Article published online: 2022-08-25



# Use of Artificial Intelligence for the Development of Predictive Model to Help in Decision-Making for Patients with Degenerative Lumbar Spine Disease

Gaurav Purohit[1] Madhur Choudhary[1] V. D. Sinha[1]

[1] Department of Neurosurgery, SMS Medical College, Jaipur, Rajasthan, India

**Address for correspondence** Madhur Choudhary, MBBS, MS, MCh, B-193, Bapu Nagar, University Marg, Jaipur 302015, Rajasthan, India (e-mail: madhurchoudhary969@gmail.com).

AJNS 2022;17:274–279.

## Abstract

**Context** The aim of the study was to develop a prognostic model using artificial intelligence for patients undergoing lumbar spine surgery for degenerative spine disease for change in pain, functional status, and patient satisfaction based on preoperative variables included in following categories—sociodemographic, clinical, and radiological.

**Methods and Materials** A prospective cohort of 180 patients with lumbar degenerative spine disease was included and divided into three classes of management—conservative, decompressive surgery, and decompression with fixation. Preoperative variables, change in outcome measures (visual analog scale—VAS, Modified Oswestry Disability Index—MODI, and Neurogenic Claudication Outcome Score—NCOS), and type of management were assessed using Machine Learning models. These were used for creating a predictive tool for deciding the type of management that a patient should undergo to achieve the best results. Multivariate logistic regression was also used to identify prognostic factors of significance.

**Results** The area under the curve (AUC) was calculated from the receiver-operating characteristic (ROC) analysis to evaluate the discrimination capability of various machine learning models. Random Forest Classifier gave the best ROC-AUC score in all three classes (0.863 for VAS, 0.831 for MODI, and 0.869 for NCOS), and the macroaverage AUC score was found to be 0.842 suggesting moderate discriminatory power. A graphical user interface (GUI) tool was built using the machine learning algorithm thus defined to take input details of patients and predict change in outcome measures.

**Conclusion** This study demonstrates that machine learning can be used as a tool to help tailor the decision-making process for a patient to achieve the best outcome. The GUI tool helps to incorporate the study results into active decision-making.

**Keywords**
- neurosurgery
- arificial intelligence
- degenerative lumbar spine disease
- machine learning
- Random Forest Classifier
- lumbar canal stenosis

**DOI** https://doi.org/10.1055/s-0042-1750785.
**ISSN** 2248-9614.



Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

## Introduction

Low back pain (LBP) is one of the leading causes of nonfatal health loss.[1] LBP can be caused by the involvement of any spinal structure that is innervated and is susceptible to disease or injury.[2]

In the context of LBP, degenerative changes in the spine were subcategorized into spondylolisthesis, disc degeneration, and spinal stenosis to outline the types of pathology and the potential ramification of surgical intervention.

Surgical treatment is effective—lumbar discectomy being the standard surgical procedure for patients with lumbar disc herniation.[3] Spontaneous regression of herniated disc tissue can occur in most patients, and can be treated with conservative strategies.[4]

Bony decompression by laminectomy is still considered the gold standard of surgery and the most common technique for lumbar spinal stenosis.[5] Leg pain relief and better back-related functional status favored those initially receiving surgical treatment.[6]

Laminectomy and posterior instrumented spinal fusion are the standard of care and are the most commonly performed surgical procedure for the management of degenerative spondylolisthesis.[7] Patients with degenerative spondylolisthesis and spinal stenosis treated surgically showed considerably more improvement in pain and function than patients treated conservatively.[8]

The surgical outcome can be measured as a change in symptoms intensity, resumption of activity, and patient satisfaction.

Machine learning (ML) classification is a domain of artificial intelligence that enables algorithms to learn patterns in large, complex datasets and generate useful predictive outputs.[9] It represents a set of powerful technologies capable of effectively predicting outcomes to support decision-making in neurosurgery.[10]

## Objective

The aim of the study was to develop a prognostic model using artificial intelligence for patients undergoing lumbar spine surgery for degenerative spine disease for change in pain, functional status, and patient satisfaction based on preoperative variables (sociodemographic, clinical, and radiological).

## Methods and Materials

### Study Population

This prospective study was conducted in the Neurosurgery Department of Sawai Man Singh Medical College, Jaipur, after obtaining ethical clearance from University Ethics Committee. Patients included those undergoing lumbar spine surgery for degenerative spine disease. These included Open discectomy for disk degeneration with prolapsed intervertebral disk, Decompression – Open laminectomy with foraminotomy for patients with Spinal canal stenosis, Decompression with posterolateral fixation for those with spinal canal stenosis with instability. All patients failed to respond to at least 6 weeks of conservative management, including physical therapy, anti-inflammatory medications, and analgesics. A cohort of patients managed conservatively were followed up for 6 months.

### Inclusion Criteria

Patients older than 18 years of age were admitted from January 2019 to January 2021 for complaints of LBP with or without radiculopathy, diagnosed as having degenerative spine disease using symptomatology and magnetic resonance imaging (MRI).

### Exclusion Criteria

Exclusion criteria included trauma, neoplasm, infection, congenital deformation, and chronic illness such as rheumatoid arthritis. Patients with an extraspinal cause of back/neck pain or radiculopathy were excluded.

### Prognostic Factors

#### Sociodemographic Factors

Sociodemographic factors included age, gender, body mass index (BMI), occupation (sedentary, light, medium, heavy), and smoking status.

#### Symptomatology

A reliable questionnaire was filled out by patients before surgery that included severity of back pain as compared with leg pain, duration of symptoms, whether ambulant independently or with support or bedridden, history of previous lumbar spine surgery, Hamilton Anxiety scale (HAM-A), visual analog scale (VAS), Neurogenic Claudication Outcome Score (NCOS) and Modified Oswestry Disability Index (MODI). A thorough neurological examination was done and the presence of neurological deficit was noted that included objective weakness and cauda equina syndrome.

#### Psychosocial Risk Factors

HAM-A was developed to measure the severity of anxiety symptoms.[11] It was used as an attempt to remove the confounding effect of subjective patient-reported scores.

#### Radiological Factors

Pfirrmann grading system is used as a standardized and reliable assessment of disc morphology based on MRI.[12] Disc herniation was categorized into normal, symmetric disc bulging, disc protrusion, disc extrusion, and free fragment.[3]

Modic et al described the types of signal changes, classification criteria of the lumbar endplate, and bone marrow changes on MRI scans.[13]

Central canal area at the level of maximum compression was used to grade central spinal canal stenosis.

Lateral spinal canal stenosis was graded according to Bartynski and Lin.[14]

Foraminal spinal stenosis was graded according to perineural intraforaminal fat,[15] hypertrophic facet degeneration,[16] foraminal nerve root impingement,[17] and size and shape of the foramen.[16]

### Outcome Measures

The primary outcome measures were the VAS, MODI, and NCOS collected preoperatively and at 6 months postoperatively during clinic review.

### Surgical Procedures

The study surgeons had an average of 10 years' experience in spinal surgery in the neurosurgery department and regularly performed the surgeries employing standard procedures. Bony decompression by laminectomy and foraminotomy for lumbar canal stenosis and discectomy for disc herniation were classified as decompression (management—one) and laminectomy with posterior instrumented spinal fixation for degenerative spondylolisthesis as decompression with fixation (management—two). Patients managed conservatively were included in management—zero.

### Data Analysis

For the prediction of improvement in various indices, we subtracted the index value before and after surgery. Later the improvement was classified into various classes. VAS improvement was divided into four classes, that is,, 0, 1, 2, and 3 for a range of 0 to 2, 3 to 4, 5 to 6, and 7 to 10, respectively. MODI improvement was also divided into four classes for a range of 0 to 5, 5 to 10, 11 to 15, and more than 15. NCOS improvement was further divided into four classes for a range of 0 to 10, 11 to 20, 21 to 30, and more than 30. Label encoding was used for other discrete input classes that could be used to train the model.

For classification, different algorithms used were Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Support Vector Machine, and K-Nearest Neighbor. Two different models for each type of classifier were trained. The first model was trained with the sociodemographic factors, symptomatology scoring, psychosocial factors, and radiological factors as mentioned previously. The second model was trained with the type of management used.

The dataset was divided into training and testing datasets. The first subset is used to fit the model and is referred to as the training dataset that is 80% of the complete dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared with the expected values. This second dataset is referred to as the test dataset that is the remaining 20% of the dataset left apart. The dataset was divided in such a way that the class ratio of all classes remained proportional in the training and test dataset. The objective was to estimate the performance of the ML model on new data (data not used to train the model and have all the classes present for validation).

Example: Training of VAS improvement model with logistic regression—the first logistic regression model was trained with features like sex and BMI, as mentioned previously. The second model was trained with management class as input. The final predicted probability was the average of individual probability and the outcome was evaluated based on ROC (receiver-operating characteristic) – AUC (area under the curve) score. A similar approach was followed to train all other models.

## Results

There were a total of 180 patients with lumbar spine disease over 18 years of age enrolled in our study. Our study had 53 patients treated conservatively, 96 patients underwent decompression, and 31 patients underwent decompression with posterior instrumented spinal fixation. The average age of patients in the study was 50.26 years. About 40.55% of patients had complaints of back pain without radiculopathy, whereas the rest of the patients complained of radiculopathy with or without back pain. Forty-two (23.33%) patients had symptoms for less than 3 months of which 15 were treated conservatively, 22 underwent decompression, and 5 underwent decompression with fixation. Fifty-eight (32.22%) patients had symptoms for 3 to 6 months of which 22 were treated conservatively and 30 underwent decompression and 6 underwent fixation. Eighty (44.44%) patients had symptoms for more than 6 months of which 16 were treated conservatively, 44 underwent decompression, and 20 underwent decompression with fixation. Nine (5%) patients had a history of previous lumbar spine surgery. Twenty-six (14.4%) patients presented with neurological deficit out of which 25 were operated on and only 4 had improvement in their weakness postoperatively. These four had symptoms for less than 3 months.

The average improvement in VAS, NCOS, and MODI scores categorized according to the management is given in ►**Table 1**.

For patients managed conservatively, multiple logistic regression applied to the dataset obtained showed active occupation associated with improved outcome in VAS, lesser BMI, back pain not associated with radiculopathy, and no foraminal nerve root impingement on MRI associated with improved outcome in MODI scores. The preoperative scores in the patients treated conservatively were not that poor, which could explain the lack of significant improvement in the scores.

For patients undergoing lumbar decompression, multiple logistic regression analysis suggested younger age, independent ambulation, duration of symptoms less than 3 months, and back pain without associated radiculopathy and no foraminal nerve root impingement on MRI associated with improved outcome in VAS; younger age, independent ambulation, and lesser disc degeneration grading on MRI were associated with improved outcome in MODI scores; back pain without radiculopathy was associated with improved outcome in NCOS.

For patients undergoing lumbar decompression with fixation, multiple logistic regression applied showed back pain not associated with radiculopathy and no foraminal nerve root impingement on MRI associated with improved outcome in VAS, lesser BMI, involvement in sports activities, no past spine surgery, and no neurological deficit associated with improved outcome in NCOS.

 

**Table 1** Average improvement in VAS, NCOS, and MODI scores according to management

| | Conservative | Decompression | Decompression + fixation | *p*-Value |
|---|---|---|---|---|
| VAS score (0–10) | | | | |
| Before | 6.09 ± 1.19 | 8.22 ± 1.17 | 7.97 ± 1.14 | 0.02 |
| After | 3.34 ± 1.48 | 2.73 ± 1.50 | 3.42 ± 1.63 | < 0.001 |
| Improvement | 2.75 ± 1.81 | 5.49 ± 1.89 | 4.55 ± 1.69 | < 0.001 |
| MODI (6–60); 60—maximum disability | | | | |
| Before | 14.94 ± 4.07 | 16.60 ± 3.89 | 17.84 ± 3.62 | < 0.001 |
| After | 20.04 ± 5.11 | 29.47 ± 5.11 | 29.97 ± 3.82 | < 0.001 |
| Improvement | 5.09 ± 3.06 | 12.86 ± 4.95 | 12.13 ± 2.91 | < 0.001 |
| NCOS (0–100); 100—asymptomatic, full function | | | | |
| Before | 85.15 ± 7.30 | 82.70 ± 7.65 | 80.23 ± 7.87 | 0.011 |
| After | 78.66 ± 8.46 | 61.17 ± 10.48 | 59.81 ± 8.32 | < 0.001 |
| Improvement | 6.49 ± 4.32 | 21.53 ± 11.04 | 20.42 ± 8.55 | < 0.001 |

Abbreviations: MODI, Modified Oswestry Disability Index; NCOS, Neurogenic claudication Outcome Score; VAS, visual analog scale.
Average improvement in VAS score, NCOS, and MODI scores in patients between preoperative period and at 6 months follow-up is described. It is categorized based on management strategy into conservative, decompression, and decompression with fixation.

The AUC was calculated from the ROC analysis to evaluate the discrimination capability of various ML models (►**Table 2**). Random Forest Classifier gave the best ROC-AUC score in all three classes and was therefore used. The AUC score for VAS, MODI, and NCOS was 0.863, 0.831, and 0.869, respectively, and the macroaverage AUC score was found to be 0.842.

A graphical user interface (GUI) tool was built to take input details of patients. The surgeon can select the initial management used, based on the inputs the model predicts the improvement in all three indices based on management. Further, the surgeon can change the type of management used and see the difference in the improvement in the indices and find the best suitable management for the particular patient.

## Discussion

Degenerative lumbar spine disease is a multifactorial entity in its causation, pathology as well as management. Anatomy of the spine and its relationship with spinal cord and nerve roots is complex, thus degenerative disease process causing symptoms has to be dealt with keeping in mind multiple variables involved.

A longer duration of symptoms more than 3 months was associated with a less favorable outcome as measured by improvement in VAS. Similar results have been found in previous studies.[18–20]

Lesser BMI was associated with improved outcomes in patients managed conservatively and those undergoing spinal decompression with instrumented fixation as measured by improvement in MODI scores and NCOS, respectively. In the Spine Patient Outcomes Research Trial, obese patients showed less improvement from baseline with conservative management.[21]

Independent ambulation preoperatively was associated with better outcomes in patients undergoing decompression surgery as measured by improvement in VAS and MODI scores.

In our study, leg symptoms signifying radiculopathy were associated with poor outcomes in all management groups.

**Table 2** Discrimination capability of the machine learning models for VAS, NCOS, and MODI scores

| Machine learning algorithm | VAS improvement ROC-AUC score | MOD index improvement ROC-AUC score | NCOS improvement ROC-AUC score |
|---|---|---|---|
| Logistic Regression | 0.817 | 0.829 | 0.826 |
| Decision Tree Classifier | 0.753 | 0.613 | 0.657 |
| Random Forest Classifier | 0.863 | 0.831 | 0.869 |
| Support Vector Machine | 0.768 | 0.732 | 0.673 |
| K-Nearest Neighbor | 0.681 | 0.689 | 0.614 |

Abbreviations: MOD, Modified Oswestry Disability Index; NCOS, Neurogenic Claudication Outcome Score; ROC-AUC, receiver-operating characteristic-area under the curve; VAS, visual analog scale.
The AUC was calculated from the ROC analysis to evaluate the discrimination capability of various machine learning algorithms. Random Forest Classifier gave the best ROC-AUC score in all three classes and was therefore used.

 

Neurological deficit was associated with poor outcomes following decompression and fixation as denoted by NCOS. Many studies suggest that the presence of a radicular deficit (i.e., foot drop) presurgery is a negative predictive factor in terms of patient satisfaction.[22,23]

History of past lumbar spine surgery was associated with poorer outcomes in patients undergoing decompression with fixation as denoted by NCOS. A study done by Hébert et al showed previous spine surgery to be associated with poor leg pain outcome following surgery for degenerative lumbar spine surgery.[24]

Our aim with this study is to investigate prognostic factors for clinical outcome after lumbar spine surgery for degenerative spine disease and create a prognostic model using artificial intelligence that could potentially aid in decision-making.

This study shows that a ML model could be used as a predictive tool for deciding the type of management that a patient should undergo to achieve the best results. A combination of various parameters, in a ML model, could be applied to estimate the improvement in patient scores with a high degree of accuracy.

The area under the ROC curve (AUC) is widely recognized as the measure of a diagnostic test's discriminatory power. The maximum value for the AUC is 1.0, thereby indicating a (theoretically) perfect test (i.e., 100% sensitive and 100% specific). An AUC value of 0.5 indicates no discriminative value (i.e., 50% sensitive and 50% specific).[25]

The macroaverage AUC score was found to be 0.842 that is demonstrating moderate discriminatory power, therefore it suggests potential utility as a tool or a support system that could be used by experts as one of the inputs into the decision-making process.

The limitation of this model would be a small dataset. To make the model more accurate, the study should be repeated with larger dataset that would make the prediction more accurate. More variables could be added that would help in a truer prediction of outcome following the management.

Randomized clinical trials are needed to establish benefits and to confirm these findings. Finally, artificial intelligence will never replace human expert decision-makers, but they can assist in double-checking and enhancing the routine decision-making process and thus help the patient.

The toolkit can be accessed online on URL: http://134.209.148.167:5000.

## Conclusion

This study demonstrates that ML can be used as a tool to help tailor the decision-making process for a patient to achieve the best results. The GUI tool helps to incorporate the study results into active decision-making. This study would encourage further inroads into the use of artificial intelligence in the medical field for assistance in the decision-making in patient management.

**Conflict of Interest**
None declared.

## References


1 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392 (10159):1789–1858 10.1016/S0140-6736(18)32279-7 Erratum in: Lancet. 2019 Jun 22;393(10190):e44. PMID: 30496104; PMCID: PMC6227754

2 DePalma MJ, Ketchum JM, Saullo T. What is the source of chronic low back pain and does age play a role? Pain Med 2011;12(02): 224–233

3 Chen BL, Guo JB, Zhang HW, et al. Surgical versus non-operative treatment for lumbar disc herniation: a systematic review and meta-analysis. Clin Rehabil 2018;32(02):146–160

4 Chiu CC, Chuang TY, Chang KH, Wu CH, Lin PW, Hsu WY. The probability of spontaneous regression of lumbar herniated disc: a systematic review. Clin Rehabil 2015;29(02): 184–195

5 Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis: updated Cochrane Review. Spine 2005;30(20):2312–2320

6 Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the maine lumbar spine study. Spine 2005;30(08):936–943

7 Eismont FJ, Norton RP, Hirsch BP. Surgical management of lumbar degenerative spondylolisthesis. J Am Acad Orthop Surg 2014;22 (04):203–213

8 Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. N Engl J Med 2007;356(22):2257–2270

9 Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015;349(6245):255–260

10 Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev 2020;43(05):1235–1253

11 Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(01):50–55

12 Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 2001;26(17):1873–1878

13 Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;166(1 Pt 1):193–199

14 Bartynski WS, Lin L. Lumbar root compression in the lateral recess: MR imaging, conventional myelography, and CT myelography comparison with surgical confirmation. AJNR Am J Neuroradiol 2003;24(03):348–360

15 Wildermuth S, Zanetti M, Duewell S, et al. Lumbar spine: quantitative and qualitative assessment of positional (upright flexion and extension) MR imaging and myelography. Radiology 1998; 207(02):391–398. Doi: 10.1148/radiology.207.2.9577486Erratum in: Radiology 1998 Sep;208(3):834. PMID: 9577486

16 Andreisek G, Imhof M, Wertli M, et al; Lumbar Spinal Stenosis Outcome Study Working Group Zurich. A systematic review of semiquantitative and qualitative radiologic criteria for the diagnosis of lumbar spinal stenosis. AJR Am J Roentgenol 2013;201 (05):W735-46

17 Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. Spine 2008;33(14):1605–1610

18 Schoenfeld AJ, Bono CM. Does surgical timing influence functional recovery after lumbar discectomy? A systematic review. Clin Orthop Relat Res 2015;473(06):1963–1970

19 Nygaard OP, Kloster R, Solberg T. Duration of leg pain as a predictor of outcome after surgery for lumbar disc herniation: a


 

prospective cohort study with 1-year follow up. J Neurosurg 2000;92(02, Suppl):131–134
20 Ailon T, Tee J, Manson N, et al. Patient-reported outcomes following surgery for degenerative spondylolisthesis: comparison of a universal and multitier health care system. Spine J 2019;19(01):24–33
21 Rihn JA, Radcliff K, Hilibrand AS, et al. Does obesity affect outcomes of treatment for lumbar stenosis and degenerative spondylolisthesis? Analysis of the Spine Patient Outcomes Research Trial (SPORT). Spine 2012;37(23):1933–1946
22 Masuda S, Kanba Y, Kawai J, Ikeda N. Prognostic factors for drop foot due to lumbar degenerative diseases: the impact of surgical timing on postoperative recovery. Clin Spine Surg 2020;33(04):160–162
23 Takenaka S, Aono H. Prediction of postoperative clinical recovery of drop foot attributable to lumbar degenerative diseases, via a Bayesian Network. Clin Orthop Relat Res 2017;475(03):872–880
24 Hébert JJ, Abraham E, Wedderkopp N, et al. Preoperative factors predict postoperative trajectories of pain and disability following surgery for degenerative lumbar spinal stenosis. Spine 2020;45(21):E1421–E1430
25 Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. CJEM 2006;8(01):19–20

 