Artificial Intelligence

Prediction of Therapeutic Outcome after Transforaminal Epidural Steroid Injection in Patients with Radiculopathy following Lumbar Spinal Stenosis Using Machine Learning

Jeoung Kun Kim, PhD[1], Jung Hwan Park, MD[2], and Min Cheol Chang, MD[2]

From: [1]Department of Business Administration, School of Business, Yeungnam University, Gyeongsan-Si, Republic of Korea; [2]Department of Physical Medicine and Rehabilitation, College of Medicine, Yeungnam University, Daegu, Republic of Korea

Address Correspondence: Min Cheol Chang, MD Department of Physical Medicine and Rehabilitation, College of Medicine, Yeungnam University 317-1, Daemyungdong, Namku, Taegu, 705-717, Republic of Korea E-mail: wheel633@ynu.ac.kr

Disclaimer: JK KIm and JH Park contributed equally to the work. This work was supported by the National Research Foundation of Korea grant funded by the Korean government (MSIT) (No. RS2023-00219725).

Conflict of interest: Each author certifies that he or she, or a member of his or her immediate family, has no commercial association (i.e., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

Article received: 11-08-2025
Revised article received: 02-05-2026
Accepted for publication: 03-03-2026

Free full article: www.painphysicianjournal.com

Background: Transforaminal epidural steroid injections are widely used to manage radicular pain caused by lumbar spinal stenosis; however, therapeutic outcomes vary among patients. Accurately predicting a transforaminal epidural steroid injection's response may help clinicians optimize treatment planning.

Objectives: Our study aimed to develop and validate machine learning models, particularly a deep learning neural network, to predict therapeutic outcomes post transforaminal epidural steroid injections using both clinical and magnetic resonance imaging data.

Study Design: A study using machine learning analysis.

Setting: The spine center of a university hospital.

Methods: A retrospective dataset of 188 patients with lumbar spinal stenosis who underwent a single-level transforaminal epidural steroid injection was analyzed. Nine input variables were included: age, gender, injection side, injection level, pain duration, pretreatment Numeric Rating Scale pain score, predominant stenotic location, and the degrees of central and foraminal stenosis. A positive therapeutic outcome was defined as a ≥ 50% reduction in the Numeric Rating Scale score at one month postprocedure. Deep learning neural network, XGBoost, and CatBoost algorithms were used to develop the models. Model performance was evaluated using receiver operating characteristic analysis and standard classification metrics.

Results: The deep learning neural network model, configured with 5 fully connected layers, had the best performance among the 3 models, achieving a validation accuracy of 0.868 and an area under the curve (AUC) of 0.882 (95% CI, 0.733–0.994). CatBoost had an intermediate performance with an accuracy of 0.789 and an AUC of 0.765 (95% CI, 0.593–0.906), whereas XGBoost had the lowest accuracy of 0.763 and an AUC of 0.662 (95% CI, 0.458–0.853). In addition, the deep learning neural network model achieved balanced precision and recall across both favorable and poor outcome classes, with a weighted F1 score of 0.869.

Limitations: This was a single-center retrospective study with a small sample size.

Conclusions: Deep learning demonstrated superior predictive performance compared with conventional machine learning algorithms in forecasting transforaminal epidural steroid injection outcomes among patients with lumbar spinal stenosis. Integrating clinical and magnetic resonance imaging-derived features through a deep learning neural network may facilitate individualized treatment decisions and enhance therapeutic planning for lumbosacral radiculopathy. Future multicenter studies with larger datasets are warranted for external validation and clinical implementation.

Key words: Transforaminal epidural steroid injection, radiculopathy, lumbar spine, spinal stenosis, deep learning, machine learning

Pain Physician 2026: 29:E663-E670

Lumbar spinal stenosis (LSS) is a common cause of musculoskeletal pain. It occurs due to narrowing of the lumbar spinal canal or lumbar vertebral foramen, often resulting from degenerative changes such as intervertebral disc degeneration, osteophyte formation, and thickening of the spinal ligaments (1). This narrowing compresses the nerve root and surrounding vascular structures, leading to an inflammatory response mediated by inflammation-related cells and proinflammatory cytokines, which subsequently results in lumbosacral radicular pain (2). In addition, vascular compromise around the nerve root may lead to ischemia, which further contributes to lumbar radicular pain (2).

Lumbosacral radicular pain associated with LSS typically worsens during walking or prolonged standing, resulting in a reduced quality of life and physical function (3). To alleviate pain from LSS, oral medications, physiotherapy, and injection therapies are commonly used (4-6). Among these, transforaminal epidural steroid injections (TFESI) are considered to be one of the most effective and widely used procedures for pain management in clinical practice (2,7). Corticosteroids exert therapeutic effects by inhibiting the activity of inflammatory cells and suppressing the synthesis of proinflammatory mediators, thereby alleviating lumbosacral radicular pain (2,8).

To avoiding unnecessary procedures, accurately predicting TFESI's therapeutic outcomes is crucial for selecting appropriate treatment strategies for patients with LSS. Moreover, predictive insights enable clinicians to provide patients with reliable information regarding expected treatment outcomes. Although previous studies have examined associations between magnetic resonance imaging (MRI) findings and TFESIs efficacy, treatment prognosis is influenced not only by imaging findings but also by various clinical factors (2,9,10). In addition, multiple MRI parameters may collectively contribute to determining the therapeutic response to TFESI; traditional statistical methods have limitations when analyzing outcomes based on multiple variables.

Recently, deep learning has emerged as a powerful computational approach capable of handling complex, high-dimensional data and capturing nonlinear relationships among variables (11-13). Machine learning refers to computer algorithms that can automatically learn from data without explicit programming (14,15). Deep learning is an advanced machine learning that employs multiple hidden layers to construct artificial neural networks, mimicking certain structural and functional aspects of the human brain (16,17). Several studies have demonstrated that deep learning techniques can outperform conventional machine learning algorithms for predictive accuracy (18-20).

Our study aimed to apply deep learning techniques to predict the therapeutic outcomes post TFESI in patients with LSS. Clinical information and MRI features were incorporated as input variables to enhance predictive performance. Additionally, the deep learning model's performance was compared with conventional machine learning algorithms' performance to evaluate its relative effectiveness.

Methods

Patients

Our retrospective study included 188 patients (mean [SD] age, 66.3 [8.9] years; men: women = 68: 120; right: left: bilateral = 71:82:35; injection levels L2:L3:L4:L5:S1 = 1:4:33:145:5; pain duration, 17.8 [32.1] months; pretreatment Numeric Rating Scale [NRS-11, 0 = no pain, 10 = worst pain], 5.2 [1.3]; predominant stenotic location, central stenosis: foraminal stenosis: central and foraminal stenosis = 62:92:34; degree of central stenosis, 0:1:2:3 = 58:56:52:22; degree of foraminal stenosis, 0:1:2:3 = 30:46:71:41) who visited the spine center of a university hospital and underwent a lumbar TFESI for LSS from January 2013 through December 2021.

Inclusion criteria were: 1) single-level lumbar TFESI for segmental pain that radiated to the lower extremity due to LSS; 2) ≥ 50% temporary pain relief following a diagnostic nerve root block with 1 mL of 2% lidocaine; and 3) MRI and electrophysiological findings consistent with clinical manifestations. The study protocol was approved by the Institutional Review Board of Yeungnam University Hospital, which waived the requirement for written informed consent due to the retrospective nature of the study.

TFESI Procedures

A TFESI was conducted using a standardized method as previously described. All injections were performed by a single interventional physiatrist specializing in spinal injections, under strict aseptic conditions. Patients were positioned prone, and C-arm fluoroscopy was used to identify the target level and guide needle placement. Local anesthesia was achieved with 1% lidocaine at the needle insertion site. A 25G 90-mm spinal needle with a bent tip was positioned between the

lateral vertebral body and the 6 o'clock position below the pedicle. Lateral fluoroscopic imaging confirmed the needle tip location between the spinal laminar margin and the posterior vertebral body. Under anteroposterior fluoroscopy, 0.3 mL of nonionic contrast medium was injected to confirm the absence of vascular uptake and the spread of contrast medium into the foramen. Subsequently, an additional injection of the contrast medium was administered under real-time fluoroscopic monitoring, followed by a 20 mg (0.5 mL) injection of triamcinolone, 0.5 mL of bupivacaine hydrochloride, and one mL of normal saline.

Input Data Used for the Deep Learning Algorithm

Nine variables were used as input features for the deep learning model development: age, gender, side, injection level, pain duration, pretreatment NRS-11, predominant stenotic location, degree of central stenosis, and degree of foraminal stenosis.

LSS severity was assessed by a single physiatrist with more than 15 years of experience treating musculoskeletal disorders. We assessed central stenosis severity using lumbar axial MRI according to the criteria of Lee et al (21): Grade 0 indicates no central stenosis; Grade 1 corresponds to mild stenosis, with each cauda equina nerve root clearly separated; Grade 2 denotes moderate stenosis, characterized by partial aggregation of the cauda equina; And Grade 3 represents severe stenosis, in which the entire cauda equina forms a single bundle.

We assessed foraminal stenosis severity using lumbar sagittal MRI findings, following the classification described by Lee et al (22); Grade 0 indicates no foraminal stenosis; Grade 1 corresponds to mild foraminal stenosis, characterized by perineural fat obliteration surrounding the nerve root in the 2 opposing directions (vertical or transverse), involving contact with either the superior and inferior portions or the anterior and posterior portions of the nerve root, without morphologic alteration; Grade 2 denotes moderate foraminal stenosis, with perineural fat obliteration in all 4 directions, still without morphologic changes in either vertical or transverse directions; And Grade 3 represents severe foraminal stenosis, evidenced by nerve root collapse or notable morphologic changes.

Output Data for the Deep Learning algorithm (Therapeutic Outcome)

Pain severity was assessed using the NRS-11 at pretreatment and at the one-month follow-up post lumbar TFESI. NRS-11 data were obtained through a chart review. A favorable outcome was defined as a ≥ 50% reduction in the NRS-11 score at one month post-TFESI compared with the pretreatment score, whereas a poor outcome was defined as a < 50% reduction. To validate changes in pain severity, the percentage change in NRS-11 scores was calculated as follows:

Change in NRS-11 (%)=(Pretreatment NRS-11 score - one-month post-TFESI NRS-11 score)/(Pretreatment NRS-11 score)×100

Among the study population, 104 patients had a favorable outcome; 84 patients had a poor outcome.

Data Processing and Model Development

The dataset comprising 188 patients was preprocessed and partitioned into training (n = 150; 79.8%) and validation (n = 38; 20.2%) sets. The binary variable outcome served as the classification target. Nine variables were included as input features, comprising 5 numerical features (age, pain duration, pretreatment NRS-11 score, degree of central stenosis, and degree of foraminal stenosis) and 4 categorical features (gender, side, injection level, and predominant stenotic location). The output variable (therapeutic outcome) was treated as a categorical feature.

Three machine learning models were developed to predict therapeutic outcomes following TFESI: deep learning neural network (DNN), an XGBoost classifier, and a CatBoost classifier.

Model-specific encoding strategies were applied. For the DNN, numerical features were standardized to a zero mean and unit variance, and categorical features were one-hot encoded. For XGBoost, categorical features were label-encoded. CatBoost handled categorical features natively, without additional encoding (23). To address class imbalance, the scale_pos_weight parameter, calculated as the ratio of majority to minority class instances, was applied during CatBoost model training (24). The final architectures and optimized hyperparameters for each model are summarized in Table 1.

Statistical Analysis

Model performance was evaluated by calculating the area under the curve (AUC) from the receiver operating characteristic curve analysis. Ninety-five percent CIs for the AUC were estimated using a 5,000-iteration bootstrapping method. All analyses were conducted in Python 3.9, employing scikit-learn 1.4.2 (scikit-learn.

Table 1. *Model architectures and performance metrics for DNN, XGBoost, and CatBoost models.*

ML model	Model structure and performance
DNN	- Five fully connected layers with 1024-512-256-128-64 neurons, learning rate 0.000809, AdamW optimizer, SILU activation, batch size 8, weight decay 0.0001, drop out 0.2. - Training accuracy: 0.933; Validation accuracy: 0.868. - Validation AUC: 0.882 (95% CI, 0.733–0.994).
DNN Validation Performance	Class Precision Recall F1-score Support 0 (poor outcome) 0.833 0.882 0.857 17 1 (favorable outcome) 0.900 0.857 0.878 21 Macro average 0.867 0.870 0.868 38 Weighted average 0.870 0.868 0.869 38
XGBoost	- n_estimators: 100, max_depth: 7; learning_rate: 0.021145067; subsample: 0.5; colsample_bytree: 0.55; colsample_bylevel: 0.75; colsample_bynode: 0.55; gamma: 0.631578947; reg_alpha: 1.473684211; reg_lambda: 9; min_child_weight: 7; - Training accuracy: 0.629; Validation accuracy: 0.763. - Validation AUC: 0.662 with (95% CI, 0.458–0.853).

Table 1 cont. *Model architectures and performance metrics for DNN, XGBoost, and CatBoost models.*

ML model	Model structure and performance
XGBoost Validation Performance	Class Precision Recall F1-score Support 0 (poor outcome) 0.733 0.647 0.688 17 1 (favorable outcome) 0.739 0.810 0.773 21 Macro average 0.736 0.728 0.730 38 Weighted average 0.737 0.737 0.735 38
CatBoost	- Iterations: 20; depth: 3; learning_rate: 0.15; l2_leaf_reg: 3; border_count: 32; bagging_temperature: 0.5; random_strength: 1; scale_pos_weight: 0.96923. - Training accuracy: 0.860; Validation accuracy: 0.789. - Validation AUC: 0.765 (95% CI, 0.593–0.906).

org) for statistical metrics and TensorFlow 2.16 (tensorflow.org) for deep learning development.

Results

Regarding the DNN, the developed DNN model consisted of 5 fully connected layers (1024-512-256-128-64 neurons) employing Swish (SiLU) activation functions. The model was trained using the AdamW optimizer with a learning rate of 8.09×10^{-4} and a batch size of 8. Additionally, minimal weight decay (1.0×10^{-4}) and dropout regularization ($P = 0.2$) were applied to mitigate overfitting. The DNN achieved the highest validation accuracy of 0.868 (training accuracy: 0.933), and the highest validation AUC of 0.882 (95% CI, 0.733–0.994), outperforming both XGBoost and CatBoost (Fig. 1). Classification performance metrics indicated balanced discrimination between poor and favorable outcomes. Precision values were 0.833 and 0.900 for poor and favorable outcomes, respectively, with corresponding recall values of 0.882 and 0.857. This observation resulted in F1-scores of 0.857 and 0.878, with a weighted average F1-score of 0.869, indicating robust overall classification performance. On the validation set, the DNN produced 2 false negatives and 3 false positives (Fig. 2).

Regarding XGBoost, the model—configured with 100 estimators, a maximum depth of 7, and a learning rate of 0.021—exhibited the lowest performance metrics among the 3 algorithms. Additional hyperpa-

rameters included colsample_bytree (0.55), colsample_bylevel (0.75), colsample_bynode (0.55), gamma (0.632), regularization alpha (1.474), lambda (9.0), and a minimum child weight set to 7. The model achieved a validation accuracy of 0.763 (training accuracy: 0.629), suggesting potential underfitting. The validation AUC was 0.662 (95% CI, 0.458–0.853), which was lower than that of the other developed models with the widest CI, indicating greater prediction uncertainty (Fig. 1). Classification metrics demonstrated reduced performance, with precision values of 0.733 and 0.739 for poor and favorable outcomes, respectively, and corresponding recall values of 0.647 and 0.810. The weighted average F1-score was 0.735. On the validation set, XGBoost produced 4 false negatives and 6 false positives (Fig. 2).

The CatBoost model, implemented with 200 iterations, a depth of 3, and a learning rate of 0.15, demonstrated intermediate performance. Additional hyperparameters included L2 leaf regularization (3.0), border count (32), bagging temperature (0.5), random strength (1.0), and scale position weight (0.969). The model achieved a validation accuracy of 0.789 (training accuracy: 0.860), indicating reasonable generalization capability. Additionally, the validation AUC was 0.765 (95% CI. 0.593–0.906), which was intermediate between the DNN and XGBoost models (Fig. 1). Classification metrics revealed precision values of 0.786 and 0.750 for poor and favorable outcomes, respectively, with corresponding recall values of 0.647 and 0.857. The weighted average F1-score was 0.760. On the validation set, CatBoost produced 3 false negatives and 6 false positives (Fig. 2).

Table 1 cont. *Model architectures and performance metrics for DNN, XGBoost, and CatBoost models.*

ML model	Model structure and performance
CatBoost Validation Performance	Class Precision Recall F1-score Support 0 (poor outcome) 0.786 0.647 0.710 17 1 (favorable outcome) 0.750 0.857 0.800 21 Macro average 0.768 0.752 0.755 38 Weighted average 0.766 0.763 0.760 38

Note: Class 0 indicates a poor outcome; Class 1 indicates a favorable outcome.
Abbreviation: ML, machine learning; DNN, deep neural network; XGBoost, extreme gradient boosting; CatBoost, categorical boosting; AdamW, adaptive moment estimation; AUC, area under the curve

Discussion

In this study, we developed and validated machine learning models to predict the therapeutic outcomes of TFESI for patients with LSS using both clinical and MRI-derived variables. Among the 3 algorithms evaluated, the DNN demonstrated the highest predictive performance, achieving a validation AUC of 0.882 and an accuracy of 86.8%, outperforming XGBoost and CatBoost.

The superior performance of the DNN model may be attributed to its ability to capture nonlinear relationships and interactions among multiple clinical and imaging variables. Unlike conventional machine learning algorithms, such as XGBoost and CatBoost (23,24), DNNs can automatically extract hidden patterns from complex datasets, allowing for a more nuanced interpretation of the combined effect of clinical parameters, including age, gender, side, injection level, pain duration, pretreatment NRS-11 scores, and MRI findings, such as predominant stenotic location and stenosis grades (25,26). Overall, the deep learning model appears capable of effectively integrating multimodal inputs to predict TFESI outcomes.

XGBoost and CatBoost were selected as representative conventional machine learning models because of their widespread use in structured clinical data. XGBoost efficiently handles numerical and categorical variables through gradient boosting with strong regularization (23), whereas CatBoost provides optimized handling of categorical features and mitigates overfitting in small datasets (24). Including these models allowed for a direct comparison between decision tree–based learners and the DNN's nonlinear feature learning, confirming the superior performance of deep

learning in integrating complex clinical and MRI features for TFESI outcome predictions.

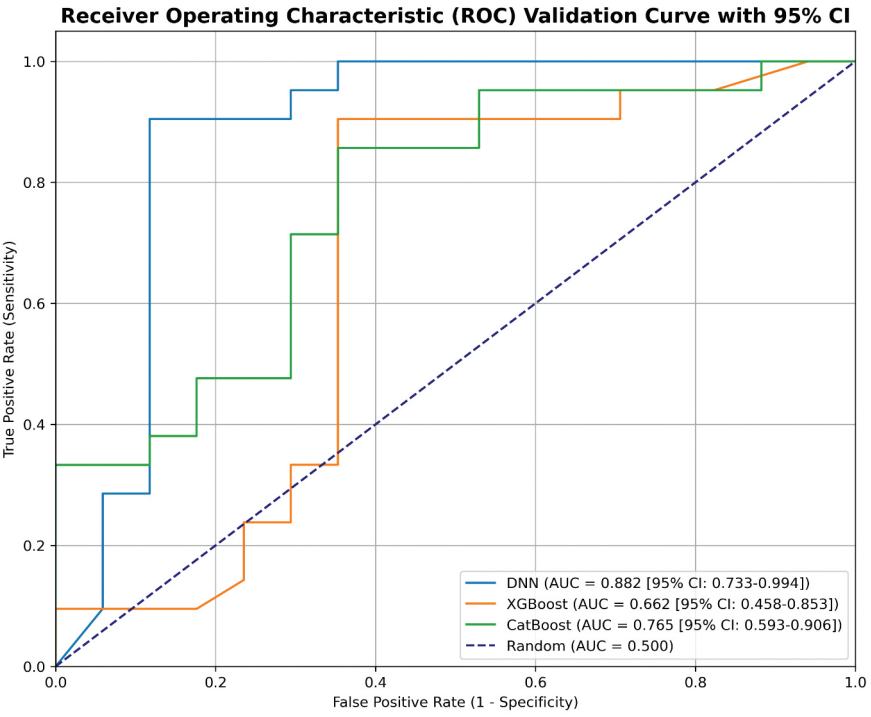


Fig. 1. *Receiver operating characteristic curves for data validation. The deep learning neural network model achieved the highest performance with an area under the curve (AUC) of 0.882, followed by CatBoost with an AUC of 0.765, and XGBoost with an AUC of 0.662.*

From a clinical perspective, early identification of patients with a poor expected response to a TFESI could enable clinicians to individualize treatment planning, minimize unnecessary procedures, and facilitate a timely transition to alternative therapies, such as surgical decompression or regenerative interventions. Given the modest accuracy of traditional regression-based models reported in prior studies, our findings underscore the potential utility of DNN-based approaches for real-world decision-making for patients with lumbosacral radiculopathy due to LSS (9,10).

To date, only 2 previous studies have applied DNNs to predict TFESI therapeutic outcomes in patients with LSS (27,28). In 2023, Wang et al (27) used MRI data; however, their study was limited to cases of foraminal stenosis,

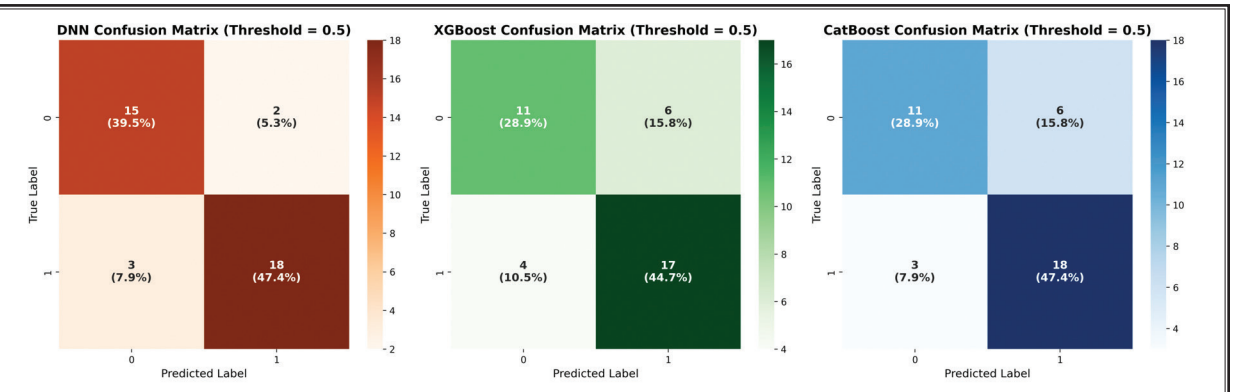


Fig. 2. *Confusion matrices illustrating the classification performance of the deep neural network, XGBoost, and CatBoost models in the validation dataset. Each matrix displays the number of true positives, true negatives, false positives, and false negatives for predicting therapeutic outcomes post transforaminal epidural steroid injection. A favorable outcome was defined as a ≥ 50% reduction in the Numeric Rating Scale pain score at one month postinjection, whereas a poor outcome was defined as a < 50% reduction. Label 0 indicates a poor outcome and Label 1 indicates a favorable outcome. The deep learning neural network model had fewer misclassifications compared with the other models, indicating superior discriminative performance.*

achieving a predictive performance (AUC = 0.827) that was lower than that of our model (27). In 2024, 3 of the same authors (28) used anteroposterior, lateral, and bilateral oblique lumbar spine radiographs as input data, achieving an AUC of 0.920 and an accuracy of 87.2%. Unlike our study, their model analyzed radiographs rather than MRI data and did not incorporate detailed clinical parameters. The high accuracy achieved using only radiographs suggests that lumbar spine radiographs contain substantial predictive information relevant to TFESI outcomes.

Our study has some limitations. First, its retrospective and single-center design may limit generalizability. Second, the sample size, particularly in the validation cohort, was relatively small, increasing the risk of model overfitting. Third, although MRI grading systems for central and foraminal stenosis were standardized, inter-rater variability in grading was not analyzed. Fourth, despite including multiple clinical and imaging variables, important confounding factors—such as prior treatments, psychosocial factors, and medical comorbidities—were not fully accounted for due to the retrospective design. Fifth, outcomes were assessed solely based on pain reduction; we did not evaluate functional status or quality of life. Lastly, heterogeneity in disease characteristics and injection levels, combined with the relatively small sample size, precluded meaningful subgroup analyses. Future studies using larger, prospective, multicenter cohorts with comprehensive clinical, functional, and psychosocial data are needed to address these limitations and to enable more robust subgroup analyses and external validation.

In conclusion, our findings demonstrate that deep learning can effectively predict the therapeutic response to a TFESI in patients with LSS using combined clinical and MRI features. The DNN model demonstrated superior discriminative ability compared with conventional machine learning algorithms, suggesting that deep learning-based prediction models may serve as valuable tools for personalized pain management and treatment stratification for patients with LSS.

References

1. Abdou A, Kades S, Masri-Zada T, Asim S, Bany-Mohammed M, Agrawal DK. Lumbar spinal stenosis: Pathophysiology, biomechanics, and innovations in diagnosis and management. *J Spine Res Surg* 2025; 7:1-17.
2. Chang MC, Lee DG. Outcome of transforaminal epidural steroid injection according to the severity of lumbar foraminal spinal stenosis. *Pain Physician* 2018; 21:67-72.
3. Munakomi S, Cruz R. Lumbar spinal stenosis. *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2025. www.ncbi.nlm.nih.gov/books/NBK531493/
4. Katz JN, Zimmerman ZE, Mass H, Makhni MC. Diagnosis and management of lumbar spinal stenosis: A review. *JAMA* 2022; 327:1688-1699.
5. Malik KN, Giberson C, Ballard M, Camp N, Chan J. Pain management interventions in lumbar spinal stenosis: A literature review. *Cureus* 2023; 15:e44116.
6. Webb CW, Aguirre K, Seidenberg PH. Lumbar spinal stenosis: Diagnosis and management. *Am Fam Physician* 2024; 109:350-359.
7. Davis N, Hourigan P, Clarke A. Transforaminal epidural steroid injection in lumbar spinal stenosis: An observational study with two-year follow-up. *Br J Neurosurg* 2017; 31:205-208.
8. Liu H, Ji M, Xiao P, et al. Glucocorticoids-based prodrug design: Current strategies and research progress. *Asian J Pharm Sci* 2024; 19:100922.
9. Fromm MK, Liu SH, Carr C, Stefancic E, Rogalski M. Factors associated with improved outcomes after lumbar transforaminal epidural steroid injections for radicular pain: A systematic review. *N Am Spine Soc J* 2025; 21:100592.
10. McCormick Z, Cushman D, Casey E, Garvan C, Kennedy DJ, Plastaras C. Factors associated with pain reduction after transforaminal epidural steroid injection for lumbosacral radicular pain. *Arch Phys Med Rehabil* 2014; 95:2350-2356.
11. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021; 8:53.
12. Choudhary K, DeCost B, Chen C. et al. Recent advances and applications of deep learning methods in materials science. *npj Comput Mater* 2022; 8:59.
13. Mienye ID, Swart TG. A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information* 2024; 15:755.
14. Choo YJ, Chang MC. Use of machine learning in stroke rehabilitation: A narrative review. *Brain Neurorehabil* 2022; 15:e26.
15. Sarker IH. Machine learning: Algorithms, real-world applications and research directions. *SN Comput Sci* 2021; 2:160.
16. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol* 2020; 9:14.
17. Wang MX, Kim JK, Choi JW, Park D, Chang MC. Deep learning algorithm for automatically measuring Cobb angle in patients with idiopathic scoliosis. *Eur Spine J* 2024; 33:4155-4163.
18. Choo YJ, Kim JK, Kim JH, Chang MC, Park D. Machine learning analysis to predict the need for ankle foot orthosis in patients with stroke. *Sci Rep* 2021; 11:8499.
19. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke* 2019; 50:1263-1265.
20. Kim JK, Choo YJ, Chang MC. Prediction of motor function in stroke patients

using machine learning algorithm: Development of practical models. *J Stroke Cerebrovasc Dis* 2021; 30:105856.

21. Lee GY, Lee JW, Choi HS, Oh KJ, Kang HS. A new grading system of lumbar central canal stenosis on MRI: An easy and reliable method. *Skeletal Radiol* 2011; 40:1033-1039.
22. 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.
23. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016; 785-794.
24. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems 2018; 31:6639-6649.
25. Islam MT, Zhou Z, Ren H, et al. Revealing hidden patterns in deep neural network feature space continuum via manifold learning. *Nat Commun* 2023; 14:8506.
26. Wen P, Zhang Y, Wen G. Learning features from irrelevant domains through deep neural network. *Complex Intell Syst* 2024; 10:627-638.
27. Wang MX, Kim JK, Chang MC. Deep learning algorithm trained on cervical magnetic resonance imaging to predict outcomes of transforaminal epidural steroid injection for radicular pain from cervical foraminal stenosis. *J Pain Res* 2023; 16:2587-2594.
28. Wang MX, Kim JK, Kim CR, Chang MC. Deep learning algorithm trained on oblique cervical radiographs to predict outcomes of transforaminal epidural steroid injection for pain from cervical foraminal stenosis. *Pain Ther* 2024; 13:173-183.