scientific reports

Check for updates

OPEN

A stacking ensemble machine learning model for predicting postoperative axial pain intensity in patients with degenerative cervical myelopathy

Xu Chu[1,4], Jiajun Song[2,4], Jiandong Wang[3] & Hui Kang[1]✉

Machine learning (ML) has been extensively utilized to predict complications associated with various diseases. This study aimed to develop ML-based classifiers employing a stacking ensemble strategy to forecast the intensity of postoperative axial pain (PAP) in patients diagnosed with degenerative cervical myelopathy (DCM). A total of 711 consecutive postoperative DCM patients were included between 2016 and 2024, and after excluding patients who did not meet the inclusion criteria and those who met the exclusion criteria, a total of 484 patients were ultimately included in this study. The intensity of PAP was assessed using a standardized Numerical Rating Scale (NRS) score one year following surgery. Participants were randomly allocated into training and testing sub-datasets in a ratio of 8:2. 91 initial ML classifiers were developed, from which the top three highest-performing classifiers were subsequently integrated into an ensemble model utilizing 13 different machine learning models. The area under the curve (AUC) served as the primary metric for evaluating the predictive performance of all classifiers. The classifiers EmbeddingLR-RF (AUC = 0.81), EmbeddingRF-MLP (AUC = 0.81), and RFE-SVM (AUC = 0.80) were recognized as the leading three models. By implementing an ensemble learning approach such as stacking, an enhancement in performance for the ML classifier was observed after amalgamating these three models, with SVM ensemble classifier performed the best (AUC = 0.91). Decision curve analysis underscored the advantages conferred by these ensemble classifiers; notably, prediction curves for PAP intensity among DCM patients exhibited significant variability across the top three initial classifiers. The ensemble classifiers effectively predicted PAP intensity in DCM patients, showcasing substantial potential to aid clinicians in managing DCM cases while optimizing medical resource utilization.

Keywords Degenerative cervical myelopathy, Postoperative axial pain, Machine learning, Stacking, Ensemble learning

Abbreviations

ML	Machine learning
PAP	Postoperative axial pain
DCM	Degenerative cervical myelopathy
NRS	Numerical rating scale
SVM	Support vector machine
AUC	Area under the curve
ROM	Range of motion
SVA	Sagittal vertical axis
HADS	Hospital Anxiety and Depression Scale
HADS-A	Hospital Anxiety and Depression Scale-Anxiety

[1]Department of Shoulder and Elbow of Sports Medicine, Honghui Hospital, Xi'an Jiaotong University, Xi'an 710054, China. [2]Department of Orthopedics Surgery, Tianjin Medical University General Hospital, Tianjin 300052, China. [3]Hefei Metrology and Testing Center, Hefei 230088, Anhui, China. [4]These authors contributed equally: Xu Chu and Jiajun Song. ✉email: drkanghui@163.com

HADS-D	Hospital Anxiety and Depression Scale-Depression
SF-36	Short Form-36 survey
JOA	Japanese Orthopedic Association
JOARR	Japanese Orthopedic Association recovery rate
nPAP	Non-postoperative axial pain
RFE	Recursive feature elimination
MIC	Maximal information coefficient
mRMR	Minimal-redundancy-maximal-relevance
LR	Logistic regression
RF	Random forest
LSVC	Linear support vector classifier
LDA	Linear discriminant analysis
AdaBoost	Adaptive boosting
DNN	Deep neural network
MLP	Multilayer perceptron
NB	Naïve Bayes
KNN	K-Nearest Neighbor
DT	Decision tree
MICE	Multiple imputation by chained equations

Degenerative cervical myelopathy (DCM) is a prevalent condition in clinical practice, characterized by the acquired narrowing of the spinal canal, which leads to non-traumatic spinal cord injury[1]. The primary treatment for DCM involves spinal canal decompression surgery[2]. While there remains ongoing debate regarding the optimal surgical approach, posterior laminoplasty and laminectomy continue to be the principal methods employed for effective spinal cord decompression[3]. However, despite the success of these procedures in alleviating cord compression, a substantial number of patients experience postoperative axial pain (PAP), a complication that often results in significant discomfort and hinders recovery[4]. Research has indicated that PAP may be influenced by factors such as anatomical abnormalities, neurological function, and preoperative pain intensity[5–8]. Despite this understanding, there exists a notable deficiency in effective predictive models capable of identifying patients at elevated risk for developing PAP post-surgery. Given the impact of PAP on patient outcomes, it is imperative to further explore the association between these factors and PAP intensity. Such investigations could facilitate more personalized surgical approaches and postoperative care strategies aimed at mitigating the incidence of PAP.

In this context, machine learning (ML) algorithms have emerged as powerful tools for analyzing large and complex datasets, rendering them particularly well-suited for developing predictive models within clinical settings. The capacity of ML to manage extensive data volumes while uncovering nonlinear relationships between predictors and outcomes has garnered considerable attention[9]. As ML techniques have advanced over time, researchers have increasingly applied them to predict various types of pain—including PAP—with promising results[10–12]. Unlike traditional statistical models that may struggle with intricate high-dimensional data sets, ML models possess an inherent ability to identify subtle patterns that can significantly enhance predictive accuracy[13]. This capability is especially valuable in guiding clinicians toward tailored perioperative management strategies based on individualized risk predictions.

Although numerous ML-based prediction models have been developed across different types of pain conditions, most studies primarily focus on comparing individual model predictive accuracies with an aim to identify a single best-performing model[12,14]. However, recent advancements in ML have introduced ensemble learning—a more sophisticated methodology that amalgamates predictions from multiple models into a cohesive system designed to yield stronger and more reliable predictions[15]. Ensemble learning presents several key advantages: improved accuracy; reduced risk of overfitting; greater robustness; and enhanced stability[16]. By integrating diverse model strengths through ensemble learning techniques holds promise for surpassing individual model performance while providing more accurate predictions applicable within clinical contexts.

Consequently, this study's objective was to develop an ensemble learning-based predictive model capable of analyzing clinical data to elucidate key relationships while predicting PAP intensity among DCM patients. This approach seeks not only to enhance prediction accuracy but also provide critical insights aimed at optimizing perioperative management practices ultimately improving patient outcomes.

Materials and methods

Patient cohort

The data for this study was retrospectively gathered from the Orthopedic Department of Honghui Hospital from 2016 to 2024, encompassing a total of 711 consecutive patients who underwent posterior cervical decompression for symptomatic degenerative cervical myelopathy (DCM). This study was approved by the institutional ethics board of Honghui Hospital and conducted in compliance with recognized ethical standards. All methods were performed in accordance with the relevant guidelines and regulations. All patients provided written informed consent prior to each procedure. The inclusion criteria were as follows: (1) evidence of myelopathy along the cervical spine (C3–C7) on cervical MRI; (2) associated symptoms and signs of myelopathy, including sensory and motor deficits, bladder/bowel dysfunction, and gait disturbances; (3) willingness to undergo posterior decompression surgery (e.g., laminoplasty); and (4) age of 18 years or older. Exclusion criteria included: (1) history of previous cervical surgery; (2) stenosis of the extracranial vertebral or carotid arteries as determined by Doppler ultrasound; (3) any indications of other neurological, psychiatric, ophthalmic, or systemic conditions such as hypertension or diabetes; (4) a history of alcohol or substance abuse; (5) lost to follow up.

Baseline data

Machine learning models were trained utilizing a variety of potential preoperative predictors, including patient demographics, baseline axial neck pain intensity, functional status, mental health assessment, and other pertinent radiological and surgery-related factors.

The radiographic parameters measured on preoperative neutral standing lateral X-rays included the following: (1) the anteroposterior diameter of the spinal canal at C5 as determined from plain preoperative lateral radiographs[7]; (2) the C2–C7 lordosis angle, defined as the angle between the lower endplate of C2 and the upper endplate of C7[12]; (3) range of motion (ROM) for C2–C7 assessed through flexion-extension lateral radiographs using the C2–C7 lordosis angle[12]; (4) sagittal vertical axis (SVA) for C2–C7, representing the distance between the plumb line at C2 and the posterior superior endplate of C7, with positive sagittal alignment characterized by anterior deviation[11]; (5) T1 slope measured as the angle formed between a horizontal line and the upper endplate of T1[12]; and (6) definitions for anterolisthesis and retrolisthesis as anterior or posterior vertebral body slippage exceeding 2 mm in a neutral position[12].

Preoperative neck pain intensity was evaluated using a standardized Numerical Rating Scale (NRS), which ranges from 0 (no pain) to 10 (worst imaginable pain). Patients were instructed to rate their average axial neck pain intensity experienced over the preceding month. At one-year follow-up conducted via telephone interview, patients utilized the same NRS scale to reassess their average neck pain intensity over that period[17]. All patients underwent evaluation for anxiety and depression within one week prior to surgery using a Chinese version of Hospital Anxiety and Depression Scale (HADS), comprising 14 items—seven assessing anxiety symptoms (HADS-A) and seven evaluating depressive symptoms (HADS-D)—with scores ranging from 0 to 3[18]. Generic health-related quality of life was measured using Short Form-36 survey v2.0 (SF-36). Neurological function was assessed both preoperatively and one year postoperatively by a senior spine surgeon employing JOA scoring system, with recovery rates calculated according to Hirabayashi's formula[19]:

$$\text{JOA recovery rate (JOARR)} = \frac{\text{postoperative JOA score} - \text{preoperative JOA score}}{(17 - \text{preoperative JOA score})}$$

Surgical techniques

Under general anesthesia, patients were positioned prone, and a standard posterior midline exposure was performed for all procedures. In the laminoplasty group, the supraspinous ligaments were preserved, and the open side was determined based on symptomatic or severely compressed areas, or at the surgeon's discretion when unclear. Using a high-speed burr, a full-thickness trough was drilled at the junction of the lateral mass and lamina on the open side, and a partial-thickness trough was created on the hinge side. The lamina was then elevated toward the hinge side and stabilized with miniplates and screws. For the laminectomy and laminectomy with fusion groups, complete removal of the lamina and ligamentum flavum was performed at the target levels. In the fusion group, lateral mass screws and rods were placed, and autologous bone from the laminectomy was grafted onto the lateral mass.

Predicted outcomes

The primary outcome of this study was the incidence of moderate-to-severe axial neck pain at the one-year follow-up, operationally defined as a Numerical Rating Scale (NRS) score of 4 or higher[12]. Patients were subsequently categorized into two groups: the postoperative axial pain (PAP) group, which included individuals experiencing moderate-to-severe pain (NRS ≥ 4), and the non-postoperative axial pain (nPAP) group, comprising those reporting mild or no pain (NRS < 4)[12].

Model development

The analysis workflow is illustrated in Fig. 1. After excluding patients who did not meet the inclusion criteria (36 patients) and those who met the exclusion criteria (191 patients, of which 105 were lost to follow-up), a total of 484 patients were ultimately included in this study. To identify the most relevant features, we employed seven widely recognized feature-selection techniques: recursive feature elimination (RFE), maximal information coefficient (MIC), minimal-redundancy-maximal-relevance (mRMR), embedding logistic regression (embedding LR), embedding random forest (RF), embedding tree, and embedding linear support vector classifier (embedding LSVC). Subsequently, thirteen machine learning (ML) algorithms were applied, including linear discriminant analysis (LDA), random forest (RF), gradient boosting, adaptive boosting (AdaBoost), deep neural network (DNN), multilayer perceptron (MLP), bagging, support vector machine (SVM), Gaussian naïve Bayes (NB), extra trees, K-Nearest Neighbor (KNN), logistic regression (LR), and decision tree (DT)[15,20]. Given that many of these algorithms do not natively provide probability estimates, Platt scaling was utilized to transform raw model outputs into probabilities for enhanced interpretability. All input features were standardized using z-score normalization to ensure consistency across variables. Categorical variables were encoded with one-hot encoding. The dataset was then split into training and testing sets using stratified sampling to maintain the outcome distribution in both sets.

From the combination of these seven feature-selection methods and thirteen ML algorithms, a total of 91 classifiers were generated. The dataset was partitioned into training, and testing sets in a ratio of 8:2. A single iteration of 10-fold cross-validation was performed, followed by independent testing on a reserved test set. The cross-validation procedure involved several steps: (1) Data Partitioning: The training dataset was partitioned into ten subsets of roughly equal size; (2) Training and Validation Cycles: For each iteration, nine subsets were used for training while one subset served as the validation set—ensuring that each subset functioned as the validation set exactly once; (3) Performance Aggregation: Classifier performance was averaged over all ten

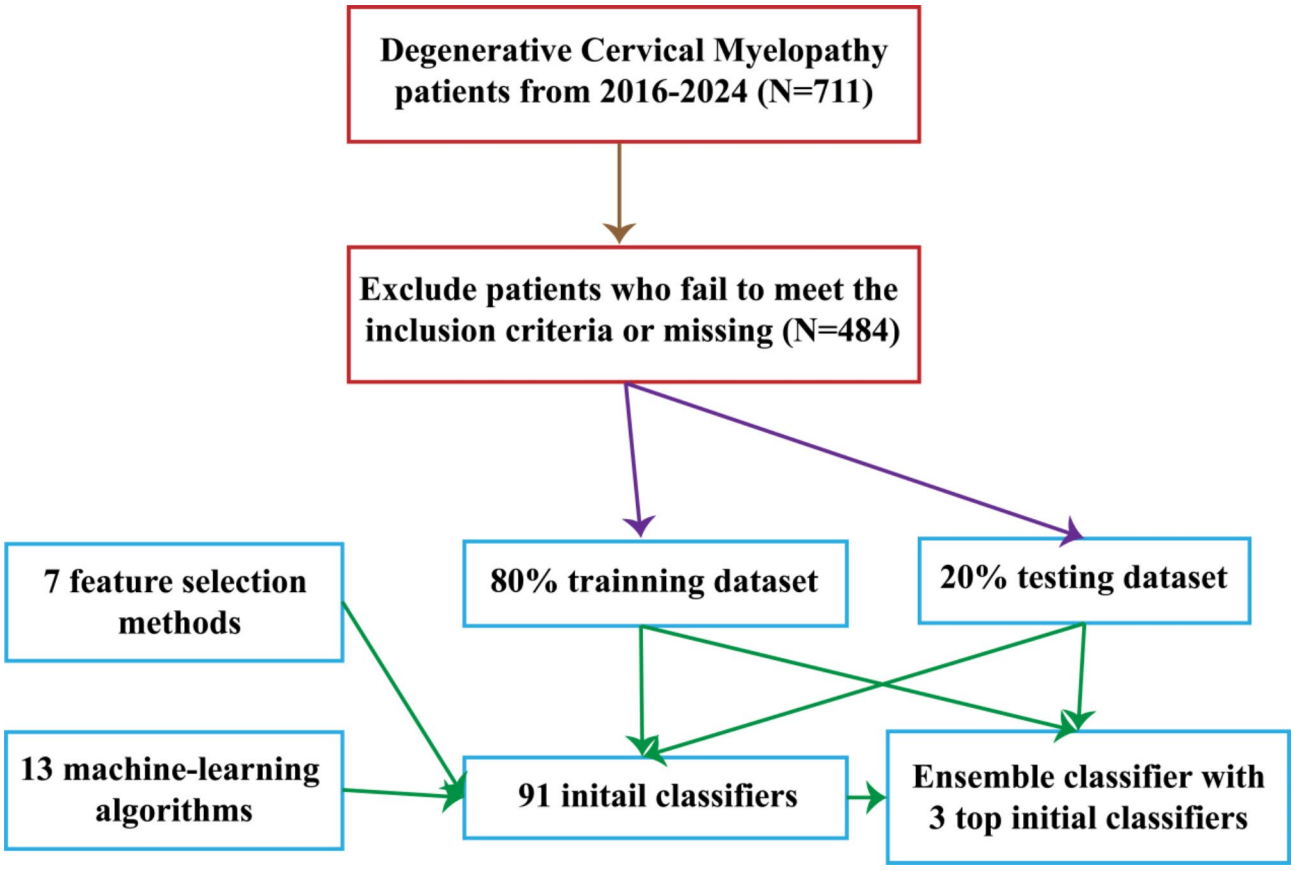

Fig. 1. Flowchart of the analysis pipeline for this study.

folds to mitigate bias from any single data split; (4) Independent Testing: After cross-validation completion, final model evaluation occurred on a distinct test set excluded from both training and validation stages. This comprehensive approach involving repeated 10-fold cross-validation provided an extensive assessment of classifier performance. Additionally, model hyperparameters were optimized during cross-validation through grid search methodology with parameter ranges detailed in Supplementary Table 1.

The models' ability to distinguish between patients with and without the target condition was evaluated using various evaluation parameters, including area under the curve (AUC), accuracy, sensitivity, specificity, and Brier Score. Among these metrics, AUC was selected as the primary measure for evaluating the models. Based on the AUC scores obtained during cross-validation, we identified the top three classifiers, which were then combined into an ensemble model. The ensemble model employed a stacked approach, where a meta-classifier synthesized predictions from the top three base classifiers to evaluate performance. Initially, each base classifier generated individual predictions on both the validation and testing datasets. Subsequently, these validation predictions were combined to create a new dataset, which was then used to train the meta-classifier with 13 different machine learning models. The meta-classifier was developed to enhance predictions by integrating the outputs of the three highest-performing classifiers. Finally, we assessed the performance of the stacked models on the testing set, providing a robust measure of its predictive power.

Results

Table 1 provides a summary of the clinical characteristics of the patients included in this study. The average age was 63.2 years, with 35.1% being female. The mean axial pain intensity score decreased from 4.74 ± 2.1 preoperatively to 3.65 ± 1.8 postoperatively and further reduced to 2.36 ± 1.2 at the 1-year follow-up. At the 1-year follow-up, 169 patients (34.9%) reported moderate to severe axial neck pain, while 315 (65.1%) experienced mild or no pain. The mean baseline HADS-D score was 8.3 ± 1.9. A total of 26 potential features were selected and analyzed in developing the predictive machine learning models.

Performance results of all classifiers in predicting PAP in DCM patients are shown in Fig. 2A–B, and the sensitivity, specificity, and AUC for the top three individual classifiers in cross-validation and independent testing are provided in Table 2. The classifiers EmbeddingLR-RF, EmbeddingRF-MLP, and RFE-SVM were recognized as the leading three models, each attaining the highest AUC values. In the testing dataset, AUCs for these models were 0.81 for both EmbeddingLR-RF and EmbeddingRF-MLP, and 0.80 for RFE-SVM (Fig. 2C). Independent testing demonstrated that the SVM ensemble classifier performed the best in predicting PAP in

Variable	Mean ± standard deviation or frequency (proportion)	Range (minimum-maximum)
Patient characteristics		
Age (year)	63.2 ± 11.5	36–84
Sex (Female)	170 (35.1)	N/A
BMI	22.5 ± 3.4	16.6–31.8
Current smoking	271 (56)	N/A
Duration of symptom (mo)	18.2 ± 12.5	0–60
Baseline symptoms		
Axial pain intensity	4.74 ± 2.1	0–9
Pre-JOA score	8.5 ± 1.8	5–13
HADS-D	8.3 ± 1.9	3–14
HADS-A	6.8 ± 2.2	4–12
SF-36 PCS	23.6 ± 12.2	15.5–44.3
SF-36 MCS	51.0 ± 10.9	28.5–62.1
Baseline radiological assessment		
A–P canal diameter (mm)	12.3 ± 1.9	6.8–14.6
C2–C7 Cobb angle (°)	13.5 ± 8.9	8.6–33.5
C2–C7 ROM (°)	34.0 ± 12.8	18.6–54.0
C2–C7 SVA (mm)	20.6 ± 11.1	− 5.3-45.2
T1 slope (°)	28.5 ± 15.3	15.8–42.5
Anterolisthesis	33 (6.8)	N/A
Retrolisthesis	21 (4.3)	N/A
Surgical-related factors		
Operation time (min)	164.5 ± 25.6	142–218
Blood loss (ml)	284 ± 35	215–420
No. of laminoplasty levels	2.8 ± 1.3	1–5
Laminoplasty	176 (36.4)	N/A
Laminectomy	233 (48.1)	N/A
Laminectomy + fusion	75 (15.5)	N/A
Postoperative parameters		
Post-JOA score	12.7 ± 2.5	9–16
Axial pain intensity	3.65 ± 1.8	0–7
1-year follow up		
Axial pain intensity	2.36 ± 1.2	0–6

Table 1. Clinical features and outcomes of the cohort of 484 DCM patients. Continuous variables are presented as Mean ± Standard Deviation, while categorical variables are expressed as frequency and percentage (%). *DCM* Degenerative cervical myelopathy, *BMI* Body Mass Index, *Pre* Preoperative, *Post* Postoperative, *HADS-D* depression subscale of hospital anxiety and depression scale, *HADS-A* anxiety subscale of hospital anxiety and depression scale, *A–P* anterior–posterior, *ROM* range of motion, *SVA* sagittal vertical axis, *JOA* Japanese Orthopedic Association.

DCM patients, outperforming the initial classifiers, with a superior AUC of 0.91 (Fig. 3A; Table 3), while the initial classifiers had AUCs of 0.80, 0.86, and 0.79, respectively (Table 4).

Decision curve analysis emphasized the advantages of the ensemble classifier, revealing notable distinctions compared to the curves of the top three classifiers in predicting PAP in DCM patients (Fig. 3B). Feature importance was evaluated using permutation importance, which identified the top ten predictors for the ensemble model. The five most influential features were "preoperative axial pain intensity," "JOARR," "preoperative C2-7 Cobb angle," "HADS-D," and "age" (Fig. 3C).

Discussion

This study identified three key findings: (1) Classifier performance was significantly improved by applying an ensemble learning approach that integrated three commonly used ML models—EmbeddingLR-RF, EmbeddingRF-MLP, and RFE-SVM; (2) The implementation of an ensemble learning approach, such as stacking, led to an improvement in the performance of the machine learning classifier, with the SVM ensemble classifier achieving the best results; (3) Finally, preoperative axial pain intensity, JOARR, preoperative C2-7 Cobb angle, HADS-D, and age were identified as the five most important clinical predictors of postoperative axial pain following posterior cervical decompression surgery in DCM patients.

Determining the optimal NRS cut-off between mild and moderate pain is essential for identifying patients who require pain management (moderate-to-severe pain) versus those who do not (mild pain)[21]. However, the

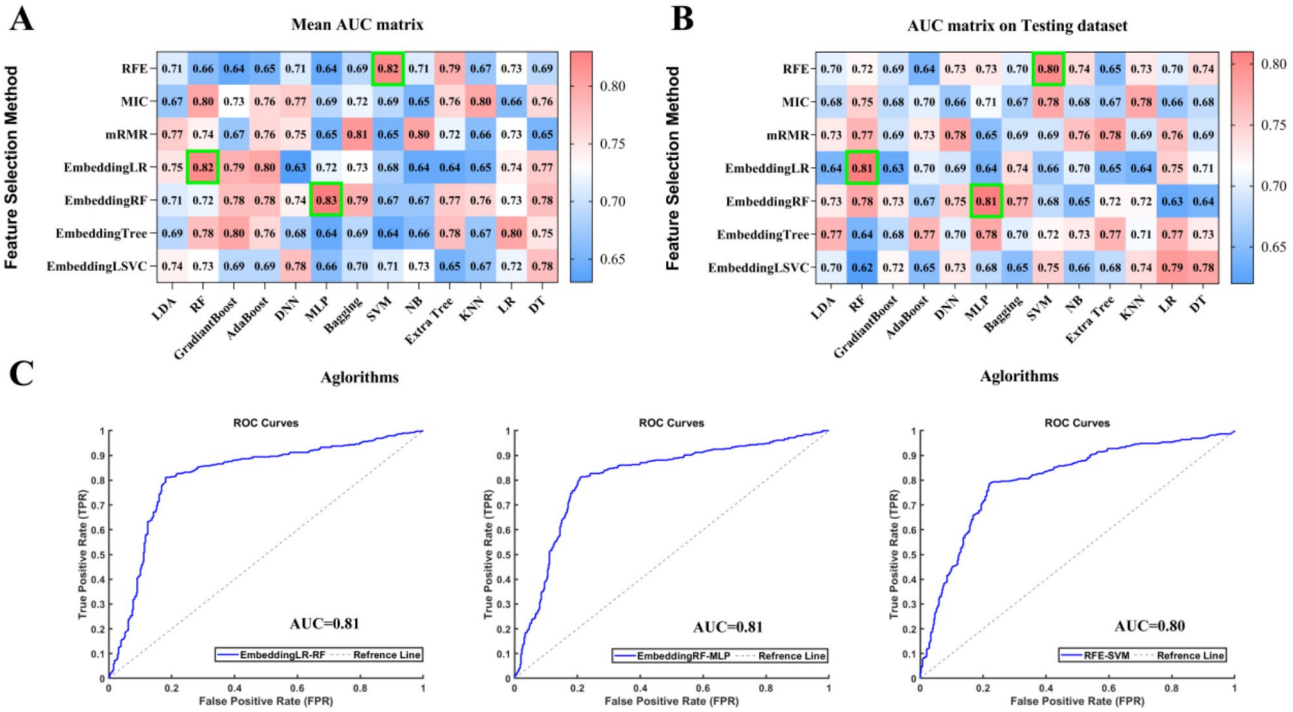

Fig. 2. Prediction performance of postoperative axial pain in patients with degenerative cervical myelopathy. (**A**) AUC values of all initial classifiers during cross-validation; (**B**) AUC values of all initial classifiers during independent testing; (**C**) ROC curves of the top 3 classifiers with the highest predictive performance.

Model	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	Brier score
Cross-validation					
EmbeddingLR-RF	0.82	79.59	0.82	0.77	0.18
EmbeddingRF-MLP	0.83	82.69	0.83	0.83	0.18
RFE-SVM	0.82	78.42	0.77	0.8	0.18
Independent testing					
EmbeddingLR-RF	0.81	79.97	0.77	0.83	0.18
EmbeddingRF-MLP	0.81	80.49	0.80	0.81	0.18
RFE-SVM	0.80	76.10	0.78	0.74	0.19

Table 2. Comparison of the AUCs, accuracies, sensitivities, specificities and BrierScores of the top three individual classifiers in both cross-validation and independent testing. *EmbeddingLR* embedding logistic regressor, *RF* random forest, *EmbeddingRF* embedding random forest, *MLP* multilayer perceptron, *RFE* recursive feature elimination, *SVM* supported vector machine.

exact threshold remains debated. Current literature suggests varying cut-off points for moderate-to-severe pain, with NRS thresholds ranging from 3 to 6, depending on diagnostic criteria and analytical methods[21]. Recent studies show a tendency to set the threshold at 4 for identifying postoperative patients needing pain treatment[22]. Accordingly, this study defines postoperative axial pain as an NRS score of 4 or higher.

In this study, we converted a continuous variable (e.g., postoperative axial pain) into a binary one in the machine learning classifiers for three primary reasons[23]. First, with a limited sample size, binary variables simplify model construction by reducing data complexity. Second, this transformation minimizes the impact of outliers, enhancing predictive reliability. Finally, binary variables are less prone to noise and measurement errors, thereby improving the model's robustness and stability—qualities critical for clinical applications.

Early identification of patients with PAP enables clinicians to develop new perioperative management strategies and early interventions to reduce postoperative pain incidence. Over recent decades, efforts have focused on creating clinical prediction models to forecast PAP severity in DCM patients undergoing posterior cervical decompression surgery[24–26]. An accurate prediction model would help spine surgeons identify individuals at higher risk of severe pain postoperatively, facilitating the creation of personalized treatment plans based on each patient's risk profile. However, current predictive models for PAP following this procedure remain limited. Prior studies have offered some insights. Kimura et al. used logistic regression to identify predictors

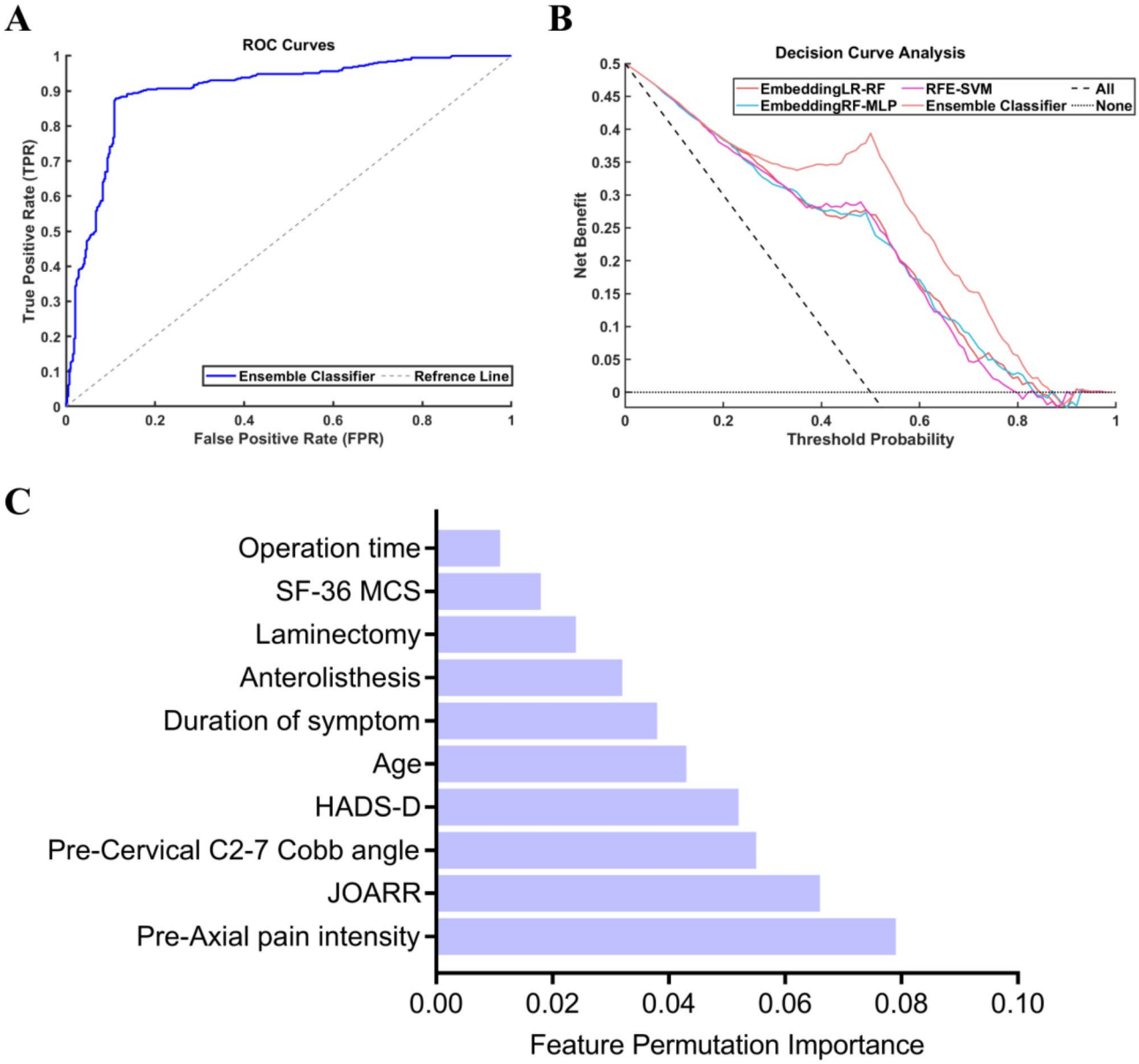

Fig. 3. Decision curve analysis and feature permutation importance. (**A**) ROC curve of the ensemble classifier, (**B**) Decision curve analysis (DCA) for predicting postoperative axial pain in patients with degenerative cervical myelopathy; (**C**) Top 10 features contributing to postoperative axial pain prediction in the ensemble classifier. *Pre* preoperative, *JOARR* Japanese Orthopedic Association recovery rate, *HADS-D* Hospital Anxiety and Depression Scale—Depression, *SF-36* Short form-36 survey.

of PAP, including anterolisthesis, smoking, moderate-to-severe baseline neck pain, and lower SF-36 Mental Component Summary scores[12]. Ionse et al. found that preoperative and postoperative cervical lordosis in extension independently predicted postoperative neck pain[11]. Additionally, Cao et al. used logistic regression on 144 patients to predict PAP after cervical decompression surgery, achieving an AUC of 0.78, with a sensitivity of 0.77 and specificity of 0.65, and identified preoperative C2-C7 Cobb angle as an independent PAP risk factor[27]. This study systematically examined and compared various widely used machine learning algorithms to identify the most effective predictive models for PAP in DCM patients. Future research should build on this work, focusing on model refinement and optimization to enhance predictive accuracy and clinical relevance.

In this study, the three top-performing classifiers, selected based on AUC, were integrated into an ensemble model utilizing an SVM classifier. Ensemble learning provides distinct advantages over individual machine learning models by harnessing the collective strengths of multiple algorithms. By combining diverse models, ensemble methods reduce the risk of overfitting and improve generalization, as different algorithms capture unique patterns within the data. This diversity allows the ensemble to compensate for the weaknesses of any single model, resulting in more robust predictions. Additionally, ensemble techniques enhance model stability by aggregating multiple outputs, thereby minimizing the influence of performance variability. Another

Model	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	Brier score
Stacking-LDA	0.82	79.65	0.80	0.80	0.18
Stacking-RF	0.86	81.40	0.84	0.79	0.17
Stacking-GradiantBoost	0.83	83.37	0.82	0.85	0.17
Stacking-AdaBoost	0.88	86.98	0.89	0.85	0.16
Stacking-DNN	0.86	83.68	0.85	0.82	0.17
Stacking-MLP	0.89	86.78	0.86	0.87	0.16
Stacking-Bagging	0.87	84.50	0.85	0.84	0.16
Stacking-SVM	0.91	88.76	0.92	0.85	0.15
Stacking-NB	0.81	78.51	0.79	0.78	0.19
Stacking-Extra-Tree	0.85	81.92	0.80	0.84	0.17
Stacking-KNN	0.87	84.50	0.87	0.82	0.16
Stacking-LR	0.88	87.09	0.87	0.87	0.16
Stacking-DT	0.83	81.30	0.79	0.84	0.18

Table 3. Model performance for stacking-learning of top 3 machine learning models in independent testing set. *LDA* linear discriminant analysis, *RF* random forest, *AdaBoost* adaptive boosting, *DNN* deep neural network, *MLP* multilayer perceptron, *SVM* support vector machine, *NB* naïve Bayes, *KNN* K-Nearest Neighbor, *LR* logistic regression, *DT* decision tree.

Model	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	Brier score
Cross-validation					
EmbeddingLR-RF	0.78	75.58	0.77	0.74	0.20
EmbeddingRF-MLP	0.88	89.02	0.89	0.89	0.16
RFE-SVM	0.81	79.59	0.80	0.80	0.18
SVM Ensemble Classifier	0.89	87.47	0.87	0.88	0.16
Independent testing					
EmbeddingLR-RF	0.80	77.91	0.75	0.80	0.19
EmbeddingRF-MLP	0.86	82.43	0.84	0.80	0.17
RFE-SVM	0.79	78.81	0.78	0.79	0.19
SVM Ensemble Classifier	0.91	88.76	0.92	0.85	0.15

Table 4. The AUCs, accuracies, sensitivities, specificities and BrierScores of each individual classifier in the ensemble model in both cross-validation and independent testing. *EmbeddingLR* embedding logistic regressor, *RF* random forest, *EmbeddingRF* embedding random forest, *MLP* multilayer perceptron, *RFE* recursive feature elimination, *SVM* supported vector machine.

critical benefit is their ability to explore the solution space more comprehensively, helping prevent models from becoming trapped in local minima or overly influenced by specific data patterns. The synergistic effect of ensemble learning ultimately yields more accurate and reliable predictive outcomes, especially valuable in complex clinical scenarios.

To enhance methodological rigor, this study incorporated insights from prior research and introduced independent testing, which was absent in similar studies. Independent testing enabled objective assessment of model performance, reduction of overfitting, and increased reliability of findings. For example, Jiang et al. developed a predictive model for forecasting JOARR in DCM patients, achieving an AUC improvement from 0.78 to 0.81 through an ensemble learning approach. Similarly, by stacking the top three predictive models in this study, an AUC increase from 0.81 to 0.92 was observed[20]. This improvement underscores the potential of ensemble learning to enhance classification accuracy and affirms the model's practical applicability in clinical settings.

To further assess the role of each feature in the final ensemble model, the relative importance of predictors was evaluated. Machine learning-driven feature selection offers significant advantages, particularly in reducing the subjective bias commonly associated with manual selection methods[28–30]. By algorithmically determining the significance of predictors, this approach ensures an objective and data-driven process. It is especially effective in managing large datasets with numerous variables, facilitating the identification of the most relevant features. This reduction in dimensionality minimizes redundancy, irrelevant information, and the risk of overfitting. Streamlined models resulting from this process not only operate more efficiently—requiring fewer data and computational resources—but also maintain high levels of generalizability. Moreover, isolating the most impactful factors provides valuable scientific insights by revealing key causal relationships, which can guide future research by emphasizing high-value variables that drive outcomes. In summary, automated and unbiased

feature selection enhances model performance and efficiency while deepening the understanding of complex phenomena, making it essential for advancements across various scientific fields.

Our analysis identified preoperative axial pain intensity, JOARR, preoperative cervical C2–7 Cobb angle, HADS-D score, and age as the most predictive features, aligning with established predictors of postoperative axial pain (PAP) in patients with degenerative cervical myelopathy (DCM)[7,14,26]. Notably, preoperative axial pain emerged as the most significant risk factor for predicting postoperative pain. A study revealed that approximately 40% of patients experienced postoperative axial pain, predominantly affecting those with a history of preoperative pain[31]. Research by Su et al. linked preoperative pain hypersensitivity to the development of PAP in DCM patients, indicating that individuals with preoperative axial pain exhibited higher levels of pain hypersensitivity, leading to increased postoperative pain intensity[8]. Additionally, our findings highlighted the JOARR as a significant predictor of prognosis, serving as an important indicator of neurological recovery and functional outcomes post-surgery. Yoshida et al. found that patients with poor JOARR often experience more severe postoperative axial symptoms, suggesting that degenerative changes in the dorsal horn of the spinal cord may contribute to chronic axial pain in those with impaired neurological function[7]. Toyama et al. also indicated that certain cervical myelopathic pain types may stem from abnormalities in second-order neurons located in the dorsal horn[32]. The preoperative cervical C2–7 Cobb angle (CCA) is another risk factor for predicting PAP, consistenting with previous studies. Previous studies have demonstrated a negative correlation between preoperative CCA and postoperative axial symptoms[14]. Biomechanically, a cervical spine with reduced lordosis or kyphotic alignment experiences increased flexural stress, contributing to postoperative axial pain[33]. Chavanne et al. found that a cervical lordosis of less than 7.5° resulted in elevated intramedullary pressure within the spinal cord, hypothesizing a higher likelihood of developing postoperative axial pain[34]. Notably, our study identified the HADS-D score as a key predictive factor for PAP intensity, indicating that depressive states in DCM patients are associated with postoperative pain severity. Previous research highlights a bidirectional relationship between chronic pain and comorbid depression, with Kroenke et al. demonstrating that pain predicts depression severity while depression predicts pain intensity[35]. In this context, chronic neck pain in DCM patients may lead to psychological comorbidities such as depression, exacerbating the severity of PAP after posterior cervical decompression surgery. The role of age as a risk factor for PAP in DCM patients remains controversial[7,26,36]. Kato et al. discovered that an older age (greater than 63 years) significantly reduced the risk of postoperative axial pain, while another study reported that patients over 70 years of age experienced significantly higher levels of axial pain[26]. Our findings suggest that age is one of the most important predictive factors for PAP intensity, as age-related spinal cord changes and comorbidities may hinder elderly patients from achieving the same level of functional improvement as younger patients, potentially leading to PAP[7]. Additionally, degenerative changes in the dorsal horn of the spinal cord may be associated with postoperative axial pain in patients with cervical spondylosis. In conclusion, our thorough analysis not only confirmed established predictors but also identified new determinants, including the HADS-D score, thereby improving the accuracy of PAP predictions in patients with DCM.

It is worth noting that the choice of surgical techniques in this study may also influence the occurrence of PAP. Simple laminectomy achieves decompression by removing the lamina but may cause biomechanical instability by disrupting posterior muscles and ligaments, leading to atrophy, dysfunction, and persistent neck pain[37]. Laminoplasty, which preserves part of the posterior structures, reduces cervical instability but may still result in muscle atrophy and pain due to muscle detachment or damage to facet joints and ligaments[31,36,38]. Laminectomy with fusion combines decompression and stabilization using internal fixation, minimizing instability-related pain but restricting cervical motion, which can cause compensatory stress and new sources of discomfort[37]. Additionally, recent advances, such as the application of endoscopic techniques and minimally invasive tubular approaches in DCM patients, have shown promising outcomes in reducing postoperative pain and improving recovery. These approaches minimize tissue disruption, preserve posterior muscle and ligament integrity, and shorten recovery times, offering potential advantages over traditional surgical methods[39,40].

Limitations

This study has several limitations that warrant acknowledgment. First, the sample size is relatively small, and all patients were recruited from a single center, which limits the generalizability of our findings. Future studies utilizing larger, multicenter datasets are necessary to validate the predictive performance and robustness of our machine learning-based ensemble model across diverse patient populations. Second, the retrospective nature of this study may have introduced biases, particularly in the selection of predictor variables. For instance, certain key preoperative radiographic parameters, such as global sagittal alignment, which significantly influence patient outcomes and quality of life, were not included in our model. Given that overall sagittal balance is a critical determinant of postoperative recovery, future research should incorporate full-spine standing radiographs to evaluate the impact of global sagittal alignment on postoperative axial pain. Additionally, prior studies have underscored the benefits of modified surgical techniques and paraspinal muscle preservation in reducing postoperative axial pain. However, our retrospective study did not capture predictive data related to these surgical techniques. Prospective studies with more comprehensive clinical and imaging data are warranted to address these gaps in future research. Furthermore, postoperative changes in cervical alignment, alongside preoperative alignment, should also be considered as potential predictors of postoperative pain. In clinical practice, uncorrected cervical kyphosis and a high SVA may lead to poorer outcomes for patients[37,41,42]. Due to the limitations of the existing data, we did not analyze the changes in postoperative alignment. Future studies should consider these factors. Finally, the inclusion of different surgical types (laminoplasty, laminectomy, and fusion) introduces heterogeneity. Due to the small sample size, stratification was not feasible, though surgical type was analyzed as a risk factor. Future studies with larger cohorts should address this issue.

Conclusion

The ensemble classifiers successfully predicted postoperative axial pain intensity in DCM patients, and preoperative axial pain intensity was identified as the most relevant predictor. This finding can assist clinicians in identifying patients with postoperative axial pain preoperatively, thereby enabling more rational formulation of perioperative management strategies.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Received: 2 November 2024; Accepted: 17 March 2025
Published online: 22 March 2025

References

1. Nurick, S. The pathogenesis of the spinal cord disorder associated with cervical spondylosis. *Brain: J. Neurol.* **95**, 87–100. https://doi.org/10.1093/brain/95.1.87 (1972).
2. Fehlings, M. G. et al. Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy: results of the AOSpine North America prospective multi-center study. *J. Bone Joint Surg. Am. Vol.* **95**, 1651–1658. https://doi.org/10.2106/jbjs.L.00589 (2013).
3. Toledano, M. & Bartleson, J. D. Cervical spondylotic myelopathy. *Neurol. Clin.* **31**, 287–305. https://doi.org/10.1016/j.ncl.2012.09.003 (2013).
4. Wang, S. J., Jiang, S. D., Jiang, L. S. & Dai, L. Y. Axial pain after posterior cervical spine surgery: a systematic review. *Eur. Spine J.* **20**, 185–194 (2011). https://doi.org/10.1007/s00586-010-1600-x
5. Higashino, K. et al. Preservation of C7 spinous process does not influence the long-term outcome after laminoplasty for cervical spondylotic myelopathy. *Int. Orthop.* **30**, 362–365. https://doi.org/10.1007/s00264-005-0062-y (2006).
6. Hyun, S. J., Rhim, S. C., Roh, S. W., Kang, S. H. & Riew, K. D. The time course of range of motion loss after cervical laminoplasty: a prospective study with minimum two-year follow-up. *Spine* **34**, 1134–1139. https://doi.org/10.1097/BRS.0b013e31819c389b (2009).
7. Yoshida, M. et al. Does reconstruction of posterior ligamentous complex with extensor musculature decrease axial symptoms after cervical laminoplasty? *Spine* **27** 1414–1418 (2002). https://doi.org/10.1097/00007632-200207010-00008
8. Su, Q., Li, J., Chu, X. & Zhao, R. Preoperative pain hypersensitivity is associated with axial pain after posterior cervical spinal surgeries in degenerative cervical myelopathy patients: a preliminary resting-state fMRI study. *Insights into imaging.* **14**, 16. https://doi.org/10.1186/s13244-022-01332-2 (2023).
9. Obermeyer, Z. & Emanuel, E. J. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N. Engl. J. Med.* **375**, 1216–1219. https://doi.org/10.1056/NEJMp1606181 (2016).
10. Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29. https://doi.org/10.1038/s41591-018-0316-z (2019).
11. Inose, H. et al. Factors contributing to neck pain in patients with degenerative cervical myelopathy: A prospective multicenter study. *J. Orthop. Surg.* **30**, 10225536221091848. https://doi.org/10.1177/10225536221091848 (2022).
12. Kimura, A., Shiraishi, Y., Inoue, H., Endo, T. & Takeshita, K. Predictors of Persistent Axial Neck Pain After Cervical Laminoplasty. *Spine* **43**, 10–15. https://doi.org/10.1097/brs.0000000000002267 (2018).
13. Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. *Nat. Methods.* **15**, 233–234. https://doi.org/10.1038/nmeth.4642 (2018).
14. Liu, Y. et al. Preoperative Factors Affecting Postoperative Axial Symptoms After Single-Door Cervical Laminoplasty for Cervical Spondylotic Myelopathy: A Prospective Comparative Study. *Med. Sci. monitor: Int. Med. J. experimental Clin. Res.* **22**, 3746–3754. https://doi.org/10.12659/msm.900954 (2016).
15. Fan, G. et al. Machine Learning-based Prediction of Prolonged Intensive Care Unit Stay for Critical Patients with Spinal Cord Injury. *Spine* **47**, E390–e398. https://doi.org/10.1097/brs.0000000000004267 (2022).
16. Antonucci, L. A. et al. An Ensemble of Psychological and Physical Health Indices Discriminates Between Individuals with Chronic Pain and Healthy Controls with High Reliability: A Machine Learning Study. *Pain therapy*. **9**, 601–614. https://doi.org/10.1007/s40122-020-00191-3 (2020).
17. Li, J. et al. Abnormal preoperative fMRI signal variability in the pain ascending pathway is associated with the postoperative axial pain intensity in degenerative cervical myelopathy patients. *spine journal: official J. North. Am. Spine Soc.* **24**, 78–86. https://doi.org/10.1016/j.spinee.2023.09.003 (2024).
18. Lin, X. et al. Rasch analysis of the hospital anxiety and depression scale among Chinese cataract patients. *PloS one.* **12**, e0185287. https://doi.org/10.1371/journal.pone.0185287 (2017).
19. Hirabayashi, K., Miyakawa, J., Satomi, K., Maruyama, T. & Wakano, K. Operative results and postoperative progression of ossification among patients with ossification of cervical posterior longitudinal ligament. *Spine* **6**, 354–364. https://doi.org/10.1097/00007632-198107000-00005 (1981).
20. Cai, Z., Sun, Q., Li, C., Xu, J. & Jiang, B. Machine-learning-based prediction by stacking ensemble strategy for surgical outcomes in patients with degenerative cervical myelopathy. *J. Orthop. Surg. Res.* **19**, 539. https://doi.org/10.1186/s13018-024-05004-3 (2024).
21. Gerbershagen, H. J., Rothaug, J., Kalkman, C. J. & Meissner, W. Determination of moderate-to-severe postoperative pain on the numeric rating scale: a cut-off point analysis applying four different methods. *Br. J. Anaesth.* **107**, 619–626. https://doi.org/10.1093/bja/aer195 (2011).
22. Boonstra, A. M. et al. Cut-Off Points for Mild, Moderate, and Severe Pain on the Numeric Rating Scale for Pain in Patients with Chronic Musculoskeletal Pain: Variability and Influence of Sex and Catastrophizing. *Front. Psychol.* **7**, 1466. https://doi.org/10.3389/fpsyg.2016.01466 (2016).
23. Vittinghoff, E. & McCulloch, C. E. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am. J. Epidemiol.* **165**, 710–718. https://doi.org/10.1093/aje/kwk052 (2007).
24. Chen, R. et al. Predictive value of preoperative short form-36 survey scale for postoperative axial neck pain in patients with degenerative cervical myelopathy. *Glob. Spine J.* 21925682231200136 https://doi.org/10.1177/21925682231200136 (2023).
25. Hosono, N., Sakaura, H., Mukai, Y., Fujii, R. & Yoshikawa, H. C3-6 laminoplasty takes over C3-7 laminoplasty with significantly lower incidence of axial neck pain. *Eur. Spine J.* **15**, 1375–1379. https://doi.org/10.1007/s00586-006-0089-9 (2006).
26. Kato, M. et al. Effect of preserving paraspinal muscles on postoperative axial pain in the selective cervical laminoplasty. *Spine* **33**, E455–459. https://doi.org/10.1097/BRS.0b013e318178e607 (2008).
27. Cao, Y. et al. Preoperative Cervical Cobb Angle Is a Risk Factor for Postoperative Axial Neck Pain after Anterior Cervical Discectomy and Fusion with Zero-Profile Interbody. *Orthop. Surg.* **14**, 3225–3232. https://doi.org/10.1111/os.13552 (2022).

28. Gholizadeh, M., Saeedi, R., Bagheri, A. & Paeezi, M. Machine learning-based prediction of effluent total suspended solids in a wastewater treatment plant using different feature selection approaches: A comparative study. *Environ. Res.* **246**, 118146. https://doi.org/10.1016/j.envres.2024.118146 (2024).
29. Alfraihat, A., Samdani, A. F. & Balasubramanian, S. Predicting radiographic outcomes of vertebral body tethering in adolescent idiopathic scoliosis patients using machine learning. *PloS one*. **19**, e0296739. https://doi.org/10.1371/journal.pone.0296739 (2024).
30. Rezvantalab, S., Mihandoost, S. & Rezaiee, M. Machine learning assisted exploration of the influential parameters on the PLGA nanoparticles. *Sci. Rep.* **14**, 1114. https://doi.org/10.1038/s41598-023-50876-w (2024).
31. Ohnari, H. et al. Investigation of axial symptoms after cervical laminoplasty, using questionnaire survey. *spine journal: official J. North. Am. Spine Soc.* **6**, 221–227. https://doi.org/10.1016/j.spinee.2005.10.014 (2006).
32. Toyama, Y. & Fujumura, H. K. Central pain of the spinal cord origin: Pain of the spinal cord origin in cervical myelopathy. *Jpn Dtsch. Med. Ber.* **33**, 43–56 (1988).
33. Harrison, D. E. et al. Comparison of axial and flexural stresses in lordosis and three buckled configurations of the cervical spine. *Clin. Biomech. (Bristol, Avon)*. **16**, 276–284. https://doi.org/10.1016/s0268-0033(01)00006-7 (2001).
34. Chavanne, A., Pettigrew, D. B., Holtz, J. R., Dollin, N. & Kuntz, C. Spinal cord intramedullary pressure in cervical kyphotic deformity: a cadaveric study. *Spine* **36**, 1619–1626. https://doi.org/10.1097/BRS.0b013e3181fc17b0 (2011).
35. Kroenke, K. et al. Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. *J. pain*. **12**, 964–973. https://doi.org/10.1016/j.jpain.2011.03.003 (2011).
36. Kawaguchi, Y., Matsui, H., Ishihara, H., Gejo, R. & Yoshino, O. Axial symptoms after en bloc cervical laminoplasty. *J. Spinal Disord.* **12**, 392–395 (1999).
37. Rhee, J. M. & Basra, S. Posterior surgery for cervical myelopathy: laminectomy, laminectomy with fusion, and laminoplasty. *Asian spine J.* **2**, 114–126. https://doi.org/10.4184/asj.2008.2.2.114 (2008).
38. Ruan, C. et al. Analysis of risk factors for axial symptoms after posterior cervical open-door laminoplasty. *J. Orthop. Surg, Res.* **18**, 954. https://doi.org/10.1186/s13018-023-04426-9 (2023).
39. Huang, C. C., Fitts, J., Huie, D., Bhowmick, D. A. & Abd-El-Barr, M. M. Evolution of cervical endoscopic spine surgery: Current progress and future directions-a narrative review. *J. Clin. Med.* **13**. https://doi.org/10.3390/jcm13072122 (2024).
40. Cai, R. Z., Wang, Y. Q., Wang, R., Wang, C. H. & Chen, C. M. Microscope-assisted anterior cervical discectomy and fusion combined with posterior minimally invasive surgery through tubular retractors for multisegmental cervical spondylotic myelopathy: A retrospective study. *Medicine* **96**, e7965. https://doi.org/10.1097/md.0000000000007965 (2017).
41. Dru, A. B. et al. Cervical Spine Deformity Correction Techniques. *Neurospine* **16**, 470–482. https://doi.org/10.14245/ns.1938288.144 (2019).
42. Tang, J. A. et al. The impact of standing regional cervical sagittal alignment on outcomes in posterior cervical fusion surgery. *Neurosurgery* **76** (Suppl 1), S14–S21. https://doi.org/10.1227/01.neu.0000462074.66077.2b (2015). (Discussion S21).

Author contributions

XC designed the study and authored the manuscript. XC and JJS contributed to data collection and statistical analysis. JDW participated in software and data visualization. HK revised the manuscript. All authors have read and approved the final version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

Approval for this study was granted by the Institutional Review Board of Honghui Hospital. Written informed consent was obtained from each participant before the procedures were conducted.

Consent for publication

The authors confirm that informed consent for the publication of images in all figures was secured from the human research participants.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94755-y.

Correspondence and requests for materials should be addressed to H.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025