Cai *et al. Journal of Orthopaedic Surgery and Research* (2024) 19:539
https://doi.org/10.1186/s13018-024-05004-3

Journal of Orthopaedic Surgery and Research

RESEARCH ARTICLE **Open Access**

Machine-learning-based prediction by stacking ensemble strategy for surgical outcomes in patients with degenerative cervical myelopathy

Zhiwei Cai[1†], Quan Sun[2†], Chao Li[2†], Jin Xu[2*] and Bo Jiang[2*]

Abstract

Background Machine learning (ML) is extensively employed for forecasting the outcome of various illnesses. The objective of the study was to develop ML based classifiers using a stacking ensemble strategy to predict the Japanese Orthopedic Association (JOA) recovery rate for patients with degenerative cervical myelopathy (DCM).

Methods A total of 672 patients with DCM were included in the study and labeled with JOA recovery rate by 1-year follow-up. All data were collected during 2012–2023 and were randomly divided into training and testing (8:2) sub-datasets. A total of 91 initial ML classifiers were developed, and the top 3 initial classifiers with the best performance were further stacked into an ensemble classifier with a supported vector machine (SVM) classifier. The area under the curve (AUC) was the main indicator to assess the prediction performance of all classifiers. The primary predicted outcome was the JOA recovery rate.

Results By applying an ensemble learning strategy (e.g., stacking), the accuracy of the ML classifier improved following combining three widely used ML models (e.g., RFE-SVM, EmbeddingLR-LR, and RFE-AdaBoost). Decision curve analysis showed the merits of the ensemble classifiers, as the curves of the top 3 initial classifiers varied a lot in predicting JOA recovery rate in DCM patients.

Conclusions The ensemble classifiers successfully predict the JOA recovery rate in DCM patients, which showed a high potential for assisting physicians in managing DCM patients and making full use of medical resources.

Keywords Machine learning, Ensemble learning, Degenerative cervical myelopathy, Prognostic prediction, Japanese orthopedic association

†Zhiwei Cai, Quan Sun and Chao Li contributed equally to this work.

*Correspondence:
Jin Xu
516289565@qq.com
Bo Jiang
jianbohbu@163.com
[1]Department of Orthopedics, The Forth Medical Center of Chinese PLA General Hospital, Beijing 100142, China
[2]Department of Orthopedics, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang 441000, Hubei, China

Introduction

Degenerative cervical myelopathy (DCM) is frequently encountered in clinical settings and is characterized by the acquired narrowing of the spinal canal leading to non-traumatic injury to the spinal cord [1]. Currently, the main treatment for DCM is decompression surgery for the spinal canal [2]. Timely and effective decompression surgery can halt the deterioration of neurological function, resulting in an improvement in surgical outcomes for individuals with DCM [2, 3]. Despite surgery to decompress the cervical spinal canal, certain patients may still experience neurological deficits following the procedure [4, 5]. Prediction of postoperative outcomes in DCM patients can aid clinicians in making informed decisions and creating customized rehabilitation plans, thereby avoiding unnecessary surgeries for individuals with a high risk of unfavorable outcomes [4].

With the development of machine learning (ML) algorithms, researchers have been developing ML-based predictive models for DCM. While ML has been thoroughly explored in the context of medical diagnostics and imaging, its application to epidemiological datasets for predicting various health outcomes represents a recent advancement [6–8]. It has been emphasized that ML has various benefits in contrast to statistical models, indicating its capacity to manage extensive datasets and discern nonlinear connections between potential predictors and observational outcomes [7].

In the past years, research has developed numerous ML-based prognostic prediction models for DCM. However, previous studies have primarily concentrated on comparing prediction accuracy among various models and selecting a ML model exhibited the highest prediction performance as the main reported result. Ensemble learning is an ML technique that involves combining the predictions of multiple models to create a stronger, more robust model and offers several advantages over individual machine learning models, including improved accuracy, reduced overfitting, increased robustness, versatility across algorithms, and enhanced stability. Therefore, in our current study, our primary goal is to develop a predictive model to predict postoperative outcomes in DCM patients via ML-based ensemble learning.

Table 1 Clinical characteristics and outcomes of the 476 DCM Patient Cohort

Variable	Mean ± standard deviation or frequency (proportion)	Range(minimum-maximum)
Age(yr)	61.2 ± 13.17	35–83
Sex (Female)	223(44.8)	N/A
Preoperative JOA	9.30 ± 2.03	6–13
Axial pain intensity	4.34 ± 2.45	0–8
ISI (Occurrence)	267(56.1)	N/A
MSCC	1.56 ± 0.80	0–3
BMI	10.77 ± 1.85	5.0–15.0
Neutrophil	2.08 ± 1.09	0.77–8.35
Lymphocyte	2.02 ± 0.74(*10^9)	0.89–5.25(*10^9)
NLR	3.74 ± 1.27(*10^9)	1.38–8.6(*10^9)
Surgical Approach:		
Anterior discectomy	37(7.7)	N/A
Anterior corpectomy	38(7.8)	N/A
Anterior fusion	71(15.0)	N/A
Anterior fixation	72(15.1)	N/A
Laminectomy	146(30.1)	N/A
Laminectomy + Fusion	52(10.9)	N/A
Laminoplasty	61(12.8)	N/A
Level of operation:		
Operation at C1	12(1.9)	N/A
Operation at C2	13(2)	N/A
Operation at C3	32(5)	N/A
Operation at C4	203(33.4)	N/A
Operation at C5	131(21.6)	N/A
Operation at C6	125(20.1)	N/A
Operation at C7	91(14.9)	N/A

Mean ± Standard Deviation is Used to Represent Continuous Variables While Frequency and Proportion (%) Are Used to Represent Categorical Variables. DCM: Degenerative cervical myelopathy; NLR: neutrophil-to-lymphocyte ratio; ISI: increased signal intensity; MSCC: maximum spinal cord compression; JOA: Japanese Orthopedic Association;

Materials and methods

Study design and patient cohort

The dataset for our current study was retrospectively collected from the Orthopedic department at Xiangyang Central Hospital between 2012 and 2023. It comprised 672 patients who received surgical decompression due to symptomatic DCM. The study received approval from the ethics board, and the research was carried out in accordance with ethical guidelines. Patients were deemed eligible for participation upon furnishing written informed consent and satisfying the specified criteria: (1) symptomatic DCM exhibiting a minimum of one clinical sign of myelopathy; (2) imaging that verifies compression of the cervical cord; (3) no prior surgery for DCM; and (4) 18 years of age or older.

Baseline data and predicted outcomes

Machine learning models were trained using clinical measurements, such as age, gender, comorbidities, and other relevant factors (Table 1). The Japanese Orthopedic Association (JOA) score served as the metric for assessing functional status preoperatively. Two senior spine surgeons determined the JOA score to evaluate the severity of neurological symptoms [9], and the average JOA scores were employed for subsequent analyses (Table 2). Additionally, patients' JOA scores were also assessed one-year post-surgery. The Hirabayashi method was utilized to calculate the JOA recovery rate (JOARR).

Table 2 Comparison of the Japanese Orthopedic Association scores (JOA) and JOA recovery rate calculated from two senior spine surgeons

Model	Surgeon 1	Surgeon 2	Paired T values	*P* values
Preoperative JOA	9.3 ± 1.8	9.3 ± 2.1	0.19	0.84
Postoperative JOA	14.3 ± 2.7	14.2 ± 2.7	0.53	0.59
JOA recovery rate	65.4%±21.3%	66.1%±20.7%	-0.16	0.87

$$JOARR = \frac{\text{postoperative JOA score} - \text{preoperative JOA score}}{17 - \text{preoperative JOA score}} \times 100\%$$

Patients were categorized based on their JOA recovery rate (JOARR) into two groups: individuals scoring below 60% on JOARR were assigned a score of 1 (indicating poor recovery) [10]. In contrast, individuals scoring above 60% on JOARR were given a score of 0 (indicating good recovery) [11]. We employed this categorical variable as the dependent variable in constructing machine learning classifiers. There are three reasons for converting this continuous variable into a binary one: (1) Considering our limited sample size, binary variables generally exhibit a narrower range of values, which facilitates model simplification; (2) Transforming continuous variables into binary ones is an effective way to handle outliers; (3) Continuous variables may be susceptible to the impact of noise or measurement errors. Converting continuous variables into binary ones aids in mitigating the influence of noise present in the data, leading to increased robustness and stability of the model.

Model development

We adhered to the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) checklist [12] and guidelines for the analysis of machine learning predictive models [13]. The analyses pipeline could be found in Fig. 1. Data preprocessing involved the removal of patients without follow-up JOA scores, leaving a total of 476 patients. A total of seven commonly used feature-selection methods were adopted including maximal information coefficient (MIC), embedding logistic regressor (embedding LR), embedding linear supported vector classifier (embedding LSVC), embedding random forest (RF), embedding tree, minimal-redundancy-maximal-relevance (mRMR), and recursive feature elimination (RFE).

Thirteen ML algorithms were employed including linear discriminant analysis (LDA), gradient boosting, adaptive boosting (AdaBoost), multilayer perceptron (MLP), deep neural network (DNN), supported vector machine (SVM), Gaussian naïve Bayes (NB), decision tree, logistic regression, random forest (RF), bagging, extra tree, and K-Nearest Neighbor (KNN). The rationale behind the choice of ML algorithms and feature-selection methods was based on a previous study [14]. This strategy included most used methods for ML analyses for developing the prognostic prediction model in clinical practice.

Moreover, probability estimates are not provided by many machine learning algorithms in contrast to logistic regression models. Platt scaling was utilized to transform the less interpretable output scores of the model into probabilities to tackle the issue. Consequently, a sum of 91 initial classifiers was generated from the 7×13 combinations. The data used for this study were divided into

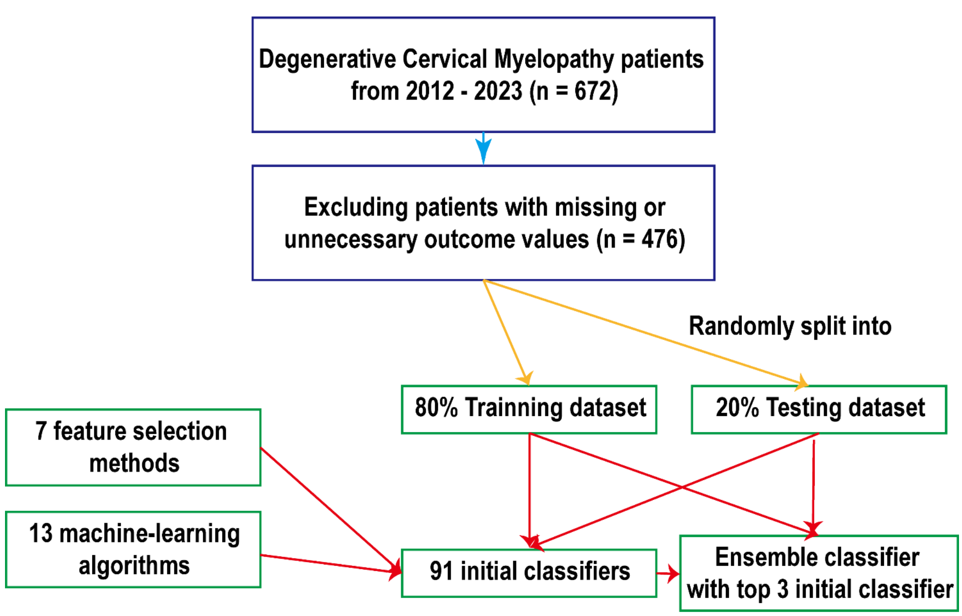

Fig. 1 Flowchart of the analyses pipeline for the current study

three sub-datasets, with a ratio of 8:2 for training, and testing respectively. Each initial classifier underwent three repetitions of 10-fold cross-validation and independent testing using the training and testing dataset. The detailed procedures for cross-validation were as follows: (1) Data Splitting: The entire training dataset was divided into ten subsets of approximately equal size. (2) Training-Validation Splits: The classifier was trained and validated ten times, each time using a different combination of nine subsets for training and the remaining one subset for validation. This process was repeated until each subset had been used as a validation set exactly once. (3) Performance Calculation: The performance of the classifier was then calculated as the average performance across the ten folds. This approach helps mitigate the potential bias introduced by a single partition of the data. (4) Independent Testing: Following the cross-validation, the final evaluation involved testing the classifier on the testing subset not used during the training or validation phases. By employing 10-fold cross-validation and repeating the process three times, we aimed to ensure a thorough and reliable assessment of the performance of our initial classifiers. More importantly, the hyperparameters for the ML models were tuned using a grid search strategy during cross-validation and the detailed information for the hyperparameters that were tuned, the range of values considered could be found in supplementary materials Table 1.

Various discrimination parameters were employed to evaluate the model's performance on the training set. The ability of the model to differentiate between patients who experienced functional recovery in JOA and those who did not was assessed using these metrics. Area under the curve (AUC) or area under the receiver operating characteristic curve (ROC), accuracy, sensitivity, and specificity were encompassed in the metrics. The AUC was employed as the metric to evaluate the performance of all classifiers. The top performers were identified by selecting the three initial classifiers with the highest average AUC during cross-validation. These top three classifiers were subsequently stacked into an ensemble classifier using an SVM classifier. In the stacked ensemble model, the performance measurement is assessed through a meta-classifier (e.g., ensembled classifier) based on the combination of the predictions from the base classifiers (e.g., Top 3 classifiers). The process involves the following steps: Base Classifiers: The top 3 classifiers (including the SVM) generate individual predictions for the validation and testing dataset. Stacking: The predictions for the validation dataset from the base classifiers are combined or stacked to form a new dataset to train the meta-classifier (in this case, the SVM). This meta-classifier learns to make predictions based on the outputs of the top 3 classifiers and was then tested on the testing set using the predictions for the testing-set as features. Performance Measurement: The performance of the stacked model is then assessed using accuracy, and AUC.

Results

Clinical characteristics

In our current study, the average age of the patients included was 58.5 years, and 54.3% of them were male. The mean pre-JOA score was 9.3 and the average preoperative axial pain intensity was 4.3 at baseline. Other measures such as neutrophil-to-lymphocyte ratio, occurrence of increased signal intensity in the spinal cord and maximum spinal cord compression of the spinal cord were also calculated and showed in Table 1. Furthermore, to ensure that the JOA recovery rate was not influenced by measurement errors educed by a single measurer, the preoperative and postoperative JOA scores were obtained from two spine surgeons and compared. No significant differences in terms of preoperative JOA score ($P=0.84$), postoperative JOA score ($P=0.59$) and JOA recovery rate ($P=0.87$) were observed between two spine surgeons (Table 2).

Machine learning prediction model performances

In the process of developing predictive machine learning models, a set of 23 potential features was incorporated. In our study, RFE-SVM, Embedding LR-logistic, and RFE-AdaBoost were identified as the three initial classifiers with the highest average AUC during cross-validation (Fig. 2.A-B). The AUC for the three models are: 0.78 for Embedding-LR, 0.79 for RFE-SVM and 0.81 for RFE-AdaBoost. Furthermore, in independent testing, the ensemble classifier for predicting JOA recovery rate in DCM patients exhibited a superior AUC of 0.92 (Fig. 3.A) compared to that of the initial classifiers (AUC was 0.796, 0.799, 0.802 during independent testing, Table 3). Furthermore, the performance metrics for each individual classifier in the ensemble model were shown in Table 4. The Hosmer-Lemeshow tests were also performed for statistical analyses to assess the agreement between the predicted and observed probabilities of the outcome and the results could be found in supplementary materials Table 2. Decision curve analysis highlighted the advantages of the ensemble classifiers, with significant variations in the curves of the top 3 initial classifiers when predicting surgical outcomes in DCM patients (Fig. 3.B).

Feature importances

Permutation importance was utilized to rank the top 10 features for the ensemble classifier. The top 5 important features for the ensemble classifier were "preoperative JOA scores," "Age," "Smoking status," "Duration," and "T2 ISI" (Fig. 3.C).

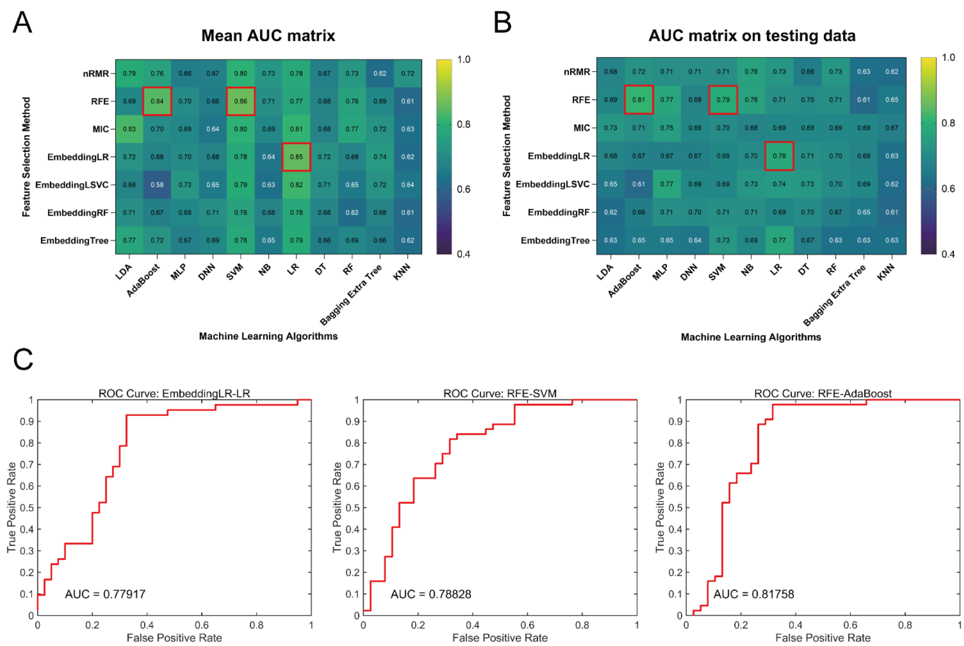

Fig. 2 Prediction performance of JOA recovery rate in degenerative cervical myelopathy patients. A: AUC for all initial classifiers during cross-validation; B: AUC for all initial classifiers during independent testing; C: ROC curves for the initial classifiers with Top 3 predictive performance

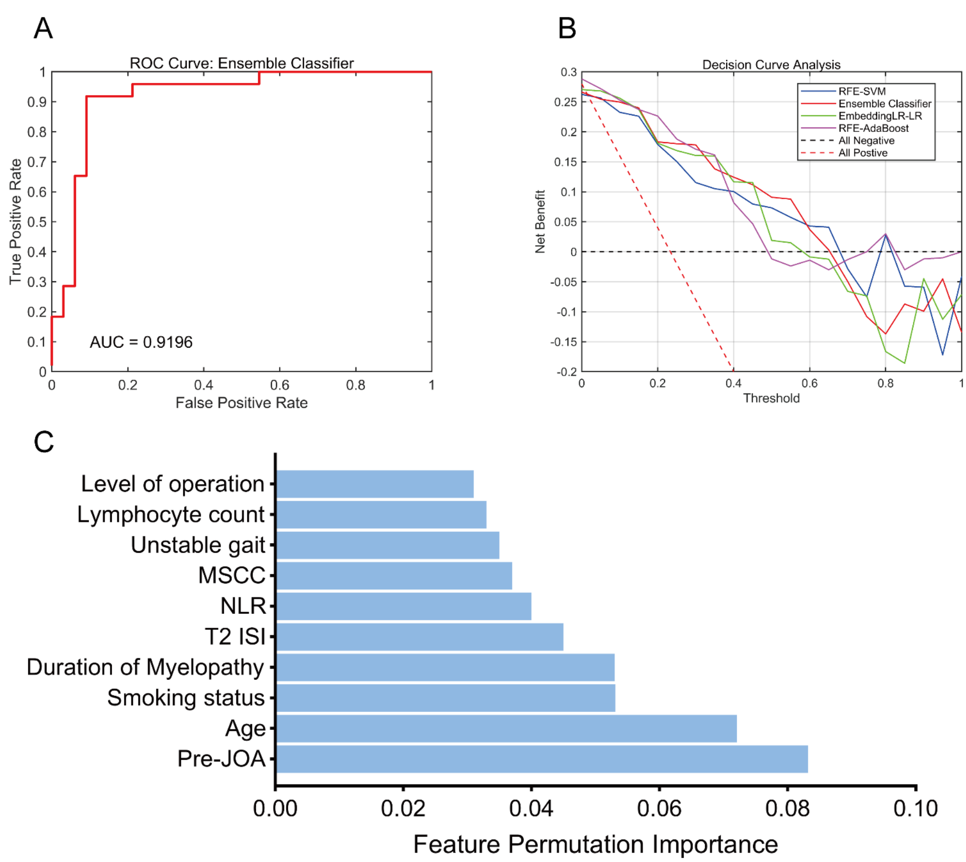

Fig. 3 Decision curve analysis and feature permutation importance. A: Decision Curve Analysis (DCA) for JOA recovery rate in degenerative cervical myelopathy patients; B: ROC curve for ensemble classifier; C: Top 10 Features of the ensemble classifiers for JOA recovery rate prediction. JOA: Japanese Orthopedic Association; ISI: increased signal intensity, NLR: Neutrophil to lymphocyte ratio

Table 3 Comparison of the AUCs, accuracies, sensitivities, and specificities of initial classifiers and ensemble classifier

Model	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
RFE-SVM	0.79	73.2	64.8	87.7
RFE-AdaBoost	0.81	84.8	68.4	93.7
EmbeddingLR- LR	0.78	79.3	65.0	92.8
Ensemble Classifier	0.92	90.2	90.1	89.8

RFE: recursive feature elimination; EmbeddingLR: embedding logistic regressor; SVM: supported vector machine; LR: logistic regressor

Table 4 The AUCs, accuracies, sensitivities, and specificities of each individual classifier in the ensemble model

Model	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
Internal validation				
RFE-SVM	0.78	71.8	63.2	88.2
RFE-AdaBoost	0.89	83.2	67.4	93.4
EmbeddingLR- LR	0.81	80.1	71.8	94.1
Independent testing				
RFE-SVM	0.79	71.2	65.6	88.6
RFE-AdaBoost	0.83	82.6	67.4	90.1
EmbeddingLR- LR	0.77	77.3	67.0	91.8

RFE: recursive feature elimination; EmbeddingLR: embedding logistic regressor; SVM: supported vector machine; LR: logistic regressor

Discussion

In the current study, three main findings were observed: (1) Machine learning classifiers could successfully identify DCM patients with poor surgical outcomes preoperatively; (2) By applying an ensemble learning strategy (e.g., stacking), the predictive performance of the ML classifier improved following combining three widely used ML models (e.g., RFE-SVM, EmbeddingLR-logistic, and RFE-AdaBoost); (3) Finally, preoperative JOA scores, Age, Smoking status, Myelopathy duration and Increased T2 intensity on spinal cord were identified as the most important clinical features for poor clinical outcomes following decompression surgery in DCM patients.

Predicting the surgical outcomes, which is frequently assessed by JOA recovery rate, has been a longstanding concern in the field of spinal surgery. Early identification of patients with poor postoperative recovery has significant implications. Firstly, promptly recognizing patients struggling to recover provides physicians with crucial information to adjust care plans accordingly. This includes refining rehabilitation protocols, managing patient expectations regarding prognosis and feasibility of improvement, and deploying psychological resources when necessary. Accurately predicting patient trajectories enables the optimization of intervention efficacy and the reduction of unnecessary medical costs and emotional distress. Over the past few decades, researchers have been dedicated to developing clinical prediction models for forecasting the prognosis of DCM. Creating a predictive algorithm for DCM that assesses functionality has the potential to enhance clinical care efficiency and profoundly influence patient management [15]. A precise prediction model would enable spine surgeons to identify patients with an increased risk of experiencing deteriorating functional outcomes after decompression surgery. Early identification could support positive interventions, including specific preventive interventions, aimed at improving functional outcomes in patients. Additionally, personalized treatment plans can be created by healthcare providers based on the unique risk profile of each patient. A dependable predictive model would enable surgeons to have informed conversations with patients regarding their prognosis and potential risks. To address these issues, several researches have employed machine learning methods to develop predictive model. Qmar et al. applied a polynomial support vector machine with default parameters (utilizing a training sample size of 561) to predict the poorer postoperative functional condition in patients with degenerative cervical myelopathy, with an accuracy rate of 74.3% and an AUC of 0.78. Their findings surpassed those of previous studies (refer to [16]) for details), in which Zamir G et al. utilized a random forest framework, obtained an average AUC of 0.70, a classification accuracy of 77%, and a sensitivity of 78% [17]. Using XGBoost, Satoshi showcased the highest AUC (0.72) and a substantial accuracy (67.8%) in predicting surgical outcomes 1-year postoperatively [18]. However, in their studies, the performance of the widely used machine learning models were not compared directly. In comparison to these results, our current study systematically examined and compared commonly used ML algorithms for developing predictive models in DCM patients for predicting the JOA recovery rate. More importantly, we incorporate several feature-selection approaches to improve the predictive power of our ML models.

Moreover, in our present study, the three initial classifiers with the highest average AUC were further stacked into an ensemble classifier using an SVM classifier. Employing ensemble learning provides critical advantages that enhance predictive performance beyond individual models. Combining various algorithms serves to alleviate their inherent limitations through complementarity, leveraging the power of diversity to reduce collective blindness. Singular models often succumb to overfitting to noise, but ensembles counteract such idiosyncrasies by filtering information from spurious artifacts. Furthermore, integrating diverse perspectives guards against fixating on local optima. Singular models easily become trapped at suboptimal solutions; ensemble escape relies on divergence. Fan G et al. constructed a predictive model aimed at predicting extended stays in the intensive care unit (ICU) and prolonged hospital stays among patients with spinal cord injury. By applying a resemble learning approach, they enhanced the AUC from 0.799 to 0.802 [14]. Likewise, through the stacking of the top 3 predictive models, we observed an increase in the AUC from 0.81 to 0.92. The final model's performance was enhanced, indicating the capability of ensemble learning to improve classification accuracy. This enhancement lays the groundwork for the subsequent application and practical use of the model in clinical scene. To our knowledge, our study is the first to develop the ML-based prediction model using stacking-ensemble approach in DCM population.

Additionally, to determine the importance of features in the final ensemble model, we calculated the relative importance of each feature. The application of machine learning in assessing feature importance presents various advantages [19–23]. By algorithmically evaluating the significance of predictors, it eliminates subjective bias that may arise from manual selection. Moreover, this data-driven prioritization efficiently handles large datasets with numerous attributes, effectively identifying the most informative features. Furthermore, by identifying the risk factors contributing to outcomes, this methodology provides valuable scientific insights by providing potential causal mechanisms and guiding future research directions through the identification of high-value variables.

Our analysis has identified age, gender, disease duration, and preoperative neurological status as most predictive features, aligning with reported predictors of DCM outcomes. [2, 7, 8, 24]. In a study conducted by Lindsay A. Tetreault et al., gender, preoperative function, and disease duration were also identified as pivotal factors. It is noteworthy that this study has unveiled that advanced age is linked to poorer outcomes, particularly in elderly patients, even though most surgeons do not tailor treatment depending on age. Nevertheless, surgeons should be aware that elderly individuals may not attain equivalent functional improvement compared to their younger counterparts, even in the presence of neurological recovery, due to factors such as age-related spinal cord changes or comorbidities [25–28]. Additionally, our findings have highlighted other significant predictors, namely the heightened T2-signal intensity and neutrophil-lymphocyte ratio (NLR). The NLR serves as an inflammatory marker that encompasses ratios of immune cells and has been utilized in assessing inflammation and predicting outcomes in conditions such as spinal cord injury. The increased support garnered by the NLR implies its potential effectiveness in predicting outcomes for DCM. The investigation has explored the correlation between NLR and outcomes of spinal injury, wherein injury disrupts the blood-spinal cord barrier, allowing infiltration of immune cells and initiating inflammation [29–31]. Likewise, individuals with DCM exhibit disruption of the blood-spinal cord barrier at sites of compression, potentially triggering similar neuroinflammatory mechanisms [32–34]. Concerning intramedullary signal intensity (ISI), the existence of heightened T2-weighted magnetic resonance imaging signal is commonly identified in DCM cases, indicating the occurrence of either reversible or irreversible spinal cord alterations due to compression. Numerous studies have thoroughly investigated the prognostic significance of this phenomenon using various classification frameworks [35–37]. Importantly, a recently developed cervical myelopathy MRI classification system (Ax-CCM) was introduced by You et al., relying on axial images. This system identifies a specific ISI subtype associated with unfavorable clinical outcomes [11]. We utilized Ax-CCM to classify ISI subtypes, thereby offering insights into varying recovery capacities and predicting DCM outcomes. In summary, our comprehensive analysis not only reaffirmed established predictors but also revealed novel prognostic determinants such as NLR and ISI subtypes, enhancing the accuracy of DCM prediction. As for clinical implications, our current findings.

Limitations

Our study is subject to several limitations that warrant discussion. Firstly, the relatively small sample size drawn from a single medical center and limited to a specific ethnic group may restrict the generalizability of our findings. Future investigations should aim to include larger and more diverse cohorts to validate and extend our results across different populations. It should be noted that in our current study, to minimize the impact of small sample sizes as much as possible, we have not included a large number of features. We also illustrated the AUC of all models during model training to make sure the model is not underfitted. Therefore, the models in this study possess a certain level of reliability. Secondly, the retrospective nature of our research introduces inherent limitations, including the absence of supplementary clinical evaluations that could impact post-surgery outcomes in patients with DCM. This retrospective design also raises concerns about selection bias and uncontrolled confounding variables. To address these limitations, future studies should consider employing a prospective study design that incorporates a broader range of clinical parameters to provide a more comprehensive understanding of DCM outcomes. Additionally, our analysis was limited to axial images of cervical magnetic resonance imaging (MRI), omitting other valuable imaging modalities such as sagittal images and advanced techniques like Diffusion Spectrum Imaging (DSI), Diffusion Tensor Imaging (DTI), and functional MRI (fMRI). Integrating these additional imaging modalities could offer deeper insights into the pathophysiology of DCM and should be considered in future research endeavors.

Having said this, these features could optimize the predictive accuracy of the model to a certain extent. However, current results indicate that using conventional MRI indicators can also predict prognosis. Lastly, our study primarily focused on the Japanese Orthopaedic Association (JOA) score as a measure of neurological function, overlooking a comprehensive assessment of frailty. Given the importance of frailty in guiding patient management and expectations, future research should prioritize its inclusion and explore its impact on DCM outcomes in greater detail.

Conclusion

Our results indicate that utilizing machine learning classifiers, like support vector machines (SVM), is proficient in foreseeing surgical outcomes in DCM patients. Simultaneously, it enables the identification of associated predictors through a multivariate analysis.

Abbreviations

ML	Machine learning
JOA	Japanese Orthopedic Association
DCM	Degenerative cervical myelopathy
AUC	Area under the curve
SVM	Supported vector machine
JOARR	Japanese Orthopedic Association recovery rate
MIC	Maximal information coefficient
LR	Logistic regressor
LSVC	Linear supported vector classifier
RF	Random forest
mRMR	Minimal-redundancy-maximal-relevance
RFE	Recursive feature elimination
AdaBoost	Adaptive boosting
MLP	Multilayer perceptron
DNN	Deep neural network
NB	Naïve Bayes
KNN	K-Nearest Neighbor
ROC	Receiver operating characteristic curve
PPV	Positive predictive value
NPV	Negative predictive value
ISI	Intramedullary signal intensity
NLR	Neutrophil-lymphocyte ratio
ICU	Intensive care unit

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13018-024-05004-3.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

CZW designed the study and wrote the manuscript. SQ and CZW participated in the collection of data and data statistics. XJ, CL and JB revised the manuscript. All authors read and approved the final manuscript.

Funding

This study has received no funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of the local medical center. Written informed consent was obtained from each participant prior each procedure.

Consent for publication

The authors affirm that human research participants provided informed consent for publication of the images in all Figures.

Competing interests

The authors declare no competing interests.

Received: 14 January 2024 / Accepted: 16 August 2024
Published online: 04 September 2024

References

1. Nurick S. The pathogenesis of the spinal cord disorder associated with cervical spondylosis. Brain. 1972;95(1):87–100.
2. Fehlings MG, 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. 2013;95(18):1651–8.
3. Vidal PM et al. Delayed decompression exacerbates ischemia-reperfusion injury in cervical compressive myelopathy. JCI Insight. 2017;2(11).
4. Karadimas SK, et al. Riluzole blocks perioperative ischemia-reperfusion injury and enhances postdecompression outcomes in cervical spondylotic myelopathy. Sci Transl Med. 2015;7(316):316ra194.
5. Fehlings MG, et al. Perioperative and delayed complications associated with the surgical treatment of cervical spondylotic myelopathy based on 302 patients from the AOSpine North America Cervical Spondylotic Myelopathy Study. J Neurosurg Spine. 2012;16(5):425–32.
6. Khan O, et al. Predictive modeling of outcomes after traumatic and nontraumatic spinal cord Injury using machine learning: review of current progress and future directions. Neurospine. 2019;16(4):678–85.
7. Khan O, et al. Prediction of worse functional status after surgery for degenerative cervical myelopathy: a Machine Learning Approach. Neurosurgery. 2021;88(3):584–91.
8. Khan O, et al. Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy. Spine J. 2021;21(10):1659–69.
9. Furlan JC, Catharine B, Craven. Psychometric analysis and critical appraisal of the original, revised, and modified versions of the Japanese Orthopaedic Association score in the assessment of patients with cervical spondylotic myelopathy. Neurosurg Focus. 2016;40(6):E6.
10. Hirabayashi K, et al. Operative Results and Postoperative Progression of Ossification Among Patients With Ossification of Cervical Posterior Longitudinal Ligament. Spine. 1981, 6(4): 354-64.
11. You JY, et al. MR classification system based on axial images for cervical compressive myelopathy. Radiology. 2015;276(2):553–61.
12. Collins GS, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
13. Luo W, et al. Guidelines for developing and reporting machine learning predictive models in Biomedical Research: a multidisciplinary view. J Med Internet Res. 2016;18(12):e323.
14. Fan G, et al. Machine learning-based prediction of prolonged Intensive Care Unit stay for critical patients with spinal cord Injury. Spine (Phila Pa 1976). 2022;47(9):E390–8.
15. Khan O, et al. Use of Machine Learning and Artificial Intelligence to drive Personalized Medicine approaches for Spine Care. World Neurosurg. 2020;140:512–8.
16. Saravi B et al. Artificial Intelligence-Driven prediction modeling and decision making in spine surgery using hybrid machine learning models. J Pers Med. 2022;12(4).
17. Merali ZG, et al. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. PLoS ONE. 2019;14(4):e0215133.

18. Maki S, et al. Machine Learning Approach in Predicting clinically significant improvements after surgery in patients with cervical ossification of the posterior longitudinal ligament. Spine (Phila Pa 1976). 2021;46(24):1683–9.
19. Alfraihat A, Samdani AF, Balasubramanian S. Predicting radiographic outcomes of vertebral body tethering in adolescent idiopathic scoliosis patients using machine learning. PLoS ONE. 2024;19(1):e0296739.
20. Gholizadeh M et al. Machine learning-based prediction of effluent total suspended solids in a wastewater treatment plants using different feature selection approaches: a comparative study. Environ Res, 2024: p. 118146.
21. Li Q, et al. Identification of diagnostic signatures for ischemic stroke by machine learning algorithm. J Stroke Cerebrovasc Dis. 2024;33(3):107564.
22. Rezvantalab S, Mihandoost S, Rezaiee M. Machine learning assisted exploration of the influential parameters on the PLGA nanoparticles. Sci Rep. 2024;14(1):1114.
23. Zhang R, et al. Automated machine learning for early prediction of acute kidney injury in acute pancreatitis. BMC Med Inf Decis Mak. 2024;24(1):16.
24. Fehlings MG, et al. A clinical practice Guideline for the management of patients with degenerative cervical myelopathy: recommendations for patients with mild, moderate, and severe disease and nonmyelopathic patients with evidence of Cord Compression. Global Spine J. 2017;7(3 Suppl):s70–83.
25. Hasegawa K, et al. Effects of surgical treatment for cervical spondylotic myelopathy in patients > or = 70 years of age: a retrospective comparative study. J Spinal Disord Tech. 2002;15(6):458–60.
26. Kim HJ, et al. Diabetes and smoking as prognostic factors after cervical laminoplasty. J Bone Joint Surg Br. 2008;90(11):1468–72.
27. Nagata K, et al. Cervical myelopathy in elderly patients: clinical results and MRI findings before and after decompression surgery. Spinal Cord. 1996;34(4):220–6.
28. Matsuda Y, et al. Outcomes of surgical treatment for cervical myelopathy in patients more than 75 years of age. Spine (Phila Pa 1976). 1999;24(6):529–34.
29. Whetstone WD, et al. Blood-spinal cord barrier after spinal cord injury: relation to revascularization and wound healing. J Neurosci Res. 2003;74(2):227–39.
30. Figley SA, et al. Characterization of vascular disruption and blood-spinal cord barrier permeability following traumatic spinal cord injury. J Neurotrauma. 2014;31(6):541–52.
31. Karadimas SK, et al. Immunohistochemical profile of NF-κB/p50, NF-κB/p65, MMP-9, MMP-2, and u-PA in experimental cervical spondylotic myelopathy. Spine (Phila Pa 1976). 2013;38(1):4–10.
32. Kalsi-Ryan S, Karadimas SK, Fehlings MG. Cervical spondylotic myelopathy: the clinical phenomenon and the current pathobiology of an increasingly prevalent and devastating disorder. Neuroscientist. 2013;19(4):409–21.
33. Hirai T, et al. The prevalence and phenotype of activated microglia/macrophages within the spinal cord of the hyperostotic mouse (twy/twy) changes in response to chronic progressive spinal cord compression: implications for human cervical compressive myelopathy. PLoS ONE. 2013;8(5):e64528.
34. Yu WR, et al. Human neuropathological and animal model evidence supporting a role for Fas-mediated apoptosis and inflammation in cervical spondylotic myelopathy. Brain. 2011;134(Pt 5):1277–92.
35. Vedantam A, Rajshekhar V. Does the type of T2-weighted hyperintensity influence surgical outcome in patients with cervical spondylotic myelopathy? A review. Eur Spine J. 2013;22(1):96–106.
36. Vedantam A, Rajshekhar V. Change in morphology of intramedullary T2-weighted increased signal intensity after anterior decompressive surgery for cervical spondylotic myelopathy. Spine (Phila Pa 1976). 2014;39(18):1458–62.
37. Chen CJ, et al. Intramedullary high signal intensity on T2-weighted MR images in cervical spondylotic myelopathy: prediction of prognosis with type of intensity. Radiology. 2001;221(3):789–94.

Publisher's note