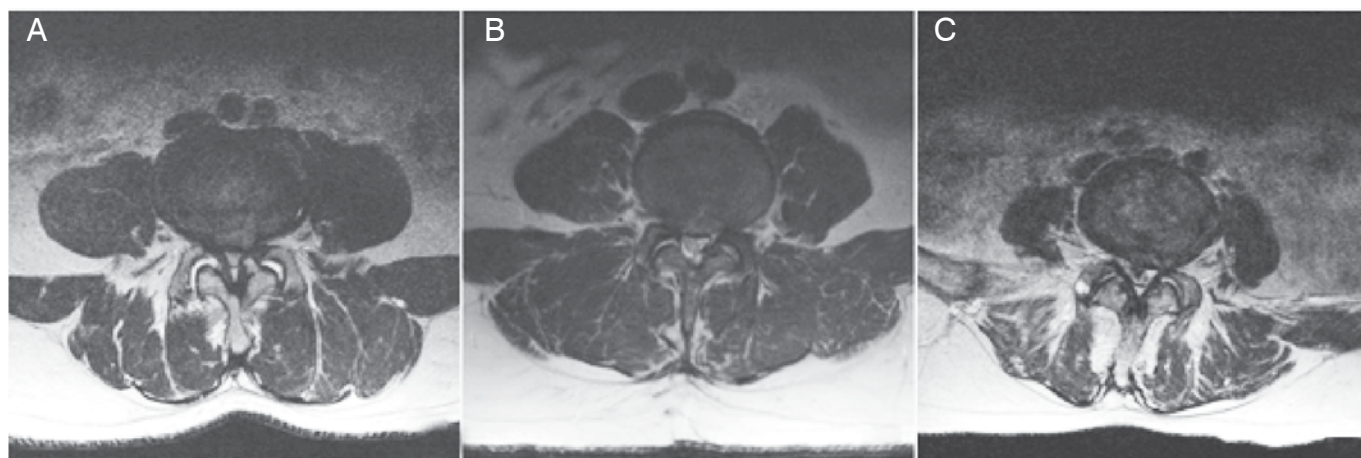
**Table 3.** Inter-observer and intra-observer reliability assessment using Kappa Statistic.

Reliability	Statistical Measure	Kappa	Interpretation
Inter-observer	Fleiss' Kappa	0.422	Moderate
Intra-observer (Resident 1)	Cohen's Kappa	0.750	Substantial
Intra-observer (Resident 2)		0.772	Substantial
Intra-observer (Resident 6)		0.859	Almost perfect

practiced by authors around the world to optimize management strategies for patients with lumbar disc herniations.<sup>2,10-12</sup> The concept of considering a single axial cut image at the level of maximal herniation may be unlike the "Lumbar disc nomenclature: version 2.0" where sagittal images are taken into consideration;<sup>13</sup> even so, the MSU classification clearly defines the shape, location and extent of the disc herniation.

Even though, our residents were comfortable with this classification system, a calibrating session was held to refine their understanding of this system. In this session, we discuss the MRI of several patients and ask residents to classify them according to the MSU classification system. They were then asked to justify why they chose a particular type for each discussed MRI. If there were disagreements, rules were framed to give the most appropriate rating for a specific MRI. By this way, we believe that the understanding of the classification system was refined.

We used the Kappa statistic to determine the reliability of the MSU classification system.<sup>4</sup> This is because the ratings given by the six residents were considered as nominal variables. Hence we determined the inter-observer reliability by calculating the Fleiss' Kappa for more than two raters which is an extension of Cohen's Kappa that is used for calculating agreement among two raters. Besides that, the intra-observer (test-retest) reliability was determined using the measure of Cohen's Kappa as it involves one previous and one recent rating by each resident. Our results were finalized based on



**Figure 2.** Examples of deceptive MRI that had least agreement among residents. A) Frequently rated as type 2a or 2ab. B) Frequently rated as type 2b or 2ab. C) Frequently rated as type 2ab or 2c



accepted interpretations of the Kappa statistic.<sup>4,14</sup> We inferred a moderate inter-observer reliability and substantial to almost perfect intra-observer reliability.

The reason for obtaining a moderate inter-observer reliability needs to be discussed. Firstly, this could be because we chose multiple raters who had to rate multiple categories.<sup>14</sup> It is an accepted fact that when multiple data collectors are required to make finer discriminations, reliability is difficult to be obtained.<sup>4</sup> Besides that, this could be because the types 2B, 2AB, 2C and 3B had mean agreement percentages ranging between 55 and 65 only. On analysing the reason for less agreement among Residents for these types, we found that this was mainly due to the herniations being broad based in an already stenosed canal due to degeneration. Apart from these factors, the learning curve to get familiarized to this classification system may also be considered; however, if appropriate rules are framed to interpret such deceptive MRI, better inter-observer reliability can be achieved and findings can be correlated with clinical presentations to guide management. It should be noted that this classification system does not take into account a bulging disc, either symmetrical or asymmetrical as described in the Lumbar Disc Nomenclature 2.0;<sup>1,13,15</sup> however, the system holds good for herniated discs. Clinical presentation of patients does not depend on the anatomy of the disc prolapse alone but rather depends on many other factors that can cause symptoms.<sup>13,16</sup> These include disc degeneration, reactive vertebral body marrow changes, ligamentum flavum hypertrophy, facet hypertrophy or associated segmental instabilities.<sup>17,18</sup> In spite of a significant MRI finding of a disc herniation, asymptomatic clinical presentations are also a possibility.<sup>16,19-20</sup>

Hence, even though MSU classification can describe the exact anatomic appearance of a herniated disc, management protocols cannot be formulated with its sole guidance. Other concomitant parameters should be given equal importance along with MSU classification type to optimize management protocols; yet, it is vital to know the anatomic appearance of the disc by an objective system like MSU classification to plan the approach and procedure if intervention is considered.

### Limitation

Our results and interpretation could be influenced by each resident's understanding and experience with this classification system. This could have biased our results of reliability.

### CONCLUSION

The inter-observer and intra-observer reliability of the MSU classification for lumbar disc herniations was calculated among orthopaedic residents. Our findings demonstrate moderate homogeneity of the ratings given by the residents; however, test-retest proved the ratings to be consistent. This observation implies that the MSU classification could be of clinical importance; however, appropriate rules need to be framed to interpret deceptive MRI which is highly essential to delineate optimal management protocols.

### ACKNOWLEDGEMENTS

The authors sincerely thank Dr. Winson Min-Teng Low, Department of Orthopaedic Surgery, Chang Gung Memorial Hospital at Linkou, Taiwan for his contributions to this research.

**AUTHORS' CONTRIBUTIONS:** Each individual author contributed significantly to the development of this work. AKKP (0000-0002-2747-4500)\*: concept, design, data acquisition, statistical analysis, interpretation of data, drafting the manuscript and approval of final version; CAL (0000-0003-1791-0820)\*: concept, design, statistical analysis, interpretation of data, approval of final version; YCY (0000-0002-3393-3996)\*: concept, design, review and approval of final version; YFT (0000-0001-6726-2572)\*: interpretation of data, review and approval of final version; MJWC (0000-0001-7445-904X)\*: interpretation of data, review and approval of final version; TTT (0000-0002-7776-7627)\*: concept, design, interpretation of data, drafting the manuscript, review and approval of final version. \*ORCID (Open Researcher and Contributor ID).

### REFERENCES

1. Fardon DF. Nomenclature and classification of lumbar disc pathology. *Spine (Phila Pa 1976)*. 2001;26(5):461-2.
2. Mysliwiec LW, Cholewicki J, Winkelpfeck MD, Eis GP. MSU classification for herniated lumbar discs on MRI: toward developing objective criteria for surgical selection. *Eur Spine J*. 2010;19(7):1087-93.
3. Wiltse LL, Berger PE, McCulloch JA. A system for reporting the size and location of lesions in the spine. *Spine (Phila Pa 1976)*. 1997;22(13):1534-7.
4. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-82.
5. Kim KY, Kim YT, Lee CS, Shin MJ. MRI classification of lumbar herniated intervertebral disc. *Orthopedics*. 1992;15(4):493-7.
6. Carlisle E, Luna M, Tsou PM, Wang JC. Percent spinal canal compromise on MRI utilized for predicting the need for surgical treatment in single-level lumbar intervertebral disc herniation. *Spine J*. 2005;5(6):608-14.
7. Kim KY, Kim YT, Lee CS, Kang JS, Kim YJ. Magnetic resonance imaging in the evaluation of the lumbar herniated intervertebral disc. *Int Orthop*. 1993;17(4):241-4.
8. Hussaini S, Karimi N, Ezzati K, Hossein Zadeh S, Rahnama L, Arslan S. Reliability of Magnetic Resonance Imaging Findings Interpretation in Patients with Lumbar Disk Herniation. *Physical Treatments: Specific Physical Therapy*. 2015;5(2).
9. Li Y, Fredrickson V, Resnick DK. How should we grade lumbar disc herniation and nerve root compression? A systematic review. *Clin Orthop Relat Res*. 2015;473(6):1896-902.
10. Moon SH, Lee JI, Cho HS, Shin JW, Koh WU. Factors for Predicting Favorable Outcome of Percutaneous Epidural Adhesiolysis for Lumbar Disc Herniation. *Pain Res Manag*. 2017;2017:1494538.
11. Al-Khawaja DO, Mahasneh T, Li JC. Surgical treatment of far lateral lumbar disc herniation: a safe and simple approach. *J Spine Surg*. 2016;2(1):21-4.
12. Arun-Kumar K, Jayaprasad S, Senthil K, Lohith H, Jayaprakash KV. The Outcomes of Selective Nerve Root Block for Disc Induced Lumbar Radiculopathy. *Malays Orthop J*. 2015;9(3):17-22.
13. Fardon DF, Williams AL, Dohring EJ, Murtagh FR, Gabriel Rothman SL, Sze GK. Lumbar disc nomenclature: version 2.0: Recommendations of the combined task forces of the North American Spine Society, the American Society of Spine Radiology and the American Society of Neuroradiology. *Spine J*. 2014;14(11):2525-45.
14. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-68.
15. Williams AL, Murtagh FR, Rothman SL, Sze GK. Lumbar disc nomenclature: version 2.0. *AJNR Am J Neuroradiol*. 2014;35(11):2029.
16. Janardhana AP, Rajagopal, Rao S, Kamath A. Correlation between clinical features and magnetic resonance imaging findings in lumbar disc prolapse. *Indian J Orthop*. 2010;44(3):263-9.
17. Stollman A, Pinto R, Benjamin V, Kricheff I. Radiologic imaging of symptomatic ligamentum flavum thickening with and without ossification. *AJNR Am J Neuroradiol*. 1987;8(6):991-4.
18. Belthur M, Thonse R. An uncommon cause of lumbar radiculopathy. *Postgrad Med J*. 2002;78(917):182, 6.
19. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic Resonance Imaging of the Lumbar Spine in People without Back Pain. *N Engl J Med*. 1994;331(2):69-73.
20. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, Deyo RA, et al. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *AJNR Am J Neuroradiol*. 2015;36(4):811-6.