Classification of expert-level therapeutic decisions for degenerative cervical myelopathy using ensemble machine learning algorithms

Dougho Park[1], Jae Man Cho[2], Joong Won Yang[2], Donghoon Yang[2], Mansu Kim[2], Gayeoul Oh[3] and Heum Dai Kwon[2]*

[1]Department of Rehabilitation Medicine, Pohang Stroke and Spine Hospital, Pohang, South Korea, [2]Department of Neurosurgery, Pohang Stroke and Spine Hospital, Pohang, South Korea, [3]Department of Radiology, Pohang Stroke and Spine Hospital, Pohang, South Korea

Background: Therapeutic decisions for degenerative cervical myelopathy (DCM) are complex and should consider various factors. We aimed to develop machine learning (ML) models for classifying expert-level therapeutic decisions in patients with DCM.

Methods: This retrospective cross-sectional study included patients diagnosed with DCM, and the diagnosis of DCM was confirmed clinically and radiologically. The target outcomes were defined as conservative treatment, anterior surgical approaches (ASA), and posterior surgical approaches (PSA). We performed the following classifications using ML algorithms: multiclass, one-versus-rest, and one-versus-one. Two ensemble ML algorithms were used: random forest (RF) and extreme gradient boosting (XGB). The area under the receiver operating characteristic curve (AUC-ROC) was the primary metric. We also identified the variable importance for each classification.

Results: In total, 304 patients were included (109 conservative, 66 ASA, 125 PSA, and 4 combined surgeries). For multiclass classification, the AUC-ROC of RF and XGB models were 0.91 and 0.92, respectively. In addition, ML models showed AUC-ROC values of >0.9 for all types of binary classifications. Variable importance analysis revealed that the modified Japanese Orthopaedic Association score and central motor conduction time were the two most important variables for distinguishing between conservative and surgical treatments. When classifying ASA and PSA, the number of involved levels, age, and body mass index were important contributing factors.

Conclusion: ML-based classification of DCM therapeutic options is valid and feasible. This study can be a basis for establishing generalizable ML-based surgical decision models for DCM. Further studies are needed with a large multicenter database.

Introduction

Degenerative cervical myelopathy (DCM) is a disease that causes progressive and nontraumatic cervical spinal cord compression due to degenerative changes in the cervical spine (1). Patients with DCM present a broad spectrum of symptoms, ranging from subtle sensory neuropathic symptoms to motor weakness with functional disability, depending on disease progression and severity (2). Patients with DCM require early diagnosis and management; specifically, it is essential to determine an appropriate therapeutic option according to the disease severity while minimizing damage to the spinal cord (3, 4).

There have been studies on determining the appropriate treatment options for DCM (5). A randomized controlled study of patients with mild-to-moderate DCM (modified Japanese Orthopaedic Association [mJOA] score >12) conducted by Kadanke et al. (6) showed that surgical treatment was not superior to conservative treatment. In contrast, studies have also suggested that surgical treatment demonstrated better functional recovery and patient satisfaction in patients with more severe DCM (7). In the guidelines presented in 2017, Fehlings et al. (8) argued that early diagnosis and surgical treatment are necessary for moderate-to-severe DCM. On the other hand, they reported that there was a knowledge gap in the selection of therapeutic options for mild DCM. Meanwhile, studies on the optimal surgical approach method in patients with DCM who decided to undergo surgery have also been reported (9). Several studies have compared outcomes after anterior and posterior surgical approaches (ASA and PSA, respectively) for DCM; however, no clear evidence of superiority has been established to date (5). A systematic review also concluded that there are no apparent differences between surgical methods in terms of neurologic recovery (10). Consequently, there has been a debate regarding the therapeutic decision for DCM (11). In addition, the surgical indications can be slightly different depending on the surgeon's practice style, preference, and health insurance system (12, 13).

Machine learning (ML) algorithms are actively applied to research for medical decisions because of their excellent classification and prediction (14). Specifically, ML algorithms can handle data with huge samples and use clinical information to make the medical decision-making process more efficient (15). In line with this, Park et al. (16) presented a disease severity classification model to minimize unnecessary electrodiagnostic testing in patients with carpal tunnel syndrome. Yoo et al. (17) reported an ML decision model for the selection of an optimal laser refractive surgery method. In addition, studies have applied ML to DCM. Merali et al. (18) predicted the 6-, 12-, and 24-month outcomes in 757 patients with DCM and showed an area under the receiver operating characteristic curve (AUC-ROC) of approximately 0.70. Hopkins et al. (19) presented models predicting DCM diagnosis and

mJOA scores using a deep neural network in 14 patients with DCM and 14 healthy controls. In these studies, ensemble ML algorithms presented valid and accurate results, proving their clinical usefulness. As a result, ML-related research in the clinical field has been growing rapidly. Nevertheless, an ML model that classifies the therapeutic decisions in DCM has not yet been reported to the best of our knowledge.

In this context, we developed an ML-based model for classifying expert-level therapeutic decisions using ensemble ML algorithms in patients with DCM to verify their performance. In addition, we investigated the contributing factors involved in the therapeutic decisions in DCM through feature importance results derived from optimal ML classification models.
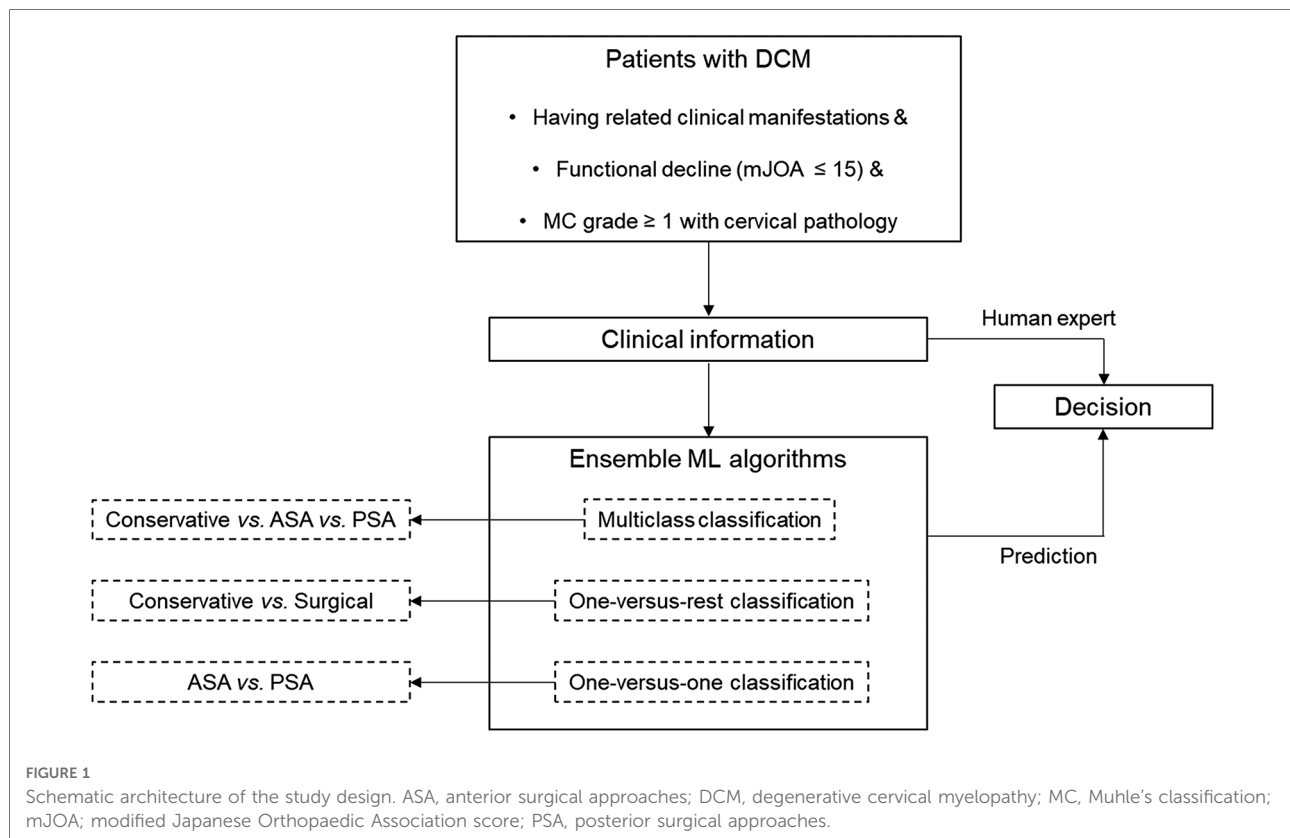
Materials and methods

Study design

This single-center, retrospective study enrolled patients diagnosed with DCM between January 2017 and December 2021. The dataset included patients of experienced neurosurgeons in our hospital who also co-authored this study. This study was reviewed and approved by the institutional review board of Pohang Stroke and Spine Hospital (PSSH-0475-202202-HR-007-01). It was performed in compliance with the Declaration of Helsinki and the International Conference on Harmonization-Good Clinical Practice Guidelines. Informed consent was not required owing to the retrospective nature of the study design.

The diagnosis of DCM was confirmed when it satisfied the following criteria: clinical manifestation, functional level, and cervical cord compression grade on magnetic resonance imaging (MRI). We utilized patient information—demographic, clinical, radiological, and electrodiagnostic characteristics—as variables for ML classification based on electrical health records (EHR). Therefore, variable definitions were primarily based on formats recorded in the EHR. Detailed information on each variable is provided in **Supplementary Table S1**.

We performed an ML algorithm-based ensemble classification using the extracted variables. The ground truth was set based on an experienced neurosurgeon's decision. We proceeded with three types of classification—multiclass, one-versus-rest, and one-versus-one—according to the target outcomes (**Figure 1**).

Patients and clinical assessments

DCM-related clinical manifestations were defined as follows: presence of upper motor neuron signs, clumsy hands,

FIGURE 1
Schematic architecture of the study design. ASA, anterior surgical approaches; DCM, degenerative cervical myelopathy; MC, Muhle's classification; mJOA; modified Japanese Orthopaedic Association score; PSA, posterior surgical approaches.

motor weakness, atrophic muscle change, progressive gait disturbance, sensory loss with altered proprioception, paresthesia, and bowel or bladder symptoms. Patients with at least one or more symptoms and signs, and a functional decline of ≤15 on the mJOA score were included in the confirmed cases of DCM. An experienced neurosurgeon or physiatrist enquired and evaluated the patient's symptoms and recorded the score. The exclusion criteria for this study were as follows: (1) brain lesion or brain surgery history, (2) history of cervical spine surgery, (3) previous thoracic myelopathy or cauda equina syndrome, (4) cervical cord compression lesion lower than the C6/7 level, (5) severe carpal tunnel syndrome, and (6) missing values.

During the first visit, in the period in which DCM was diagnosed, we measured the patient's numerical rating scale (NRS) of the neck and arm pain and recorded the subjective symptom duration in months. We identified comorbidities *via* patient or guardian interviews, medical records, and medication history.

The target classes were conservative treatment, ASA, and PSA. ASA included anterior cervical discectomy and fusion, anterior cervical corpectomy and fusion, and cervical arthroplasty. PSA included laminoplasty and laminectomy, with or without fusion. Considering poor compliance or transferred patients, we set the ground truth based on the

decision of human experts, not on whether the actual surgery was performed.

Radiologic features

Cervical MRI scans were performed using a 1.5 T Philips Achieva (Philips Medical Systems, Eindhoven, Netherlands). Imaging findings were analyzed at the time of DCM diagnosis and re-reviewed by a radiologist for this study as well. In case of discrepancy during the retrospective review, it was finally decided by the consensus of the radiologists and surgeon. The Muhle's classification (MC) grade, which was interpreted based on the most stenotic level in the midsagittal view of the MRI, was used as an index for cervical spinal cord compression (20). In addition, we counted the number of levels with an MC grade of I or higher. Lesion types were classified as ossification of the posterior longitudinal ligament (OPLL), disc herniation, spondylolisthesis, or others (including combined pathologies). We also identified the presence of high signal intensity (HSI) on T2 MRI.

We drew the k-line in the midline of the sagittal view of the cervical radiograph, which was recorded in the standing state to confirm cervical spine alignment (21). The k-line evaluation using a radiograph was retrospectively reviewed by an

experienced radiologist and neurosurgeon. In case of discrepancy, it was finally decided by the consensus of the two interpreters.

Electrodiagnostic evaluation

All electrodiagnostic tests were performed using a Sierra Wave (Cadwell Laboratories Inc., Kennewick, WA, USA). We conducted transmagnetic stimulation using MagPro Compact and a circular coil with an outer diameter of 12 cm (MagVenture Inc., Alpharetta, GA, USA). Stimulation for motor-evoked potential (MEP) was performed at the cortical and cervical levels and recorded from the abductor pollicis brevis muscle. Cortex stimulation was applied to the Cz region according to the international 10–20 system, and cervical stimulation was performed at the spinous process at the C7 level to obtain peripheral motor conduction time. The central motor conduction time (CMCT) was then calculated as follows (22):

$$CMCT \; (ms) = \text{cortical MEP onset latency}$$
$$- \text{cervical MEP onset latency}.$$

We used the CMCT value as a variable for the symptomatic side and a slower CMCT value for bilateral symptoms. CMCT was divided into four categories: normal, <11.5 ms; mildly delayed, $11.5 \leq CMCT < 15$; definitely delayed, ≥ 15; and not evoked MEP.

We evaluated the upper extremity nerve conduction study and electromyography in all patients for differential diagnosis and identified radiculopathy, which was confirmed by electromyography. We diagnosed concomitant radiculopathies with denervation potentials in two or more muscles innervated by different peripheral nerves in the specific myotome (23). All electrodiagnostic tests were performed and interpreted by experienced physiatrists.

Statistical analysis

We used R software version 4.1.3 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria) and GraphPad Prism 9.3.1 (GraphPad Software, San Diego, CA, USA) for statistical analyses. Continuous variables were analyzed using Shapiro–Wilk normality test and expressed as mean ± standard deviation or median (interquartile range). Categorical variables are expressed as frequency (proportion). For comparative analysis between the conservative, ASA, and PSA groups, one-way analysis of variance with Bonferroni's multiple comparison test or Kruskal-Wallis test with Dunn's multiple comparisons test was applied to continuous variables,

and chi-squared trend test was used for categorical variables. Statistical significance was set at $p < 0.05$.

Machine learning processing

ML modeling was performed based on the "caret" package of the R software (24). The ML processing in this study is shown in Figure 2. The R code in this study is available in the Online supplementary material. We confirmed variables with near-zero variance and multicollinearity (correlation coefficient >0.7) for variable selection. Subsequently, centering and scaling were performed for numeric variables, and one-hot encoding was performed for categorical variables.
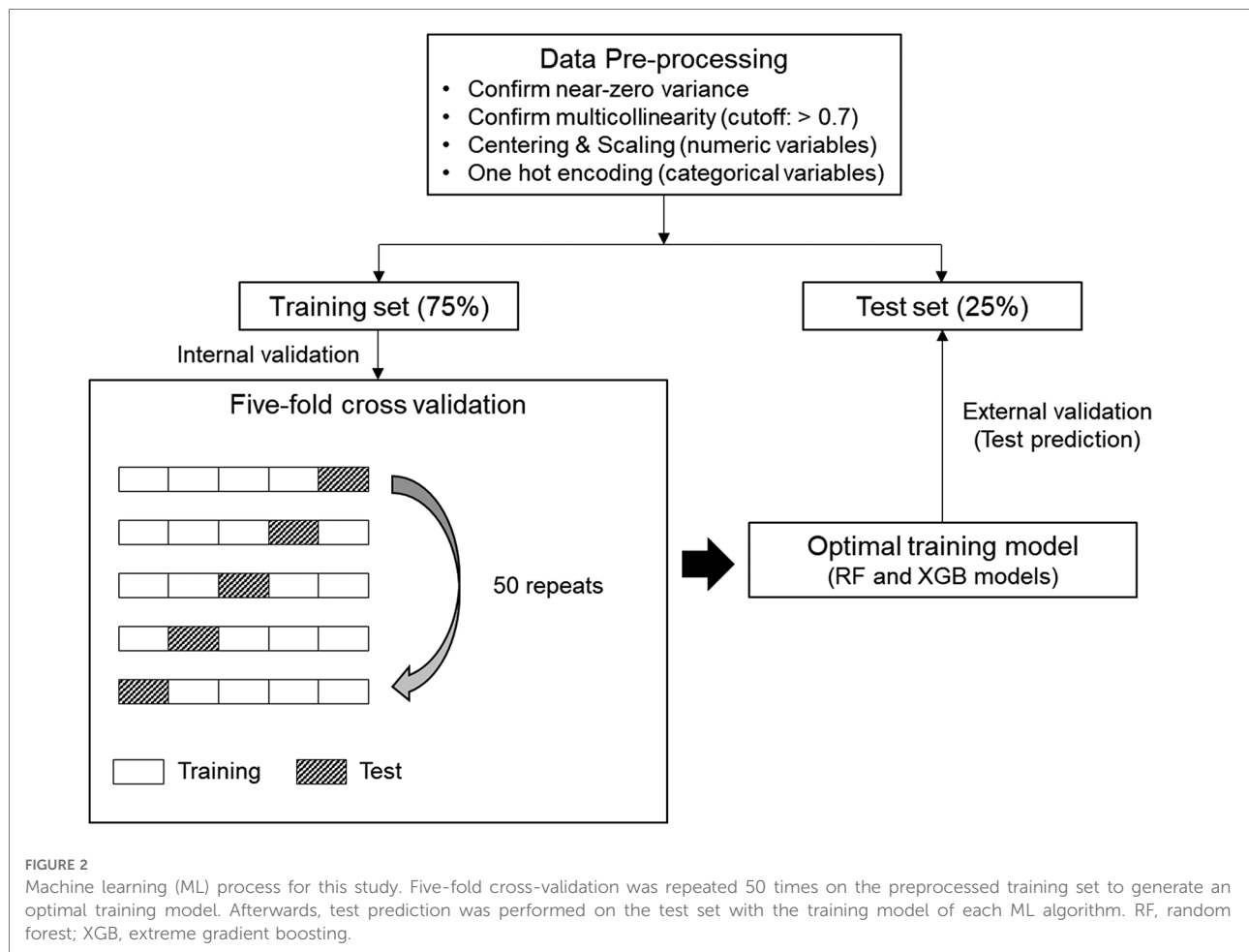
The whole dataset was randomly split into 75%-training and 25%-test sets. In the binary classifications, the target class-balanced training set was additionally formed by applying the synthetic minority oversampling technique, after which we drove the ML algorithms. Random forest (RF) and extreme gradient boosting (XGB), which are representative ensemble ML algorithms, were used (25, 26). Five-fold cross-validation was repeated 50 times to generate the optimal training model. Random and grid search methods were applied for hyperparameter tuning. The tuned hyperparameter values applied to each classification are presented in Supplementary Table S2. External validation was performed by applying the training model generated for each algorithm to the test set. AUC-ROC and overall accuracy were used as the main metrics for multiclass classification. In the multiclass classification, AUC-ROC was defined as the micro-average value calculated by converting the multiclass into the sum of binary classifications using the "multiROC" package of R software (27). In the binary classifications, AUC-ROC, F1 measure, and area under the precision-recall curve (AUC-PR) were evaluated using the "MLeval" package of the R software (28).

Results

Baseline characteristics

A total of 304 patients were included (109 conservative and 195 surgical treatments). ASA was chosen in 66 patients, PSA in 125, and a combined approach was selected as the surgical treatment in four patients.

The PSA group was significantly older than the other two groups ($p < 0.001$), and the proportion of men was relatively higher ($p = 0.023$). In addition, the proportion of patients covered by medical-aid and living in rural areas in the PSA group was significantly higher ($p = 0.045$ and $p < 0.001$, respectively). Meanwhile, the ASA group showed a significantly lower body mass index (BMI) value of 23.5

FIGURE 2
Machine learning (ML) process for this study. Five-fold cross-validation was repeated 50 times on the preprocessed training set to generate an optimal training model. Afterwards, test prediction was performed on the test set with the training model of each ML algorithm. RF, random forest; XGB, extreme gradient boosting.

(21.3–25.3) kg/m^2 compared to 24.6 (23.0–26.6) kg/m^2 in the PSA group ($p = 0.036$) (**Table 1** and **Supplementary Table S3**).

The PSA group had a significantly longer symptom duration than the other two groups ($p < 0.001$). The mJOA score was 12.0 (11.0–13.0) in the ASA and PSA groups, which was significantly lower than 14.0 (14.0–14.0) in the conservative group ($p < 0.001$). Moreover, the bilateral symptom rate was the highest in the PSA group ($p = 0.013$). The involved level count was the highest in the PSA group at 3.0 (3.0–4.0) levels, and lowest in the ASA group at 1.0 (1.0–2.0) level ($p < 0.001$). Regarding the lesion type, the disc herniation rate was relatively higher in the ASA group, while the rates for OPLL, spondylolisthesis, and other/combined etiology were relatively higher in the PSA group ($p < 0.001$). The surgical treatment group showed a higher rate of HSI on T2 images, higher MC grade, and lower rate of the k-line (+) cases ($p < 0.001$, $p < 0.001$, and $p = 0.003$, respectively). Further, the surgical treatment group presented more severe CMCT deterioration and a higher rate of radiculopathy ($p < 0.001$ and $p = 0.026$, respectively) (**Table 2** and **Supplementary Table S3**).

Multiclass classification

When multiclass classification was performed between the conservative, ASA, and PSA groups, the test prediction AUC-ROC of the RF and XGB models were 0.91 and 0.92, respectively. The overall accuracies of the RF and XGB models were 76.2% and 74.6%, respectively (**Table 3**). The confusion matrix of multiclass classification for each algorithm is presented in **Supplementary Table S4**.

Binary classifications (one-versus-rest and one-versus-one)

The results of the binary classification of each algorithm are summarized in **Table 4**. The confusion matrix for each classification is presented in **Supplementary Table S5**.

In the classification model between conservative and surgical treatments, RF and XGB showed AUC-ROC values of 0.94 and 0.93, respectively. The mJOA score and normal

TABLE 1 Baseline characteristics.

	Conservative (*n* = 109)	ASA (*n* = 66)	PSA (*n* = 125)	*p*-value
Age, years	55.9 ± 11.4	57.9 ± 13.1	64.0 ± 10.9	<0.001
Male, *n* (%)	61 (56.0)	40 (60.6)	91 (72.8)	0.023
BMI, kg/m^2	24.0 (22.2–26.7)	23.5 (21.3–25.3)	24.6 (23.0–26.6)	0.036
Medical-aid, *n* (%)	3 (2.8)	2 (3.0)	12 (9.6)	0.045
Urban residence, *n* (%)	75 (68.8)	39 (59.1)	53 (42.4)	<0.001
Comorbidities, *n* (%)				
Hypertension	34 (31.2)	25 (37.9)	54 (43.2)	0.167
Diabetes	15 (13.8)	27 (21.6)	13 (19.7)	0.287
Dyslipidemia	14 (12.8)	9 (13.6)	20 (16.0)	0.777
Heart problems[a]	5 (4.6)	0 (0.0)	10 (8.0)	0.053
Degenerative lumbar disease	28 (25.7)	23 (34.8)	56 (44.8)	0.010

ASA, anterior surgical approaches; BMI, body mass index; PSA, posterior surgical approaches.
[a]Symptomatic arrhythmia or coronary artery disease.

CMCT category were the two most important variables in both RF and XGB models for classifying conservative and surgical groups. Age and symptom duration were also included in the top five important features in both models (**Table 5**).

The ensemble ML algorithms demonstrated outstanding performance in ASA and PSA classification; both RF and XGB showed AUC-ROC of 0.99 and 0.96, respectively. The most important variable in classifying ASA and PSA was confirmed to be level count in both the RF and XGB models. In addition, age, BMI, and subjective neck pain were identified as the major common features in both models (**Table 5**).

Discussion

In this study, we suggest a strategy for therapeutic decision-making by applying an ML model in patients with DCM. Our results demonstrated that the proposed ML models showed AUC-ROC values of >0.9 in all types of classifications, proving their outstanding performance. There has been an unmet need for personalized, cost-effective surgical decisions and treatment of DCM (29). As the first attempt to utilize an ML-based approach for therapeutic selection in patients with DCM, this study is significant because we present a direction for surgical decisions using a novel method. Furthermore, we presented the possibility of a comprehensive decision process for DCM therapeutic options using various clinically available determinants, which is an advantage of the ML process, rather than relying on any specific parameter. Consequently, our ML classification models can be beneficial for spine surgeons in choosing the most effective and proper surgical approach for DCM.

TABLE 2 Disease-related features.

	Conservative (*n* = 109)	ASA (*n* = 66)	PSA (*n* = 125)	*p*-value
Symptom duration, months	3.0 (2.0–6.0)	3.5 (2.0–12.0)	10.0 (3.0–24.0)	<0.001
NRS, neck	3.0 (2.0–5.0)	4.0 (3.0–6.0)	3.0 (2.0–5.0)	0.099
NRS, arm	4.0 (3.0–5.0)	5.0 (3.0–7.0)	4.0 (3.0–6.0)	0.146
mJOA score	14.0 (14.0–14.0)	12.0 (11.0–13.0)	12.0 (11.0–13.0)	<0.001
Symptom side, *n* (%)				0.013
Right	25 (22.9)	11 (16.7)	12 (9.6)	
Left	32 (29.4)	13 (19.7)	28 (22.4)	
Bilateral	52 (47.7)	42 (63.6)	85 (68.0)	
Number of involved levels	2.0 (1.0–3.0)	1.0 (1.0–2.0)	3.0 (3.0–4.0)	<0.001
Lesion type, *n* (%)				<0.001
OPLL	44 (40.4)	14 (21.2)	41 (32.8)	
Disc herniation	56 (51.4)	33 (50.0)	17 (13.6)	
Spondylolisthesis	6 (5.5)	12 (18.2)	39 (31.2)	
Others or combined	3 (2.8)	7 (10.6)	28 (22.4)	
Most stenotic level, *n* (%)				0.082
C1/2	0 (0.0)	0 (0.0)	5 (4.0)	
C2/3	2 (1.8)	0 (0.0)	1 (0.8)	
C3/4	11 (10.1)	8 (12.1)	18 (14.4)	
C4/5	19 (17.4)	20 (30.3)	30 (24.0)	
C5/6	47 (43.1)	28 (42.4)	49 (39.2)	
C6/7	30 (27.5)	10 (15.2)	22 (17.6)	
HSI on T2 image, *n* (%)	20 (18.3)	35 (53.0)	90 (72.0)	<0.001
Muhle's classification, *n* (%)				<0.001
Grade I	33 (30.3)	4 (6.1)	4 (3.2)	
Grade II	63 (57.8)	35 (53.0)	50 (40.0)	
Grade III	13 (11.9)	27 (40.9)	71 (56.8)	
K-line (+), *n* (%)	104 (95.4)	52 (78.8)	102 (81.6)	0.003
APB-CMCT[a], *n* (%)				<0.001
Normal	96 (88.1)	19 (28.8)	20 (16.0)	
Mildly delayed	10 (9.2)	24 (36.4)	39 (31.2)	
Definitely delayed	3 (2.8)	17 (25.8)	48 (38.4)	
Not evoked MEP	0 (0.0)	6 (9.1)	18 (14.4)	
Radiculopathy, *n* (%)	29 (26.6)	30 (45.5)	49 (39.2)	0.026

APB, abductor pollicis brevis; ASA, anterior surgical approaches; CMCT, central motor conduction time; HSI, high signal intensity; MEP, motor evoked potential; mJOA, modified Japanese Orthopaedic Association scale; NRS, numerical rating scale of pain; OPLL, ossification of the posterior longitudinal ligament; PSA, posterior surgical approaches.
[a]Normal, CMCT <11.5 ms; mildly delayed, 11.5 ≤CMCT <15; and definitely delayed, CMCT ≥15.

Our variable importance results provide empirical information on the contributing factors that affect the treatment method of DCM at the experienced-neurosurgeon level. In addition, some of the strengths of this study are the creation of various types of classification models and analysis

TABLE 3 Results of multiclass classification.

Metric	RF	XGB
AUC-ROC[a]	0.91	0.92
Overall accuracy (%)	76.2	74.6

AUC-ROC, area under the receiver operating characteristic curve; RF, random forest; XGB, extreme gradient boosting.
[a]Calculated by micro-averaging method.

of the importance of these factors from multiple angles. In our ML models, the mJOA score and CMCT were the most critical factors contributing to the classification of conservative and surgical treatments. Previous studies have also attempted to identify the factors that determine conservative treatment in patients with DCM. Rhee et al. (11) suggested that functional indicators such as the mJOA score, time for a 10-meter walk, and activities of daily living could determine non-operative treatment in patients with DCM through their systematic review, which was consistent with our results. Yoshimatsu et al.'s (30) multivariable logistic regression model showed that the patient's disease duration, one of the top-five important variables for both our models, should primarily be considered when deciding conservative treatment for DCM. Meanwhile, it was reported that CMCT correlated with the degree of spinal cord compression reported in imaging findings (31, 32). CMCT has been known as one of the essential diagnostic tools in myelopathy because it reflects mechanical compression as well as the functional integrity of the motor pathway (33). In

particular, ML models that distinguish the surgical treatment group from the conservative group indicated normal CMCT as the critical feature in our results.

Through ML models, this study also revealed the factors in determining the ASA and PSA in the patient group for which surgical treatment was decided. In 2019, the World Federation of Neurosurgical Societies (WFNS) spine committee recommended the surgical approaches for DCM; they reported essential features to determine between ASA and PSA, such as the number of compression levels, cervical posture, and etiology of myelopathy (34, 35). In our results, it is notable that the functional level or CMCT findings were not important features in determining the approach method in patients who had already decided on surgery. Instead, the number of levels involved was the most important contributing factor in both ML models. Age was the second important variable in the models. In addition, BMI and subjective neck pain were variables with relatively high importance in the classification model between ASA and PSA. We inferred that these could be due to the characteristic advantages and disadvantages of ASA and PSA. One of the advantages of ASA is the reduced postoperative neck pain through muscle-sparing (36). In addition, this method allows for the correction of the kyphotic curvature of the cervical spine and the removal of anterior pathology such as disc herniation (37). However, ASA is preferred only for lesions confined to less than three levels. Moreover, this approach could be more difficult in extremely obese patients (38), and

TABLE 4 Results of binary classifications.

Classification	Algorithm	AUC-ROC	F1	AUC-PR	Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)
Conservative vs. Surgical	RF	0.94	0.93	0.96	89.6	86.7	95.6	72.2
	XGB	0.93	0.93	0.95	89.1	94.1	97.6	76.2
ASA vs. PSA	RF	0.99	0.94	0.96	100.0	88.0	76.9	100.0
	XGB	0.96	0.80	0.82	100.0	80.0	66.7	100.0

ASA, anterior surgical approaches; AUC-PR, area under the precision-recall curve; AUC-ROC, area under the receiver operating characteristic curve; NPV, negative predictive value; PSA, posterior surgical approaches; RF, random forest; XGB, extreme gradient boosting.

TABLE 5 The top five important variables in each model.

Order	Conservative vs. Surgical		ASA vs. PSA	
	Random forest	Extreme gradient boosting	Random forest	Extreme gradient boosting
1	mJOA score	mJOA score	Level count	Level count
2	Normal CMCT	Normal CMCT	Age	Age
3	HSI	Age	NRS, neck	BMI
4	Symptom duration	BMI	BMI	NRS, neck
5	Age	Symptom duration	Disc herniation lesion	Symptom duration

ASA, anterior surgical approaches; BMI, body mass index; CMCT, central motor conduction time; HSI, high signal intensity; mJOA, modified Japanese Orthopaedic Association scale; NRS, numeric rating scale of pain; PSA, posterior surgical approaches; RF, random forest; XGB, extreme gradient boosting.

severe conditions such as vascular or aerodigestive complications may also occur in rare cases (39). In contrast, one of the most significant advantages of PSA is that wide decompression is possible. Furthermore, it is particularly preferred for the treatment of cervical cord compressive lesions across multiple levels—more than three levels (40). This method can preserve the lordosis of the cervical spine; however, it is vulnerable to instability (41). However, despite considering these factors, in our ML model classifying ASA and PSA, the cervical spine alignment or presence of radiculopathy unexpectedly had lower importance. Since this study targeted a group of patients with various degenerative cervical pathologies, we inferred that the number of lesion levels played a crucial role in the decision rather than the presence of cervical kyphosis or radiculopathy. Meanwhile, since we have applied strict criteria for defining k-line (−) or concomitant radiculopathy, their proportion was measured lower than that in other studies (42, 43); therefore, we inferred the possibility that they did not show higher importance. Consequently, it is necessary to consider various factors to determine the treatment direction for patients with DCM (35). The ML models proposed in this study are significant because they can support surgeons in making the complicated decision-making process more readily, efficiently, and tailored.

This study applied ML algorithms that utilize the tree-based ensemble learning method (44). RF is an advanced tree-based ensemble model that combines multiple decision trees by the bagging algorithm and effectively reduces the variance error (45). XGB, one of the most advanced ensemble algorithms using boosting, also shows superior performance compared to other traditional ML algorithms (46). It is a gradient boosting method-based ensemble algorithm with improved scalability and performance (47). These ensemble algorithms are currently the models of choice in many ML-based clinical studies on tubular data analysis (48, 49). Thus, we utilized these two algorithms in our ML processing. In our classification models, RF slightly outperformed XGB.

This study has a few limitations. First, it was a single-center study with relatively small sample size. Therefore, the results have limited generalizability. Second, as a retrospective study, since we performed ML modeling using EHR accumulated over a relatively long duration, some variables might show ambiguity in definitions. Third, we presented a model to classify the first therapeutic decision from DCM diagnosis; however, long-term follow-up data after surgery were not provided.

We applied ML to the therapeutic decision of DCM for the first time. Despite several limitations, this study can be a basis for the generalizable clinical decision model using ML algorithms in the future and referred to as a cornerstone study for achieving this purpose for many researchers. To overcome the limitations and obtain a valid and more generalizable ML model, a large sample dataset from a systematic, multicenter, and multi-ethnic-based study is needed to utilize the advantages of ML. Moreover, the dataset should include long-term outcomes to present the prognostic results of ML-based classification models.

Conclusion

This study presented valid and feasible ML classification models for the selection of therapeutic options in patients with DCM. Our results can provide a rational basis for human experts' clinical decisions and encourage efficient and tailored decision-making. Nevertheless, further large-scale studies are required to develop a more generalizable ML model.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by The institutional review board of Pohang Stroke and Spine Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

DP contributed to the conception, design of the work, data analysis, interpretation, and writing the manuscript. JMC, JWY, DY, MK, and GO helped with the data collection, validation, and investigation. JMC, JWY, DY, MK, and HDK helped with the diagnosis and procedures of patients. HDK edited the manuscript and supervised the entire study. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsurg. 2022.1010420/full#supplementary-material.

References

1. Edwards 2nd CC, Riew KD, Anderson PA, Hilibrand AS, Vaccaro AF. Cervical myelopathy. Current diagnostic and treatment strategies. *Spine J.* (2003) 3:68–81. doi: 10.1016/s1529-9430(02)00566-1

2. Montgomery DM, Brower RS. Cervical spondylotic myelopathy. Clinical syndrome and natural history. *Orthop Clin North Am.* (1992) 23:487–93. doi: 10.1016/S0030-5898(20)31760-0

3. Law Jr MD, Bernhardt M, White 3rd AA. Cervical spondylotic myelopathy: a review of surgical indications and decision making. *Yale J Biol Med.* (1993) 66:165–77.

4. Hejrati N, Moghaddamjou A, Marathe N, Fehlings MG. Degenerative cervical myelopathy: towards a personalized approach. *Can J Neurol Sci.* (2021):1–12. doi: 10.1017/cjn.2021.214

5. Bakhsheshian J, Mehta VA, Liu JC. Current diagnosis and management of cervical spondylotic myelopathy. *Global Spine J.* (2017) 7:572–86. doi: 10.1177/2192568217699208

6. Kadanka Z, Mares M, Bednanik J, Smrcka V, Krbec M, Stejskal L, et al. Approaches to spondylotic cervical myelopathy: conservative versus surgical results in a 3-year follow-up study. *Spine (Phila Pa 1976).* (2002) 27:2205–10; discussion 10-1. doi: 10.1097/01.BRS.0000029255.77224.BB

7. Sampath P, Bendebba M, Davis JD, Ducker TB. Outcome of patients treated for cervical myelopathy. A prospective, multicenter study with independent clinical review. *Spine (Phila Pa 1976).* (2000) 25:670–6. doi: 10.1097/00007632-200003150-00004

8. Fehlings MG, Tetreault LA, Riew KD, Middleton JW, Wang JC. A clinical practice guideline for the management of degenerative cervical myelopathy: introduction, rationale, and scope. *Global Spine J.* (2017) 7:21S–7S. doi: 10.1177/2192568217703088

9. George PL, Bhaveshkumar A, Nauman C, Herbert HE. Cervical spondylotic myelopathy: an updated review. *Neurosurg Cases Rev.* (2021) 4:056. doi: 10.23937/2643-4474/1710056

10. Luo J, Cao K, Huang S, Li L, Yu T, Cao C, et al. Comparison of anterior approach versus posterior approach for the treatment of multilevel cervical spondylotic myelopathy. *Eur Spine J.* (2015) 24:1621–30. doi: 10.1007/s00586-015-3911-4

11. Rhee JM, Shamji MF, Erwin WM, Bransford RJ, Yoon ST, Smith JS, et al. Nonoperative management of cervical myelopathy: a systematic review. *Spine (Phila Pa 1976).* (2013) 38:S55–S67. doi: 10.1097/BRS.0b013e3182a7f41d

12. Kato S, Ganau M, Fehlings MG. Surgical decision-making in degenerative cervical myelopathy - anterior versus posterior approach. *J Clin Neurosci.* (2018) 58:7–12. doi: 10.1016/j.jocn.2018.08.046

13. Benton JA, Weiss BT, Mowrey WB, Yassari N, Wang B, Ramos RG, et al. Association of medicare and medicaid insurance status with increased spine surgery utilization rates. *Spine (Phila Pa 1976).* (2021) 46:E939–E44. doi: 10.1097/BRS.0000000000003968

14. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. *Med.* (2021) 2:642–65. doi: 10.1016/j.medj.2021.04.006

15. Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare[☆,☆☆]. *Injury.* (2022). doi: 10.1016/j.injury.2022.01.046. [Epub ahead of print]

16. Park D, Kim BH, Lee S-E, Kim DY, Kim M, Kwon HD, et al. Machine learning-based approach for disease severity classification of carpal tunnel syndrome. *Sci Rep.* (2021) 11:17464. doi: 10.1038/s41598-021-97043-7

17. Yoo TK, Ryu IH, Choi H, Kim JK, Lee IS, Kim JS, et al. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. *Transl Vis Sci Technol.* (2020) 9:8. doi: 10.1167/tvst.9.2.8

18. Merali ZG, Witiw CD, Badhiwala JH, Wilson JR, Fehlings MG. Using a machine learning approach to predict outcome after surgery for degenerative

cervical myelopathy. *PLoS One.* (2019) 14:e0215133. doi: 10.1371/journal.pone.0215133

19. Hopkins BS, Weber 2nd KA, Kesavabhotla K, Paliwal M, Cantrell DR, Smith ZA. Machine learning for the prediction of cervical spondylotic myelopathy: a post hoc pilot study of 28 participants. *World Neurosurg.* (2019) 127:e436–e42. doi: 10.1016/j.wneu.2019.03.165

20. Muhle C, Wiskirchen J, Weinert D, Falliner A, Wesner F, Brinkmann G, et al. Biomechanical aspects of the subarachnoid space and cervical cord in healthy individuals examined with kinematic magnetic resonance imaging. *Spine (Phila Pa 1976).* (1998) 23:556–67. doi: 10.1097/00007632-199803010-00008

21. Ijima Y, Furuya T, Ota M, Maki S, Saito J, Kitamura M, et al. The K-line in the cervical ossification of the posterior longitudinal ligament is different on plain radiographs and CT images. *J Spine Surg.* (2018) 4:403–7. doi: 10.21037/jss.2018.05.23

22. Cantone M, Lanza G, Vinciguerra L, Puglisi V, Ricceri R, Fisicaro F, et al. Age, height, and sex on motor evoked potentials: translational data from a large Italian cohort in a clinical environment. *Front Hum Neurosci.* (2019) 13:185. doi: 10.3389/fnhum.2019.00185

23. Dumitru D, Amato AA, Zwarts MJ. *Electrodiagnostic medicine.* 2nd ed Philadelphia: Hanley & Belfus (2002).

24. Kuhn M. Caret: classification and regression training. R package version 6.0-90. https://CRAN.R-project.org/package=caret. (2021).

25. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. R package version 1.5.0.2. https://CRAN.R-project.org/package=xgboost. (2021).

26. Rigatti SJ. Random forest. *J Insur Med.* (2017) 47:31–9. doi: 10.17849/insm-47-01-31-39.1

27. Wei R, Wang J. MultiROC: calculating and visualizing ROC and PR curves across multi-class classifications. R package version 1.1.1. https://CRAN.R-project.org/package=multiROC. (2018).

28. John CR. MLeval: machine learning model evaluation. R package version 0.3. https://CRAN.R-project.org/package=MLeval. (2020).

29. Ganau M, Holly LT, Mizuno J, Fehlings MG. Future directions and new technologies for the management of degenerative cervical myelopathy. *Neurosurg Clin N Am.* (2018) 29:185–93. doi: 10.1016/j.nec.2017.09.006

30. Yoshimatsu H, Nagata K, Goto H, Sonoda K, Ando N, Imoto H, et al. Conservative treatment for cervical spondylotic myelopathy. Prediction of treatment effects by multivariate analysis. *Spine J.* (2001) 1:269–73. doi: 10.1016/s1529-9430(01)00082-1

31. Funaba M, Imajo Y, Suzuki H, Nagao Y, Sakamoto T, Nishida N, et al. Radiological factors associated with the severity of corticospinal tract dysfunctions for cervical spondylotic myelopathy: an analysis of the central motor conduction time and kinematic CT myelography. *J Clin Neurosci.* (2021) 94:24–31. doi: 10.1016/j.jocn.2021.09.032

32. Park D, Kim BH, Cho JM, Yang JW, Yang DH, Kim MS, et al. Diagnostic role of flexion-extension central motor conduction time in cervical spondylotic myelopathy. *Spine (Phila Pa 1976).* (2021) 46:1564–71. doi: 10.1097/BRS.0000000000003706

33. Nardone R, Holler Y, Thomschewski A, Holler P, Bergmann J, Golaszewski S, et al. Central motor conduction studies in patients with spinal cord disorders: a review. *Spinal Cord.* (2014) 52:420–7. doi: 10.1038/sc.2014.48

34. Deora H, Kim SH, Behari S, Rudrappa S, Rajshekhar V, Zileli M, et al. Anterior surgical techniques for cervical spondylotic myelopathy: WFNS spine committee recommendations. *Neurospine.* (2019) 16:408–20. doi: 10.14245/ns.1938250.125

35. Bajamal AH, Kim SH, Arifianto MR, Faris M, Subagio EA, Roitberg B, et al. Posterior surgical techniques for cervical spondylotic myelopathy: WFNS spine

committee recommendations. *Neurospine*. (2019) 16:421–34. doi: 10.14245/ns.1938274.137

36. Memtsoudis SG, Hughes A, Ma Y, Chiu YL, Sama AA, Girardi FP. Increased in-hospital complications after primary posterior versus primary anterior cervical fusion. *Clin Orthop Relat Res*. (2011) 469:649–57. doi: 10.1007/s11999-010-1549-4

37. Wang JC, McDonough PW, Endow KK, Delamarter RB. Increased fusion rates with cervical plating for two-level anterior cervical discectomy and fusion. *Spine (Phila Pa 1976)*. (2000) 25:41–5. doi: 10.1097/00007632-200001010-00009

38. Klineberg E. Cervical spondylotic myelopathy: a review of the evidence. *Orthopedic Clinics of North America*. (2010) 41:193–202. doi: 10.1016/j.ocl.2009.12.010

39. Yee TJ, Swong K, Park P. Complications of anterior cervical spine surgery: a systematic review of the literature. *J Spine Surg*. (2020) 6:302–22. doi: 10.21037/jss.2020.01.14

40. Ito M, Nagahama K. Laminoplasty for cervical myelopathy. *Global Spine J*. (2012) 2:187–93. doi: 10.1055/s-0032-1315456

41. Abdullah KG, Yamashita T, Steinmetz MP, Lubelski D, Wang JC, Benzel EC, et al. Open-door cervical laminoplasty with preservation of posterior structures. *Global Spine J*. (2012) 2:015–20. doi: 10.1055/s-0032-1307258

42. Choi BW, Kim SS, Lee DH, Kim JW. Cervical radiculopathy combined with cervical myelopathy: prevalence and characteristics. *Eur J Orthop Surg Traumatol*. (2017) 27:889–93. doi: 10.1007/s00590-017-1972-2

43. Fujiyoshi T, Yamazaki M, Kawabe J, Endo T, Furuya T, Koda M, et al. A new concept for making decisions regarding the surgical approach for cervical ossification of the posterior longitudinal ligament: the K-line. *Spine (Phila Pa 1976)*. (2008) 33:E990–3. doi: 10.1097/BRS.0b013e318188b300

44. Jafarzadeh H, Mahdianpari M, Gill E, Mohammadimanesh F, Homayouni S. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. *Remote Sens (Basel)*. (2021) 13:4405. doi: 10.3390/rs13214405

45. Couronne R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*. (2018) 19:270. doi: 10.1186/s12859-018-2264-5

46. Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. *Front Comput Sci*. (2019) 14:241–58. doi: 10.1007/s11704-019-8208-z

47. Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM. Extreme gradient boosting as a method for quantitative structure–activity relationships. *J Chem Inf Model*. (2016) 56:2353–60. doi: 10.1021/acs.jcim.6b00591

48. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, et al. A machine-learning-based prediction method for hypertension outcomes based on medical data. *Diagnostics*. (2019) 9:178. doi: 10.3390/diagnostics9040178

49. Park D, Kim I. Application of machine learning in the field of intraoperative neurophysiological monitoring: a narrative review. *Appl Sci*. (2022) 12:7943. doi: 10.3390/app12157943