Artificial Intelligence Comparison of the Radiologist Report With Endoscopic Predictors of Successful Transforaminal Decompression for Painful Conditions of the Lumber Spine: Application of Deep Learning Algorithm Interpretation of Routine Lumbar Magnetic Resonance Imaging Scan

Kai-Uwe Lewandrowski, Narendran Muraleedharan, Steven Allen Eddy, Vikram Sobti, Brian D. Reece, Jorge Felipe Ramírez León and Sandeep Shah

This information is current as of May 17, 2025.

Email Alerts   Receive free email-alerts when new articles cite this article. Sign up at:
http://ijssurgery.com/alerts

Artificial Intelligence Comparison of the Radiologist Report With Endoscopic Predictors of Successful Transforaminal Decompression for Painful Conditions of the Lumber Spine: Application of Deep Learning Algorithm Interpretation of Routine Lumbar Magnetic Resonance Imaging Scan

KAI-UWE LEWANDROWSKI, MD,[1] NARENDRAN MURALEEDHARAN, BASME,[2]
STEVEN ALLEN EDDY, MD,[3] VIKRAM SOBTI, MD, MBA,[4] BRIAN D. REECE, MD,[5]
JORGE FELIPE RAMÍREZ LEÓN, MD,[6] SANDEEP SHAH, MSEE, MBA[7]

[1]*Staff Orthopaedic Spine Surgeon Center for Advanced Spine Care of Southern Arizona and Surgical Institute of Tucson, Tucson, Arizona,* [2]*Aptus Engineering, Inc, Scottsdale, Arizona, and Multus Medical, LLC, Phoenix, Arizona,* [3]*Multus Medical, LLC, Phoenix, Arizona,* [4]*Innovative Radiology, PC, River Forest, Illinois,* [5]*The Spine and Orthopedic Academic Research Institute, Lewisville, Texas,* [6]*Fundación Universitaria Sanitas, Bogotá, Colombia, Research Team, Centro de Columna. Bogotá, Colombia, Centro de Cirugía de Mínima Invasión, CECIMIN—Clínica Reina Sofía, Bogotá, Colombia,* [7]*Multus Medical, LLC, Phoenix, Arizona*

ABSTRACT

Background: Identifying pain generators in multilevel lumbar degenerative disc disease is not trivial but is crucial for lasting symptom relief with the targeted endoscopic spinal decompression surgery. Artificial intelligence (AI) applications of deep learning neural networks to the analysis of routine lumbar MRI scans could help the primary care and endoscopic specialist physician to compare the radiologist's report with a review of endoscopic clinical outcomes.

Objective: To analyze and compare the probability of predicting successful outcome with lumbar spinal endoscopy by using the radiologist's MRI grading and interpretation of the radiologic image with a novel AI deep learning neural network (Multus Radbot™) as independent prognosticators.

Methods: The location and severity of foraminal stenosis were analyzed using comparative ordinal grading by the radiologist, and a contiguous grading by the AI network in patients suffering from lateral recess and foraminal stenosis due to lumbar herniated disc. The compressive pathology definitions were extracted from the radiologist lumbar MRI reports from 65 patients with a total of 383 levels for the central canal – (0) no disc bulge/protrusion/canal stenosis, (1) disc bulge without canal stenosis, (2) disc bulge resulting in canal stenosis, and (3) disc herniation/protrusion/extrusion resulting in canal stenosis. Both neural foramina were assessed with either – (0) neural foraminal stenosis absent, or (1) neural foramina are stenosis present. Reporting criteria for the pathologies at each disc level and, when available, the grading of severity were extracted and assigned into two categories: "Normal," and "Stenosis." Clinical outcomes were graded using dichotomized modified Macnab criteria considering *Excellent* and *Good* results as "Improved," and *Fair* and *Poor* outcomes as "Not Improved." Binary logistic regression analysis was used to predict the probability of the AI- and radiologist grading of stenosis at the 88 foraminal decompression sites to result in "Improved" outcomes.

Results: The average age of the 65 patients was 62.7 +/- 12.7 years. They consisted of 51 (54.3%) males and 43 (45.7%) females. At an average final follow-up of 57.4 +/- 12.57, Macnab outcome analysis showed that 86.4% of the 88 foraminal decompressions resulted in *Excellent* and *Good* (Improved) clinical outcomes. The stenosis grading by the radiologist showed an average severity score of 4.71 +/- 2.626, and the average AI severity grading was 5.65 +/- 3.73. Logit regression probability analysis of the two independent prognosticators showed that both the grading by the radiologist (86.2%; odds ratio 1.264) and the AI grading (86.4%; odds ratio 1.267) were nearly equally predictive of a successful outcome with the endoscopic decompression.

Conclusions: Deep learning algorithms are capable of identifying lumbar foraminal compression due to herniated disc. The treatment outcome was correlated to the decompression of the directly visualized corresponding pathology during the lumbar endoscopy. This research should be extended to other validated pain generators in the lumbar spine.

Level of Evidence: 3.

Clinical Relevance: Validity, clinical teaching, evaluation study.

INTRODUCTION

One might ask what is the significance of artificial intelligence (AI) analysis of routine lumbar magnetic resonance imaging (MRI) scan reads of patients who contemplate endoscopic decompression for sciatica-type back and leg pain due to herniated disc and stenosis. The answer lies in the need for better or additional prognosticators in the preoperative diagnostic process to direct this minimal surgical decompression procedure at the pain generators that are causing the patient's symptoms.[1] The specific cause for spine-related disability can be quite diverse but is—in the currently established framework of medical necessity criteria for spine surgery—restricted to overt mechanical neural element compression or instability. Image-based standards of grossly abnormal findings, such as grade I or higher spondylolisthesis or severe spinal stenosis, are well-accepted triggers for traditional open spine surgery and meet Medicare coverage criteria and those for most managed care plans.[2] These criteria, however, leave a significant percentage of patients with sciatica-type low back and leg pain without treatment, as their MRI scans are erroneously graded as "normal" (false negative) or underestimate the pain generators in a multilevel degenerative segment involvement. A recent study has estimated that this diagnostic gap is between 18% and 30%.[3] These patients in pain are going to continue to look for treatments, some of which may translate into continued and repetitive use of ineffective medical and physical therapies or, simply put, waste. It may also be impractical to address each possible pain generator in 1 surgical session, so the clinician will, for practicability, choose the 1 or 2 most significant pain generators that correlate with the clinical exam.

Endoscopic visualization of previously unrecognized painful spinal pathology could reduce spine-related disability if appropriate treatment was instituted even if the routine lumbar MRI scan of the painful area suggested otherwise and was read as normal.[4–7] One example of this problem is the frequently overlooked extraforaminal disc herniations of various sizes[8–13] that may impinge on the dorsal root ganglion of the exiting nerve root, chronically inflame it, and cause severe sciatica that seems out of proportion with the associated findings on the axial MRI scan through the suspected symptomatic level.[8] However, once directly visualized with the endoscope and successfully treated, a surgeon might never forget this commonly missed entity and always include it in the differential diagnosis of unexplained spine pain. Many other such directly visualized pain generators that escape the routine lumbar MRI scan have been validated and successfully treated with the spinal endoscopy procedure, corroborating the need for more accurate diagnostic tools during the preoperative work-up.[1]

In this clinical outcome study on patients suffering from sciatica due to herniated disc, the authors present the results of a binary logistic probability analysis of the AI deep learning networks being able to predict successful outcomes with the targeted endoscopic decompression surgery as is currently implied in the radiologist's description of the compressive pathology in routine MRI reporting. Ultimately, the author's goal was to aid in the development of more useful diagnostic tools to work up low back pain patients to provide more targeted and effective treatments. This retrospective study is a stepping-stone toward that goal. The consideration of AI is a check on the meaning or accuracy of the radiologist's report to the clinician who orders the imaging scan. Radiologists are also known to emphasize or deemphasize certain aspects of imaging as significant or insignificant, depending on whether the report is for a primary care physician or for a spine specialist who ordered the imaging study and report.

MATERIALS AND METHODS

In this study, the authors focused on the application of a deep learning neural network models on herniated disc affecting the lateral recess and the neuroforamina. The feasibility of this AI approach to reliably generate MRI reports comparable to those provided by a radiologist was demonstrated in the prior literature.[14,15] We are now reporting on the probability of the deep learning neural network with computer AI and software engineers to predict clinical improvements with the endoscopic decompression procedure on the basis of AI segmentation models directly targeting compressive pathology in the lateral recess and the neuroforamen.

Patients and Inclusion and Exclusion Criteria

This study included 65 patients who underwent endoscopic transforaminal decompression for her-

niated disc. The average age of the 65 patients was 62.54 years, with a standard deviation of 12.7 years ranging from 29 to 93 years. There were 51 (54.3%) male and 43 (45.7%) female patients. The average follow-up was 57.4 months with a standard deviation of 12.57 months ranging from 42 to 86 months. Patients with symptoms that have proven refractory to nonoperative treatment were considered for this procedure using the following inclusion criteria:[16–18]

(1) lumbar radiculopathy including pain, sensory changes, or weakness;
(2) imaging evidence of foraminal or lateral recess stenosis demonstrated on preoperative MRI and computed tomography (CT) scans
(3) unsuccessful nonoperative treatment including physical therapy and transforaminal epidural steroid injections for at least 12 weeks

Patients considered not suitable for the transforaminal endoscopic decompression were stratified according to the following exclusion criteria:

(1) segmental instability with greater than grade I spondylolisthesis or translational motion of greater than 5 mm on preoperative extension flexion radiographs, severe central stenosis (less than 100 mm^2),[19] or both
(2) extensive facet arthropathy
(3) infection
(4) metastatic disease

Endoscopic Surgical Technique

In this clinical outcome study on patients suffering from sciatica due to herniated disc, the authors present the results of a binary logistic probability analysis of patients who underwent endoscopic transforaminal decompression employing the "outside- in" technique.[20] An initial foraminoplasty was performed with power drills, trephines, chisels, and rongeurs after serial dilation and placement of the working cannula in the neuroforamen following published techniques.[21–25] The endoscopic decompression procedure was directly or indirectly visualized throughout the surgery. Compressive pathology contributing to inflammation or tethering of the nerve roots was recorded for correlation to the preoperative MRI scan, which, as described below, was graded by the radiologist and the AI deep learning network.
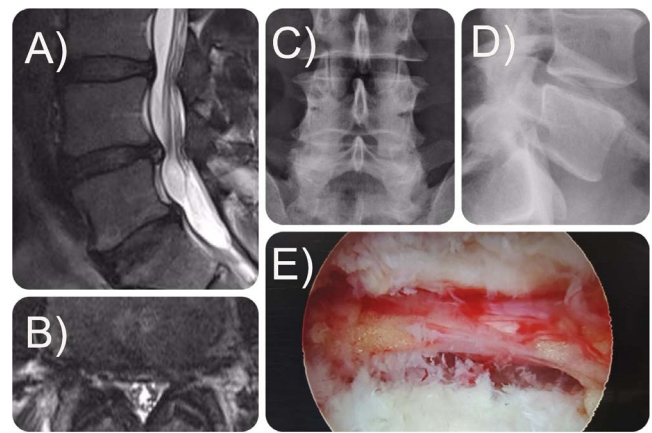

Figure 1. Preoperative sagittal (A), axial (B) magnetic resonance imaging scan, posterior-anterior (C), and lateral (D) radiograph of a 48-year-old male. The patient was treated with transforaminal outside-in endoscopic decompression with foraminoplasty and discectomy (E) for failed conservative care of an L4–L5 herniated disc.

Fluoroscopic surveillance images were occasionally taken for orientation and verification of the decompression. An illustrative case of a 48-year-old male who was treated with transforaminal outside-in endoscopic decompression with foraminoplasty and discectomy for failed conservative care of an L4–L5 herniated disc is shown in Figure 1.

Radiographic and Diagnostic Criteria

The size and location of the compressive pathology, whether from disc herniations or other types of soft tissue or bony stenosis in the spinal canal, lateral recess, and neuroforamen, were classified according to well-established radiographic classification systems.[26–29] MRI criteria used in the clinical stratification of symptomatic patients were 15 mm or less for the height of the neuroforamen, 3 mm or less measured as posterior intervertebral disc height, or the width of the neuroforamen.[30] Diagnostic and therapeutic selective nerve root blocks and foraminal epidurography[31] with favorable therapeutic response (or selective nerve root blocks) were used to confirm the pain level.[32–37] The type of disc herniation was classified as central, paracentral, or combined central and paracentral.[29] Moreover, they were graded as contained or extruded.

Clinical Outcome Measures

Outcome assessment with the endoscopic decompression procedure was done employing the modified Macnab criteria: *excellent*—little pain and return to desired activities with few limitations; *good*—occasional pain or dysesthesias with daily

activities with minor restrictions, without needing pain medication; *fair*—improved but needing pain medication postoperatively; or *poor*—worse function prompting additional surgery.[38] All patients were instructed to be seen at a minimum in follow-up for examination and management of any problems at 2 and 6 weeks and then at 3, 6, 12, and 24 months postoperatively.

Grading of MRI Data

The deep learning neural network models analyzed 65 lumbar MRI scans from the same number of patients comprising a total of 383 levels. The MRI DICOM data sets were obtained from the MRI imaging center where the patient had the study. The MRI imaging centers also provided radiology reports prepared and approved by board-certified radiologists. The MRI scans and reports were screened and graded by an independent radiologist for the presence or absence of annular bulging[39] (circumferential, paracentral, posterior), disc herniation[40] (extrusion, protrusion, sequestration, fragmentation), central canal stenosis[41–43] (compromise of the thecal sac with presence or absence of ventral epidural fat), and foraminal stenosis[44] (compromise of the left, right, or both neural foramina and nerve roots). For this analysis, only parameters derived from the surgically treated level were included in the statistical computations. The independent radiologist graded the severity of the foraminal and lateral recess stenosis on an ordinal noncontiguous scale from 1 (no stenosis) to 10 (severe stenosis). Also, the location of the foraminal stenotic process was recorded from medial to lateral into the entry, mid-, and exit zone employing validated radiographic classification systems.[40]

Extraction of MRI Data and AI Detectors

The segmentation models employed by the deep learning algorithms involve extracting the location of the disc herniation, and its dimensions were extracted from the radiologist report using the following classes for the central canal: (0) no disc bulge/protrusion/canal stenosis, (1) disc bulge without canal stenosis, (2) disc bulge resulting in canal stenosis, and (3) disc herniation/protrusion/extrusion resulting in canal stenosis. One of the following classes were also extracted for each of the left and right neural foramina: (0) neural foraminal stenosis absent or (1) neural foramina is stenosis present.

The algorithm generated a contiguous severity score of foraminal and canal stenosis by employing several pathology detectors. The first pathology detector assesses the deformity of the posterior annulus to determine whether any posterior disc deformities due to bulging exist. If the deformity value is >50%, the herniation and stenosis detectors are triggered. The herniation detector is trained to identify posterior, central, and paracentral disc herniations and to classify them as protrusions, extrusions, or contained circumferential bulges. In comparison, the canal stenosis detector is trained to identify whether the disc deformity causes stenosis in the central canal. Each of the 3 detectors has a remapped contiguous confidence level of the specific AI detector from 0 to 10 representing the level of confidence that the AI segmentation models have that a particular pathology is, in fact, present in the patient's MRI scan. Hence, it is not equivalent to the linear ordinal severity scale used by the radiologist. In contrast, the employed AI detectors auto-tune the confidence level threshold to 50% by referring to the prior training data set to minimize binary cross-entropy loss to render a prediction as to whether a compressive pathology, such as a disc herniation, exists (>50% confidence level) or does not exist (≤50%). This auto-tuning of the AI detectors uses a combination of sigmoids for class activations, softmax for final layer activations, and rectified linear unit (ReLU) functions for the image kernel layers, all of which are nonlinear detector functions.[38] The deep learning algorithm uses these sets of nonlinear activation functions to learn and predict various outcomes, in this case either the sigmoid or the ReLU function. This results in the confidence level output from the algorithm for each class having a very nonlinear relationship to severity of the pathology.

Statistical Analysis

For the clinical outcome analysis, descriptive statistics (mean and standard deviation), cross-tabulation statistics of sensitivity, specificity, and measures of association were computed for 2-way tables using IBM SPSS Statistics software (version 26.0). The Pearson χ^2 and the likelihood-ratio χ^2 tests were used as statistical measures of association. The authors employed binary logistic regression to model the probability of the MRI severity grading of the compressive pathology provided by either the radiologist or the AI to predict the binary dependent

Table 1. Level distribution of spinal disc spaces treat with endoscopic discectomy.

Disc Level	Level Distributions			Cumulative Percent
	Frequency	Percent	Valid Percent	
L3–L4	9	10.2	10.2	10.2
L4–L5	59	67.0	67.0	77.3
L5–S1	20	22.7	22.7	100.0
Total	88	100.0	100.0	

Table 2. Laterality frequency distribution of endoscopic decompression procedures.

Surgical Side	Laterality of Endoscopic Decompression			Cumulative Percent
	Frequency	Percent	Valid Percent	
Bilateral, left	17	19.3	19.3	19.3
Bilateral, right	17	19.3	19.3	38.6
Left	33	37.5	37.5	76.1
Right	21	23.9	23.9	100.0
Total	88	100.0	100.0	

Macnab outcome variable: clinical improvement (excellent and good Macnab outcomes) and no improvement (fair and poor Macnab outcomes). This logistic regression was used to estimate the parameters of a binary logistic model in which the categorical dependent variable (Macnab outcome) has 2 possible values: improved or not improved. The logarithm of the odds (log odds) for the value labeled 1 is a linear combination of 1 of the 2 independent predictors variables—the stenosis severity score produced by the radiologist (an ordinal variable of increasing severity from 1 to 10) and by the AI (a continuous variable of increasing severity from 1 to 10). The basic premise of this logit model is that the odds of a successful clinical outcome with the endoscopic decompression increases by a multiple (odds ratio) of a constant rate at which it ordinarily occurs by increasing 1 of the 2 independent stenosis severity variables employed in this study. This analysis relies on the hypothesis that decompression of compressive pathology that is more accurately graded as to its severity results in more reliable symptom relief and hence improved clinical outcomes. The log odds are converted to a probability by the logistic model allowing the authors to compare the predictive value of the stenosis grading provided by either the AI or the radiologist. This type of analysis was most appropriate since each independent predictor variable could have its own parameter for the binary dependent variable (Macnab outcome), allowing one to generalize the odds ratio. The confidence intervals for the likelihood ratios were calculated using the log method.

RESULTS

The demographic and level frequency distribution observed in the 65 patients is summarized in Tables 1 and 2. Seventeen of the 65 patients had a bilateral decompression, which accounts for 88 foraminal decompression sites. As expected, L4–L5 was the most commonly operated level (59/88; 67%). Most

patients had contained disc herniations (85/88; 96.6%), and only 3 patients had an extruded disc herniation (3.4%; Table 3), most of which were centrally located (55.7%). The remaining herniations were nearly equally distributed between paracentral (21.6%) and combined central and paracentral herniations (22.7%). Measuring the width of the herniations across its base on axial MRI sections through the midsection of the disc showed that the majority of them were larger than 10 mm in width (88.6%). The posterior disc height was preserved to more than 3 mm in the majority of surgical disc levels as well (95.5%). Most patients had a central canal area larger than 100 mm^2, indicating the absence of severe central stenosis. Cross tabulation of the foraminal zone classification and the location of the herniated disc in the spinal canal revealed that approximately half of lumbar disc herniations were causing foraminal stenosis in

Table 3. Frequency distribution of diagnosis, herniated nucleus pulposus type, size, and location and stenosis parameters in patients who underwent transforaminal endoscopic decompression.

Item	Frequency	Percent	Valid Percent	Cumulative Percent
Type of herniated disc				
Contained	85	96.6	96.6	96.6
Extruded	3	3.4	3.4	100.0
Total	88	100.0	100.0	
Location of herniated disc				
Central	49	55.7	55.7	55.7
Combined	20	22.7	22.7	78.4
Paracentral	19	21.6	21.6	100.0
Total	88	100.0	100.0	
Size of herniated disc[a]				
<10 mm	10	11.4	11.4	11.4
>10 mm	78	88.6	88.6	100.0
Total	88	100.0	100.0	
Posterior disc height				
<3 mm	4	4.5	4.5	4.5
>3 mm	84	95.5	95.5	100.0
Total	88	100.0	100.0	
Central canal stenosis[b]				
<100 mm	10	11.4	11.4	11.4
>100 mm	78	88.6	88.6	100.0
Total	88	100.0	100.0	

[a]The size of the herniated disc is measured on axial magnetic resonance images defined as the size of its widest distance across the bases.
[b]Severe central stenosis was defined according to Sengupta et al[19] as <100 mm^2 cross-sectional area.[19]

Table 4. Cross tabulation of location of herniated disc and the foraminal zones with stenosis.

Foraminal Zone With Stenosis	Location of Herniated Disc			
	Central Herniation	Paracentral Herniation	Central, Paracentral Herniation	Total
Exit zone				
Count	0	2	0	2
% within herniated disc	0.0	10.0	0.0	2.3
Mid-zone				
Count	9	0	2	11
% within herniated disc	18.4	0.0	10.5	12.5
Exit zone				
Count	3	2	1	6
% within herniated disc	6.1	10.0	5.3	6.8
Entry and mid-zone				
Count	3	3	2	8
% within herniated disc	6.1	15.0	10.5	9.1
Mid- and exit zone				
Count	27	10	8	45
% within herniated disc	55.1	50.0	42.1	51.1
All zones				
Count	7	3	6	16
% within herniated disc	14.3	15.0	31.6	18.2
Total				
Count	49	20	19	88
% within herniated disc	100.0	100.0	100.0	100.0

Table 5. Macnab outcomes with the transforaminal endoscopic decompression surgery.

Item	Macnab Criteria			Cumulative Percent
	Frequency	Percent	Valid Percent	
Excellent	21	23.9	23.9	23.9
Good	55	62.5	62.5	86.4
Fair	12	13.6	13.6	100.0
Total	88	100.0	100.0	

the mid- and exit zone for central and paracentral herniations (Table 4).

Analyzing the severity of the stenosis grading by the radiologist showed higher variation in the severity assessment of foraminal stenosis in the various foraminal zones with an average severity score of 4.71 and a standard deviation of 2.626, with the highest variation in the entry zone (Figure 2). The AI severity grading of the foraminal zones on average was slightly higher at 5.65 with a standard deviation of 3.73 and more consistent across all zones except in the exit zone (Figure 3). The scatter plot of the severity grading (continuous scale) provided by the AI Multus RadBot versus the radiologist grading (ordinal scale) showed, as expected, that a nonlinear relationship between these 2 independent predictor variables existed, with the Multus RadBot consistently grading higher in nearly in all foraminal zones (Figure 4).

Macnab outcome analysis showed that 86.4% of foraminal decompressions resulted in excellent and good clinical outcomes at final follow-up (Table 5). There were no statistical correlations between clinical outcomes from the endoscopic decompression procedure and the type, size, or location of the herniation. Logit regression probability analysis of the 2 independent prognosticators employed in this study showed that both the grading by the radiologist (86.2% probability; Tables 6–9) of the foraminal stenosis and the AI grading (86.4% probability; Tables 10–13) were nearly equally predictive of a successful outcome with the endoscopic decompression. In other words, essentially every patient with an improved outcome was picked up by either 1 of the 2 independent predictor variables. The odds ratios for the predictors obtained by the exponentiation of the coefficients were nearly equal as well, with 1.267 for the radiologist grading and 1.264 for the Multus
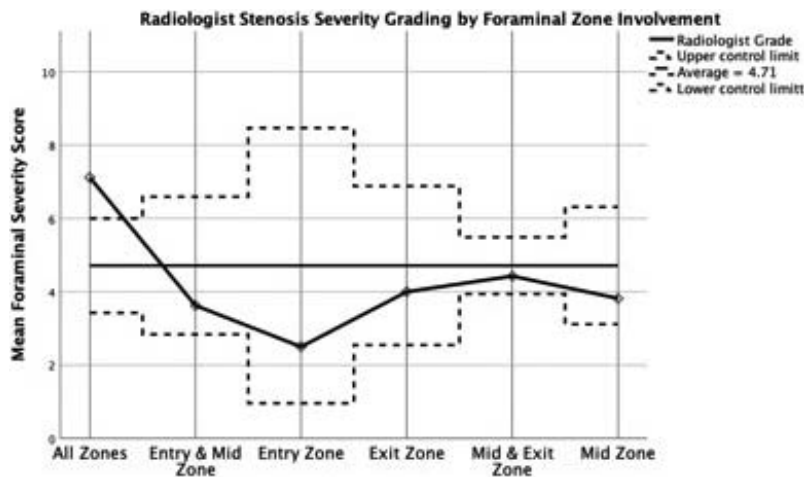

Figure 2. The foraminal severity grading provided by the radiologist plotted for the different foraminal zones showing an average score of 4.71 ± of 2.62 and variable grading across all zones.
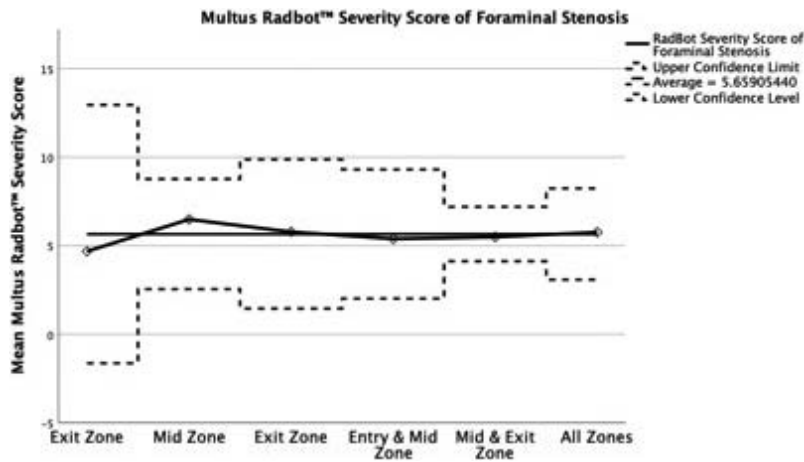
Figure 3. The AI Multus RadBot severity grading plotted for the different foraminal zones showing an average score of 5.65 ± of 3.73 and consistent grading across all zones except in the exit zone.

RadBot grading. The linear (Figure 5) and nonlinear (Figure 6) logit models for predicting improved clinical outcomes as defined by the dichotomized Macnab outcomes were graphically displayed for the radiologist ordinal grading and the contiguous AI grading by the Multus RadBot.

DISCUSSION

This correlative clinical study between spinal endoscopy outcomes and independent prognosticators of symptom relief in patients who suffer from the sciatica-type back and leg pain showed that a deep learning network is capable of identifying compressive pathology at a similar probability level as the radiologist. Successful surgical decompression proved that the Multus RadBot–generated reports on the painful pathology were equally useful as the radiologist's report in the treatment of herniated disc when using the directly visualized endoscopic decompression procedure.

The clinical outcomes that the authors found in this group of patients are comparable with previous clinical studies on the successful employment of endoscopy in the treatment of lumbar spinal stenosis and herniated disc. Most patients suffered from a contained herniated disc. The binary logistic regression analysis was best suited for this analysis of the clinical application of the AI deep learning neural network in routine lumbar MRI reading. as it allowed mixed statistical analysis of ordinal and continuous scale variables and categorical variables, such as the Macnab outcome criteria. The dichotomized use of the Macnab criteria as either improved or not improved greatly simplified the
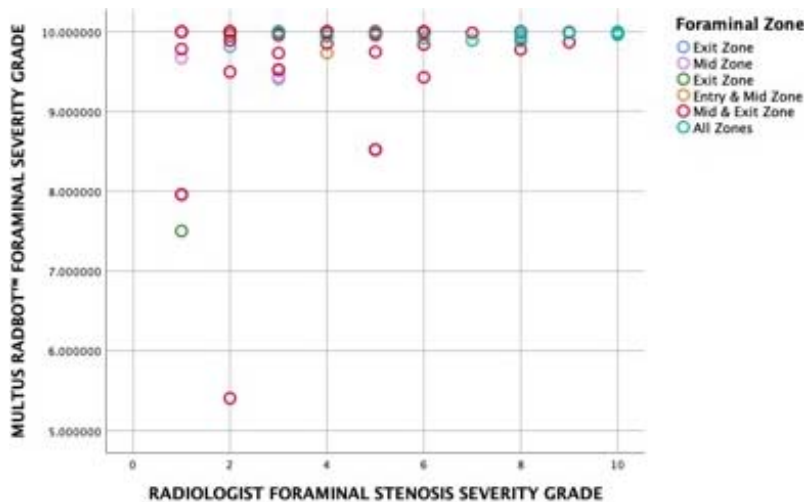

Figure 4. The scatter plot of the severity grading (continuous scale) provided by the artificial intelligence Multus RadBot versus the radiologist grading (ordinal scale) showing a nonlinear relationship between these 2 independent predictor variables, with the Multus RadBot consistently grading higher in nearly in all foraminal zones.

Table 6. Logit regression analysis of independent radiologist severity grade by dichotomized Macnab outcomes.

Observed	Predicted		Percentage Correct
	No Improvement	Improved	
No improvement	0	12	.0
Improved	0	75	100.0
Overall percentage	1 missing entry		86.2

Table 7. Variables in the equation for the logit regression analysis of independent radiologist severity grade by dichotomized Macnab outcomes.

	B	SE	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Radiologist foraminal stenosis grade[a]	.004	.119	.001	1	.974	1.004	.795	1.267
Constant	1.814	.637	8.124	1	.004	6.137		

[a]Variables entered on step 1: RadBot Foraminal Stenosis Severity Grade.
Abbreviations: B, unstandardized regression weight; SE, standard error; Wald, obtained by dividing the value of the coefficient by standard error σ; df, degrees of freedom; Sig., significance; Exp(B), odds ratios for the predictors obtained by the exponentiation of the coefficients; CI, confidence interval.

Table 8. Model summary for the logit regression analysis of independent radiologist severity grade by dichotomized Macnab outcomes.

Step	−2 Log Likelihood	Cox and Snell R^2	Nagelkerke R^2
1	69.806	.000	.000

Table 9. Hosmer and Lemeshow test for the logit regression analysis of independent radiologist severity grade by dichotomized Macnab outcomes.

Step	χ^2	df	Significance
1	3.633	6	.726

Abbreviations: df, degrees of freedom.

Table 10. Logit regression analysis of Multus RadBot severity grade by dichotomized Macnab outcomes.

Observed	Predicted		Percentage Correct
	No Improvement	Improved	
No improvement	0	12	.0
Improved	0	76	100.0
Overall percentage			86.4

Table 11. Variables in the equation for the logit regression analysis of Multus RadBot severity grade by dichotomized Macnab outcomes.

	B	SE	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
RadBot foraminal stenosis grade	.071	.083	.730	1	.393	1.074	.912	1.264
Constant	1.468	.518	8.016	1	.005	4.339		

[a]Variables entered on step 1: RadBot Foraminal Stenosis Severity Grade.
Abbreviations: B, unstandardized regression weight; SE, standard error; Wald, obtained by dividing the value of the coefficient by standard error σ; df, degrees of freedom; Sig., significance; Exp(B), odds ratios for the predictors obtained by the exponentiation of the coefficients; CI, confidence interval.

Table 12. Model summary for the logit regression analysis of Multus RadBot severity grade by dichotomized Macnab outcomes.

Step	−2 Log Likelihood	Cox and Snell R^2	Nagelkerke R^2
1	69.386[a]	.008	.015

[a]Variables entered on step 1: RadBot Foraminal Stenosis Severity Grade.

analysis and avoided a more complex and perhaps more-challenging-to-interpret multiple regression analysis. In comparison, the authors' binary linear logit analysis was practical. It did not find any statistically significant increases in odds ratios when testing any of the confounding factors in the ability of the Multus RadBot's or the radiologist's reporting to predict a successful outcome with the endoscopic decompression procedure more accurately. In other words, neither of the 2 severity scores—that provided by the radiologist and the other by the AI network—was sensitive enough to provide additional detail on the foraminal configuration the authors had hoped to stratify by cross tabulating these 2 independent predictor scores with the foraminal zone classification. It is clear from this study that additional, more detailed AI segmentation models would need to be developed that go beyond the routine MRI reporting provided by the radiologist.

The authors chose this simplified way of analyzing the level of probability of the Multus RadBot and radiologist's MRI reading, predicting successful clinical outcome with the spinal endoscopy, by applying the following assumptions: (1) the MRI report by the radiologist was employed as the gold standard in this analysis, and (2) the authors categorized the MRI findings in a straightforward manner with ordinal and contiguous severity scales of foraminal stenosis to distinguish between normal anatomy and stenosis. The authors' previous 2 research studies on the Multus RadBot employed accepted statistical methods of χ^2 testing to determine the sensitivity, specificity, positive and negative predictive value, the overall test reliability with the receiver operating characteristics (ROC) and area under the ROC curve method and the calculations of Cohen's alpha and kappa to demonstrate that the Multus RadBot is a high-quality diagnostic test

Table 13. Hosmer and Lemeshow test for the logit regression analysis of Multus RadBot severity grade by dichotomized Macnab outcomes.

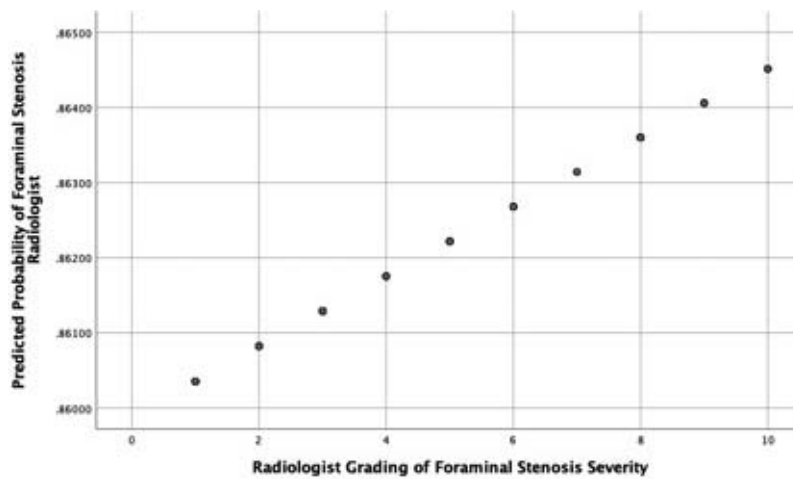| Step	χ^2	df	Significance
1	15.023	8	.059

Abbreviations: df, degrees of freedom.

Figure 5. Scatter plot of linear logit model describing the probability of improved clinical outcomes as defined by the dichotomized Macnab criteria predicted by the radiologist ordinal grading.

comparable to the routine MRI reading. A limitation of this simplified statistical analysis is assuming that the MRI reports provided by the reading radiologists were flawless. Perhaps that is one of the reasons that the probability of these 2 independent prognosticators to predict clinical success was limited to 86%. While there may have been other limitations at play, the side-by-side comparison of the radiologist and Multus RadBot predictions are similar to the real-world scenario, where routine lumbar MRI scans are read by 1 board-certified radiologist with little additional scrutiny. Clinical decision making on the most appropriate use of spinal endoscopy in the treatment of herniated disc or foraminal stenosis, particularly in the setting of multilevel lumbar degenerative disease, is currently based on a similar set of information. Ultimately, the AI deep learning neural network applications are only as smart as they were "taught" during the initial training phase. Therefore, future applications will likely be driven by "fine-tuning" the AI to clinically meaningful treatments of validated spinal pain generators. Clinician input is critical to such successful training.

CONCLUSIONS

This study demonstrated the application of AI deep learning networks to assist in the use of the lumbar MRI scan as a prognosticator of favorable clinical outcomes with the endoscopic spine surgery for foraminal stenosis due to a herniated disc. Future, more targeted AI applications in clinical decision making will have to focus on predominant pain generators, causing pain and disability in the functional context at the time when the spine care is
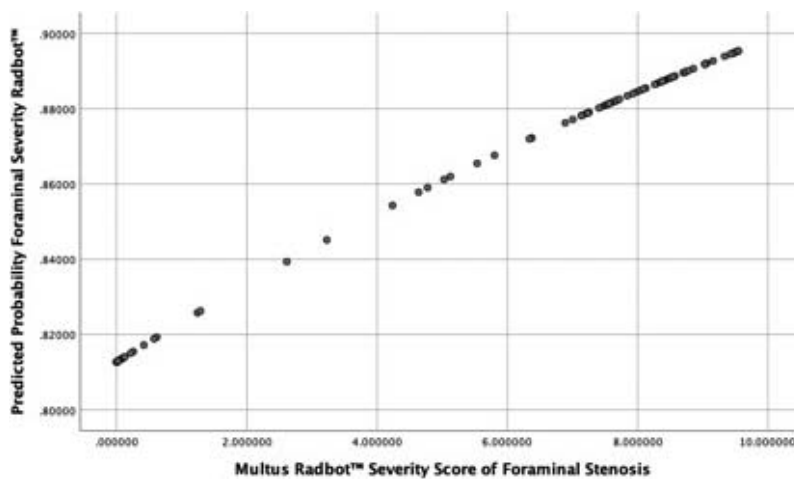

Figure 6. Scatter plot of nonlinear logit model describing the probability of improved clinical outcomes as defined by the dichotomized Macnab criteria predicted by the contiguous artificial intelligence grading by the Multus RadBot.

delivered. Further refinement of the AI segmentation models on MRI image findings of intraoperatively verified and validated pain generators responsive to treatment requires surgeon input that should be provided by only the most experienced and skilled critical opinion leader surgeons who can set the gold standard in expert endoscopic spine care.

REFERENCES

1. Yeung AT, Lewandrowski KU. Retrospective analysis of accuracy and positive predictive value of preoperative lumbar MRI grading after successful outcome following outpatient endoscopic decompression for lumbar foraminal and lateral recess stenosis. *Clin Neurol Neurosurg*. 2019;181:52.

2. Guyer R, Musacchio M, Cammisa FP, Jr., Lorio MP. ISASS recommendations/coverage criteria for decompression with interlaminar stabilization - coverage indications, limitations, and/or medical necessity. *Int J Spine Surg*. 2016;10:41.

3. Lewandrowski KU. Retrospective analysis of accuracy and positive predictive value of preoperative lumbar MRI grading after successful outcome following outpatient endoscopic decompression for lumbar foraminal and lateral recess stenosis. *Clin Neurol Neurosurg*. 2019;179:74–80.

4. Yeung AT, Yeung CA. In-vivo endoscopic visualization of patho-anatomy in painful degenerative conditions of the lumbar spine. *Surg Technol Int*. 2006;15:243–256.

5. Yeung AT, Gore S. In-vivo endoscopic visualization of patho-anatomy in symptomatic degenerative conditions of the lumbar spine ii: intradiscal, foraminal, and central canal decompression. *Surg Technol Int*. 2011;21:299–319.

6. Hua W, Zhang Y, Wu X, et al. Outcomes of discectomy by using full-endoscopic visualization technique via the interlaminar and transforaminal approaches in the treatment of L5-S1 disc herniation: An observational study. *Medicine (Baltimore)*. 2018;97(48):e13456.

7. Lewandrowski KU, Leon JFR, Yeung A. Use of "inside-out" technique for direct visualization of a vacuum vertically unstable intervertebral disc during routine lumbar endoscopic transforaminal decompression-a correlative study of clinical outcomes and the prognostic value of lumbar radiographs. *Int J Spine Surg*. 2019;13(5):399–414.

8. Lewandrowski KU, Dowling A, de Carvalho P, et al. Indication and contraindication of endoscopic transforaminal lumbar decompression. *World Neurosurg*. 2020.

9. O'Toole JE, Eichholz KM, Fessler RG. Minimally invasive far lateral microendoscopic discectomy for extraforaminal disc herniation at the lumbosacral junction: cadaveric dissection and technical case report. *Spine J*. 2007;7(4):414–421.

10. Yang D, Wu X, Zheng M, Wang J. A modified percutaneous endoscopic technique to remove extraforaminal disk herniation at the L5-S1 segment. *World Neurosurg*. 2018;119:e671–e678.

11. Choi G, Lee SH, Bhanot A, Raiturker PP, Chae YS. Percutaneous endoscopic discectomy for extraforaminal lumbar disc herniations: extraforaminal targeted fragmentectomy technique using working channel endoscope. *Spine (Phila Pa 1976)*. 2007;32(2):E93–99.

12. Sasani M, Ozer AF, Oktenoglu T, Canbulat N, Sarioglu AC. Percutaneous endoscopic discectomy for far lateral lumbar disc herniations: prospective study and outcome of 66 patients. *Minim Invasive Neurosurg*. 2007;50(2):91–97.

13. Lubbers T, Abuamona R, Elsharkawy AE. Percutaneous endoscopic treatment of foraminal and extraforaminal disc herniation at the L5-S1 level. *Acta Neurochir (Wien)*. 2012;154(10):1789–1795.

14. Safdar MF, Alkobaisi SS, Zahra FT. A comparative analysis of data augmentation approaches for magnetic resonance imaging (MRI) scan images of brain tumor. *Acta Inform Med*. 2020;28(1):29–36.

15. 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.

16. Yeung A, Lewandrowski KU. Early and staged endoscopic management of common pain generators in the spine. *J Spine Surg*. 2020;6(Suppl 1):S1–S5.

17. Lewandrowski K-U, Ransom NA. Five-year clinical outcomes with endoscopic transforaminal outside-in foraminoplasty techniques for symptomatic degenerative conditions of the lumbar spine. *J Spine Surg*. 2019:S54-S65.

18. Yeung A, Lewandrowski KU. Five-year clinical outcomes with endoscopic transforaminal foraminoplasty for symptomatic degenerative conditions of the lumbar spine: a comparative study of inside-out versus outside-in techniques. *J Spine Surg*. 2020;6(Suppl 1):S66–S83.

19. Sengupta DK, Herkowitz HN. Lumbar spinal stenosis. Treatment strategies and indications for surgery. *Orthop Clin North Am*. 2003;34(2):281–295.

20. Lewandrowski KU. "Outside-in" technique, clinical results, and indications with transforaminal lumbar endoscopic surgery: a retrospective study on 220 patients on applied radiographic classification of foraminal spinal stenosis. *Int J Spine Surg*. 2014;8:26. doi:10.14444/1026

21. Hoogland T. Percutaneous endoscopic discectomy. *J Neurosurg*. 1993;79(6):967–968.

22. Hoogland T, Scheckenbach C. [Percutaneous lumbar nucleotomy with low-dose chymopapain, an ambulatory procedure]. *Z Orthop Ihre Grenzgeb*. 1995;133(2):106–113.

23. Schubert M, Hoogland T. Endoscopic transforaminal nucleotomy with foraminoplasty for lumbar disk herniation. *Oper Orthop Traumatol*. 2005;17(6):641–661.

24. Hoogland T, Schubert M, Miklitz B, Ramirez A. Transforaminal posterolateral endoscopic discectomy with or without the combination of a low-dose chymopapain: a prospective randomized study in 280 consecutive cases. *Spine (Phila Pa 1976)*. 2006;31(24):E890–897.

25. Hoogland T, van den Brekel-Dijkstra K, Schubert M, Miklitz B. Endoscopic transforaminal discectomy for recurrent lumbar disc herniation: a prospective, cohort evaluation of 262 consecutive cases. *Spine (Phila Pa 1976)*. 2008;33(9):973–978.

26. Lee CK, Rauschning W, Glenn W. Lateral lumbar spinal canal stenosis: classification, pathologic anatomy and surgical decompression. *Spine (Phila Pa 1976)*. 1988;13(3):313–320.

27. Milette PC. Classification, diagnostic imaging, and imaging characterization of a lumbar herniated disk. *Radiol Clin North Am*. 2000;38(6):1267–1292.

28. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral

disc degeneration. *Spine (Phila Pa 1976)*. 2001;26(17):1873–1878.

29. Lee S, Kim SK, Lee SH, et al. Percutaneous endoscopic lumbar discectomy for migrated disc herniation: classification of disc migration and surgical approaches. *Eur Spine J.* 2007;16(3):431–437.

30. Hasegawa T, An HS, Haughton VM, Nowicki BH. Lumbar foraminal stenosis: critical heights of the intervertebral discs and foramina. A cryomicrotome study in cadavera. *J Bone Joint Surg Am*. 1995;77(1):32–38.

31. Yeung AT. The evolution and advancement of endoscopic foraminal surgery: one surgeon's experience incorporating adjunctive techologies. *SAS J.* 2007;1(3):108–117.

32. el-Khoury GY, Ehara S, Weinstein JN, Montgomery WJ, Kathol MH. Epidural steroid injection: a procedure ideally performed with fluoroscopic control. *Radiology*. 1988;168(2):554–557.

33. Geurts JW, Kallewaard JW, Richardson J, Groen GJ. Targeted methylprednisolone acetate/hyaluronidase/clonidine injection after diagnostic epiduroscopy for chronic sciatica: a prospective, 1-year follow-up study. *Reg Anesth Pain Med.* 2002;27(4):343–352.

34. Valat JP. Epidural corticosteroid injections for sciatica: placebo effect, injection effect or anti-inflammatory effect? *Nat Clin Pract Rheumatol.* 2006;2(10):518–519.

35. MacVicar J, King W, Landers MH, Bogduk N. The effectiveness of lumbar transforaminal injection of steroids: a comprehensive review with systematic analysis of the published data. *Pain Med.* 2013;14(1):14–28.

36. Chang MC, Lee DG. Outcome of transforaminal epidural steroid injection according to the severity of lumbar foraminal spinal stenosis. *Pain Physician*. 2018;21(1):67–72.

37. Lewandrowski KU. Successful outcome after outpatient transforaminal decompression for lumbar foraminal and lateral recess stenosis: the positive predictive value of diagnostic epidural steroid injection. *Clin Neurol Neurosurg*. 2018;173:38–45.

38. Macnab, I. (1971). "Negative disc exploration. An analysis of the causes of nerve-root involvement in sixty-eight patients." *J Bone Joint Surg Am* 53(5): 891–903.

39. Stokes IA. Surface strain on human intervertebral discs. *J Orthop Res* 1987;5:348–355.

40. Fenyo A, Shinis D, Shelef I, et al. [Lumbar disc herniation: protrusion, extrusion or bulge? The proper use of the terms - how and when will it be defined as a disease?]. *Harefuah* 2019;158:807–811.

41. Yuan S, Zou Y, Li Y, et al. A clinically relevant MRI grading system for lumbar central canal stenosis. *Clin Imaging* 2016;40:1140–1145.

42. Lee GY, Lee JW, Choi HS, et al. A new grading system of lumbar central canal stenosis on MRI: an easy and reliable method. *Skeletal Radiol* 2011;40:1033–1039.

43. Lee CK, Rauschning W, Glenn W. Lateral lumbar spinal canal stenosis: classification, pathologic anatomy and surgical decompression. *Spine (Phila Pa 1976)* 1988;13:313–320.

44. Lee S, Lee JW, Yeom JS, et al. A practical MRI grading system for lumbar foraminal stenosis. *AJR Am J Roentgenol* 2010;194:1095–1098.

Disclosures and COI: The views expressed in this article represent those of the authors and no other entity or organization. The first author has no direct (employment, stock ownership, grants, patents), or indirect conflicts of interest (honoraria, consultancies to sponsoring organizations, mutual fund ownership, paid expert testimony). He is not currently affiliated with or under any consulting agreement with any MRI vendor that the clinical research data conclusion could directly enrich. This manuscript is not meant for or intended to push any other agenda other than reporting the research data related on automated recognition of common painful spine pathologies by deep neural network learning. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Corresponding Author: Kai-Uwe Lewandrowski, MD, Staff Orthopaedic Spine Surgeon Center for Advanced Spine Care of Southern Arizona and Surgical Institute of Tucson, AZ 85712. Phone: (520) 204-1495; Fax: (623) 218-1215; Email: busniess@tucsonspine.com.