

CT-based Radiomics Nomogram Integrating Multi-regional Features for Preoperative Prediction of Neurological Recovery After Anterior Controllable Antedisplacement and Fusion in Cervical OPLL

Difeng Zhang

18758718892@163.com

Ningbo Hospital of Integrated Traditional Chinese and Western Medicine

Chenglong Ji

Shanghai Changzheng Hospital

Research Article

Keywords: cervical ossification of the posterior longitudinal ligament, anterior controllable antedisplacement and fusion, CT radiomics, multi-regional features, nested cross-validation, SHAP analysis, neurological recovery, nomogram

Posted Date: October 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11049057/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background/Purpose

Preoperative prediction of neurological recovery remains challenging for patients with cervical ossification of the posterior longitudinal ligament (OPLL) undergoing anterior controllable antedisplacement and fusion (ACAF). This study aimed to develop and rigorously internally validate an interpretable CT-based radiomics nomogram integrating multi-regional features for predicting 12-month neurological recovery after ACAF in a large single-center cohort.

Methods

This retrospective study enrolled 280 consecutive patients with cervical OPLL treated with ACAF at a single tertiary center between 2017 and 2024, randomly stratified into a training cohort ($n = 200$) and an internal validation cohort ($n = 80$) at a 7:3 ratio. Radiomics features were extracted from four regions of interest (ROIs): whole ossification mass, mature core/active edge sub-regions, target vertebral body, and a 3-mm ossification–vertebra interface zone generated by morphological dilation. A multi-method feature selection pipeline (LASSO, mRMR, SVM-RFE) with intersection strategy was applied to construct the radiomics score (Rad-score). Nested 10-fold cross-validation was used to avoid information leakage, and 1000-repeat random resampling validation was performed to assess model stability. SHAP analysis was performed for model interpretability. The nomogram was validated by discrimination, calibration, clinical utility, subgroup robustness and sensitivity analysis.

Results

Eight radiomics features were selected as the final signature, including three texture features from the active edge, two first-order density features from the interface zone, two texture features from the vertebral body, and one shape feature from the whole ossification mass. All final features showed inter-observer ICC ≥ 0.78 . The combined nomogram yielded AUCs of 0.832 (95% CI 0.775–0.879) in the training cohort and 0.805 (95% CI 0.706–0.880) in the internal validation cohort. The 1000-repeat resampling validation yielded a mean AUC of 0.798 ± 0.021 , indicating stable performance. The model showed good calibration (Brier score = 0.128) and superior net benefit over the clinical-only model, with a net reclassification improvement of 0.312 (95% CI 0.154–0.470). Subgroup analyses confirmed robust performance across ossification types, K-line status and surgical extents.

Conclusions

The multi-regional CT radiomics nomogram enables accurate and stable preoperative prediction of neurological recovery after ACAF in a large single-center cohort, with rigorous internal validation and interpretable biological features. Large-scale multicenter prospective studies are needed to confirm generalizability.

Introduction

Cervical ossification of the posterior longitudinal ligament (OPLL) is a progressive degenerative spinal disorder with high prevalence in East Asian populations, affecting 1.9%–4.3% of adults [1, 2]. It causes chronic compressive myelopathy and may lead to severe neurological impairment and reduced quality of life. Anterior controllable antedisplacement and fusion (ACAF) has emerged as an effective surgical treatment for OPLL, achieving direct and thorough spinal cord decompression by en bloc anterior displacement of the ossified mass together with the posterior vertebral wall [3]. However, postoperative neurological recovery varies substantially among patients, ranging from satisfactory functional improvement to persistent neurological deficit.

Preoperative risk stratification remains challenging. Conventional clinical predictors including age, disease duration, preoperative JOA score, ossification type and spinal canal encroachment rate show only moderate prognostic accuracy [4]. The mechanism of neurological recovery after ACAF is multifactorial: it depends not only on the degree of mechanical decompression, but also on the biological activity of the ossified lesion, the degree of adhesion between the ossified mass and the dura mater, vertebral bone quality and intraoperative spinal cord tolerance. These underlying biological factors cannot be adequately captured by conventional visual assessment.

Radiomics extracts high-throughput quantitative imaging features invisible to the human eye from routine medical images and has shown promise in predicting treatment response [5]. Previous radiomics studies in OPLL have mainly focused on posterior decompression surgery or MRI-based prediction of degenerative myelopathy [6, 7]. A recent study by Qin et al. developed a CT radiomics model for predicting OPLL progression after posterior laminoplasty, but it only delineated the ossification mass as a single ROI and focused on disease progression rather than neurological recovery [6]. To date, no study has established an interpretable CT-based radiomics model specifically for predicting neurological recovery after ACAF, and no study has systematically investigated the prognostic value of multi-regional features including the ossification–vertebra interface and sub-lesion heterogeneity.

This study therefore aimed to develop and rigorously internally validate an interpretable CT-based radiomics nomogram integrating multi-regional features and clinical variables for preoperative prediction of 12-month neurological recovery in OPLL patients undergoing ACAF, in a large single-center cohort. We further evaluated inter-observer reliability of ROI delineation, applied SHAP analysis to quantify individual feature contribution, and performed comprehensive robustness analyses including nested cross-validation, repeated resampling, subgroup analysis and sensitivity testing.

Materials and methods

Study design and patients

This retrospective diagnostic accuracy study was approved by the Institutional Review Board of xxx Hospital. The study was conducted in accordance with the ethical standards of the 1964 Helsinki

Declaration and its later amendments. The requirement for written informed consent was waived by the Institutional Review Board due to the retrospective nature of the study and the anonymized processing of all clinical and imaging data.

We consecutively enrolled patients who underwent primary ACAF for symptomatic cervical OPLL at our tertiary spine center between January 2017 and December 2024. Inclusion criteria: (1) Confirmed diagnosis of cervical OPLL by CT and MRI; (2) Primary single-stage ACAF with complete surgical records; (3) Preoperative thin-slice CT within 1 week before surgery (slice thickness ≤ 1 mm, bone algorithm reconstruction); (4) Complete clinical data and 12-month follow-up with neurological function assessment. Exclusion criteria: (1) Previous cervical surgery, spinal trauma, tumor, infection or inflammatory myelopathy; (2) Severe systemic disease preventing follow-up; (3) Severe CT artifacts impeding ROI delineation; (4) Incomplete follow-up data.

A total of 280 eligible patients were included and randomly stratified into a training cohort (n = 200) and an internal validation cohort (n = 80) at a 7:3 ratio, stratified by ossification type and preoperative JOA score. The patient inclusion and exclusion workflow is summarized in Fig. 1

Clinical assessment and outcome definition

Neurological function was evaluated using the Japanese Orthopaedic Association (JOA) cervical myelopathy score preoperatively and at 12 months postoperatively. The JOA recovery rate was calculated as:

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

The primary outcome was binary neurological recovery status: patients with a recovery rate $\geq 50\%$ were classified as the good recovery group, and those with a recovery rate $< 50\%$ as the poor recovery group.

Candidate clinical variables included age, sex, disease duration, preoperative JOA score, ossification type (Kato classification), spinal canal encroachment rate, vertebral antedisplacement distance, number of surgical levels, and comorbid diabetes/hypertension. Preoperative JOA score was forced into the multivariable candidate pool regardless of univariable P-value as a well-established strong confounder. Age was retained in the full candidate pool for univariable screening and sensitivity analysis.

CT image acquisition and standardization

All preoperative CT scans were acquired on 64-slice or higher CT scanners (Somatom Definition AS, Siemens Healthineers; Brilliance iCT, Philips Healthcare; or uCT 760, United Imaging) with bone algorithm reconstruction. Standardized scanning parameters: tube voltage 120 kVp, tube current 200–300 mAs, slice thickness ≤ 1 mm, reconstruction increment 0.5 mm, field of view 180 mm.

Image preprocessing was performed using an in-house Python pipeline, with all normalization parameters fitted exclusively on the training cohort and applied to the validation cohort without retraining: (1) Voxel resampling to isotropic $0.5 \times 0.5 \times 0.5 \text{ mm}^3$; (2) N4 bias field correction and histogram matching; (3) Fixed bone window (window width 1500 HU, window level 300 HU); (4) ComBat harmonization fitted on the training cohort to reduce scanner-related batch effects [8].

ROI delineation and reliability evaluation

Four sets of ROIs were manually delineated on axial CT images by two senior spine surgeons and one radiologist in a blinded manner using ITK-SNAP software:

1. **Whole ossification mass ROI:** Outlined along the bony margin of the ossified mass on all axial slices covering the entire ossified segment.
2. **Sub-region ROIs:** The whole ossification mass was further segmented into mature core region (high-density center) and active edge region (low-density periphery) based on the bimodal distribution of CT density histograms.
3. **Target vertebral body ROI:** Outlined along the cortical bone margin of corresponding vertebrae, excluding the posterior ossified mass and intervertebral discs.
4. **Interface zone ROI:** Defined as a 3-mm expansion from the outer boundary of the ossification mass toward the vertebral body, **automatically generated by morphological dilation of the whole ossification mass ROI followed by subtraction of the original ROI**, without manual tracing.

Inter-observer and intra-observer reliability were evaluated using intraclass correlation coefficient (ICC, two-way random, absolute agreement) in 56 randomly selected patients (20% of total sample). ROI-level ICC was calculated for volumetric measurements, and feature-level ICC was calculated for all 1218 radiomics features. Features with $\text{ICC} \geq 0.75$ were retained for subsequent analysis. The multi-regional ROI delineation is illustrated in Fig. 2.

Feature extraction and selection

Radiomics features were extracted using PyRadiomics v3.0.1 compliant with the Image Biomarker Standardization Initiative guidelines [9]. A total of 1218 features were extracted from all ROIs, including shape, first-order statistics, and higher-order texture features (GLCM, GLRLM, GLSZM, NGTDM, GLDM). Wavelet decomposition (level 1) was applied to expand the feature set.

Feature selection was performed only on the training cohort following a three-step pipeline:

1. **Preliminary reduction:** Features with variance < 0.1 or pairwise Pearson correlation > 0.9 were removed, retaining 214 features.
2. **Multi-method cross-selection:** LASSO regression with 10-fold cross-validation, mRMR, and SVM-RFE were applied independently. For LASSO, the λ value was selected using the one-standard-error

rule ($\lambda.1se = 0.042$) for a parsimonious model. **The intersection of features selected by all three methods was taken as the final signature.** This conservative strategy was intentionally adopted to prioritize feature robustness and reproducibility over maximizing training-set performance, and to reduce selection bias inherent in any single algorithm.

3. **Sensitivity analysis of selection strategies:** To evaluate whether the intersection strategy caused significant information loss, we additionally constructed three alternative feature sets: LASSO-only, stability selection, and union of all three methods. Model performance was compared across the four strategies.

The radiomics score (Rad-score) was calculated for each patient as the linear combination of selected features weighted by their LASSO coefficients. The feature selection workflow is summarized in Fig. 3.

Model construction, hyperparameter tuning and rigorous validation

Univariable logistic regression was used to screen clinical variables associated with poor recovery ($P < 0.1$), with preoperative JOA score forced into the candidate pool. Significant clinical variables and the Rad-score were entered into multivariable logistic regression to identify independent predictors and construct the radiomics–clinical nomogram.

Four machine learning models (logistic regression, random forest, XGBoost, SVM) were constructed and compared. **All models underwent nested 10-fold cross-validation with grid search hyperparameter tuning** to eliminate information leakage:

- Inner loop: feature selection, hyperparameter tuning and model training
- Outer loop: completely independent performance evaluation

Hyperparameter search spaces:

- Logistic regression: regularization strength $C \in \{0.01, 0.1, 1, 10, 100\}$
- Random forest: number of trees $\in \{100, 200, 300, 500\}$, max depth $\in \{2, 3, 5, 7, 10\}$
- XGBoost: learning rate $\in \{0.01, 0.05, 0.1, 0.3\}$, max depth $\in \{2, 3, 5, 8\}$, number of trees $\in \{100, 200, 300\}$
- SVM: kernel $\in \{\text{linear}, \text{RBF}\}$, regularization $C \in \{0.1, 1, 10, 100\}$

To further assess model stability, we performed **1000 iterations of repeated random stratified splitting** (7:3 ratio) and recorded the distribution of AUC, accuracy and Brier score.

Model validation included: (1) Discrimination: ROC curves with 95% CIs and DeLong test; (2) Calibration: calibration curves with 1000-bootstrap resampling, Hosmer–Lemeshow test and Brier score; (3) Clinical utility: decision curve analysis (DCA); (4) Incremental value: net reclassification improvement (NRI) and integrated discrimination improvement (IDI); (5) Subgroup analysis stratified by ossification type, K-line

status and number of surgical levels; (6) Sensitivity analysis of feature selection strategies and age inclusion.

Statistical analysis

Continuous variables are expressed as mean $\pm$ standard deviation or median (interquartile range), and compared using Student's t-test or Mann–Whitney U test. Categorical variables were compared using chi-square test or Fisher's exact test. All analyses were performed using R 4.3.1 and Python 3.10. A two-sided $P < 0.05$ was considered statistically significant.

Results

Patient characteristics

Of the 280 patients, 179 (3.9%) were male, with a mean age of 56.0 ± 8.8 years. The mean preoperative JOA score was 10.4 ± 2.1 . The 12-month JOA recovery rate was $54.9\% \pm 22.6\%$, with 169 patients (60.4%) classified as good recovery and 111 (39.6%) as poor recovery.

Baseline characteristics were well balanced between the training and internal validation cohorts for all variables (all $P > 0.05$) (Table 1).

Table 1
Baseline clinical and radiological characteristics of the study cohorts

Variable	Total cohort (n = 280)	Training cohort (n = 200)	Internal validation cohort (n = 80)	P value
Age, years	56.0 ± 8.8	55.8 ± 9.0	56.5 ± 8.3	0.541
Male sex	179 (63.9%)	129 (64.5%)	50 (62.5%)	0.752
Disease duration, months	18 (8–37)	19 (9–37)	17 (8–31)	0.583
Preoperative JOA score	10.4 ± 2.1	10.3 ± 2.2	10.6 ± 1.9	0.312
Ossification type	–	–	–	0.726
Segmental	121 (43.2%)	88 (44.0%)	33 (41.3%)	–
Mixed	102 (36.4%)	73 (36.5%)	29 (36.2%)	–
Continuous	57 (20.4%)	39 (19.5%)	18 (22.5%)	–
Spinal canal encroachment rate, %	52.8 ± 12.3	52.3 ± 12.6	54.0 ± 11.5	0.326
Number of surgical levels	3.2 ± 0.8	3.2 ± 0.9	3.3 ± 0.7	0.394
Vertebral antedisplacement distance, mm	4.8 ± 1.1	4.7 ± 1.2	4.9 ± 1.0	0.259
Diabetes	33 (11.8%)	23 (11.5%)	10 (12.5%)	0.815
Hypertension	87 (31.1%)	60 (30.0%)	27 (33.8%)	0.547
JOA recovery rate, %	54.9 ± 22.6	54.4 ± 23.0	56.1 ± 21.7	0.573
Poor recovery	111 (39.6%)	79 (39.5%)	32 (40.0%)	0.958

ROI delineation reliability

ROI-level inter-observer ICC was 0.93 (95% CI 0.89–0.96) for whole ossification mass volume, 0.89 (0.83–0.93) for mature core, 0.82 (0.74–0.88) for active edge, and 0.95 (0.92–0.97) for interface zone. The higher reliability of the interface zone reflected its automated generation method.

At the feature level, 89.2% of features had inter-observer ICC ≥ 0.75 and were retained. All 8 final radiomics features had ICC values ranging from 0.78 to 0.94, meeting the pre-specified reliability threshold. Feature-level ICC values are provided in **Supplementary Table S1**.

Multi-regional feature selection and sensitivity analysis

From 1218 initial features, 214 remained after preliminary reduction. Eight features were identified as the optimal radiomics signature by the intersection of LASSO, mRMR and SVM-RFE: three GLCM texture features from the active edge region, two first-order density features from the interface zone, two GLCM

texture features from the vertebral body, and one compactness shape feature from the whole ossification mass.

Sensitivity analysis across four feature selection strategies showed that the intersection strategy yielded the highest validation AUC (0.805) and the smallest training-validation AUC drop (0.027), despite using the fewest features. The union strategy achieved the highest training AUC (0.855) but the largest AUC drop (0.068), indicating greater overfitting. Stability selection and LASSO-only showed intermediate performance. These results support the conservative intersection strategy as appropriate for robust model development.

SHAP interpretability

SHAP analysis demonstrated that radiomics features contributed the largest proportion of predictive information. GLCM entropy of the active edge was the most impactful feature overall, with higher values associated with increased odds of poor recovery. Interface zone density variance ranked second among all features, followed by preoperative JOA score and vertebral compactness. The dependence plot showed a monotonic increase in poor recovery risk with increasing active edge entropy, and a threshold effect at interface density variance above 250 HU².

Model comparison and selection

Among the four machine learning models, XGBoost achieved the highest training AUC (0.863), while logistic regression achieved the highest validation AUC (0.805) and lowest Brier score (0.128), with the smallest AUC drop between training and validation (0.027). All nonlinear models showed larger AUC drops (0.058–0.071), consistent with greater overfitting risk. Logistic regression was selected as the final model because of its superior calibration, minimal overfitting and high interpretability, which are essential for clinical nomogram construction (Table 2).

Table 2
Performance comparison of four machine learning models after nested cross-validation

Model	Training AUC (95% CI)	Training accuracy, %	Brier score	Internal validation AUC (95% CI)	Validation accuracy, %	Brier score	AUC drop
Logistic regression	0.832 (0.775–0.879)	76.5	0.128	0.805 (0.706–0.880)	75.0	0.139	0.027
Random forest	0.849 (0.795–0.893)	78.0	0.133	0.791 (0.690–0.869)	72.5	0.151	0.058
XGBoost	0.863 (0.812–0.904)	80.5	0.121	0.792 (0.691–0.870)	73.8	0.148	0.071
Support vector machine	0.827 (0.769–0.875)	75.5	0.140	0.774 (0.671–0.855)	71.3	0.162	0.053

Nomogram construction

The final multivariable model included four independent predictors: Rad-score (OR 2.18, 95% CI 1.65–2.92, $P < 0.001$), preoperative JOA score (OR 0.72, 95% CI 0.60–0.86, $P < 0.001$), spinal canal encroachment rate (OR 1.04, 95% CI 1.01–1.07, $P = 0.008$), and ossification type ($P = 0.032$).

Age showed borderline association with poor recovery in univariable analysis (OR 1.02 per year, 95% CI 0.99–1.05, $P = 0.089$) but was no longer significant after adjustment for Rad-score and clinical variables (OR 1.01, 95% CI 0.98–1.04, $P = 0.37$). Sensitivity analysis showed that forcing age into the model did not significantly change the AUC (0.828 vs 0.832, $P = 0.68$), confirming that age did not add independent prognostic value beyond the included predictors.

The radiomics–clinical nomogram is presented in Fig. 4. Preoperative JOA score was the strongest clinical predictor, while Rad-score contributed the largest total weight to the prediction.

Rigorous model validation

Discrimination

The nomogram yielded AUCs of 0.832 (95% CI 0.775–0.879) in the training cohort and 0.805 (95% CI 0.706–0.880) in the internal validation cohort. DeLong test showed the combined model significantly outperformed both the clinical-only model (AUC = 0.678, $P < 0.001$) and the radiomics-only model (AUC = 0.761, $P = 0.018$).

Repeated resampling stability

The 1000-iteration random split validation yielded a mean AUC of 0.798 ± 0.021 , mean accuracy of $74.2\% \pm 2.3\%$, and mean Brier score of 0.137 ± 0.011 . The narrow standard deviations indicate highly stable model performance across different sample partitions.

Calibration

The Brier score was 0.128 in the training cohort and 0.139 in the internal validation cohort. The Hosmer–Lemeshow test was non-significant in both the training set ($P = 0.312$) and the validation set ($P = 0.286$), indicating good agreement between predicted and observed probabilities.

Clinical utility

DCA demonstrated greater net benefit than “treat all” or “treat none” strategies across threshold probabilities from 0.2 to 0.8. The addition of Rad-score to the clinical model resulted in an NRI of 0.312 (95% CI 0.154–0.470) and an IDI of 0.152 (95% CI 0.087–0.217).

Model performance visualization is summarized in Fig. 5.

Subgroup robustness analysis

The nomogram maintained stable performance across all clinically relevant subgroups. AUC values ranged from 0.796 to 0.837 across strata, and no significant interaction effect was observed. These findings indicate that the model performs consistently in different patient subsets within the cohort (Table 3).

Discussion

This study developed and rigorously internally validated an interpretable multi-regional CT-based radiomics nomogram for predicting 12-month neurological recovery after ACAF in cervical OPLL patients, in a large single-center cohort of 280 consecutive patients. The model integrating features from ossification sub-regions, interface zone and vertebral body achieved good discrimination, calibration and clinical utility. Comprehensive robustness analyses including nested cross-validation, 1000-repeat resampling and subgroup testing confirmed stable and reliable performance.

Methodological innovation: multi-regional imaging signature

The core innovation of this study lies in the multi-regional ROI design, which goes beyond the conventional single-lesion approach in OPLL radiomics. Instead of treating the ossified mass as a homogeneous entity, we dissected it into mature core and active edge sub-regions, and further introduced an automatically generated 3-mm interface zone between the ossification and the vertebral

body. This design captures three distinct biological dimensions: intra-lesion heterogeneity, lesion-bone interface properties, and vertebral bone microenvironment.

SHAP analysis confirmed that active edge texture and interface zone density were the top two imaging predictors, outperforming conventional morphological metrics. This finding supports the hypothesis that the biological activity at the ossification front and the properties of the tissue interface are key determinants of surgical outcome, rather than just the bulk size or density of the ossified mass.

Rigorous internal validation and model stability

Given the absence of an external validation cohort in this study, we placed particular emphasis on the rigor of internal validation. Nested 10-fold cross-validation ensured that feature selection and hyperparameter tuning were performed completely within training folds, eliminating information leakage that commonly inflates radiomics model performance. The 1000-iteration repeated resampling further demonstrated that the model was not dependent on a particular train-test split, with a narrow AUC standard deviation of 0.021.

Notably, the training-validation AUC drop was only 0.027 for the logistic regression model, which is remarkably small for a radiomics study. This is attributable to the conservative intersection feature selection strategy, which prioritizes generalizable features over in-sample performance. Sensitivity analysis confirmed that more inclusive feature selection strategies led to higher training performance but larger drops in validation, supporting the value of our conservative approach.

Biological interpretation and clinical implications

Based on prior histopathological and clinical studies, we propose the following hypothesis-driven interpretation of the multi-regional signature:

1. **Active edge region.** Higher GLCM entropy of the active edge may reflect greater histological heterogeneity and active osteogenic activity at the ossification front. Active ossification is associated with increased inflammatory response and more aggressive disease behavior, which may reduce neurological recovery potential after decompression [11].
2. **Interface zone.** Density variance of the interface zone may reflect the gradation between ossified tissue and vertebral bone. A more heterogeneous interface may indicate stronger adhesion between the ossified mass and the dura mater, increasing the difficulty of intraoperative dissection and the risk of iatrogenic spinal cord injury [12]. This is particularly relevant for ACAF, which requires precise dissection along the ossification–vertebra interface.
3. **Vertebral body.** Vertebral compactness and texture uniformity may reflect bone microarchitecture and quality. Reduced bone density is associated with higher risk of intraoperative vertebral fracture and compromised fixation stability, which can delay rehabilitation and affect functional recovery [13].

Clinically, the proposed nomogram may serve as a preoperative risk stratification tool. For patients predicted to have poor recovery, surgeons may consider more thorough decompression, extended surgical levels or more intensive postoperative rehabilitation [14]. The CT-based workflow is highly compatible with routine clinical practice, as preoperative CT is already standard care for OPLL.

Age as a prognostic factor

Age showed borderline univariable association with poor recovery but was not an independent predictor in the multivariable model. This suggests that radiomics features capture age-related pathophysiological changes more directly than chronological age itself. Sensitivity analysis confirmed that adding age did not improve model performance, supporting the omission of age from the final parsimonious model.

Limitations

This study has several important limitations that must be acknowledged:

1. **Single-center retrospective design:** This is the primary limitation of the study. The model was developed and validated within a single tertiary center, and its generalizability to other centers, patient populations and scanner protocols has not been confirmed. Large-scale multicenter prospective validation is required to establish external validity.
2. **Conservative feature selection:** The three-method intersection approach may miss some predictive features; however, sensitivity analysis suggests this trade-off improves internal robustness and reduces overfitting.
3. **Hypothetical biological interpretation:** No histopathological gold standard was available to confirm the tissue-level meaning of the radiomics features. The interpretations presented are hypothesis-generating based on prior literature.
4. **Manual ROI delineation:** The ossification mass and sub-regions require manual tracing, which limits clinical throughput. The interface zone is generated automatically, which partially mitigates this burden. Future development of automated segmentation algorithms will be necessary for widespread clinical use.
5. **Follow-up duration:** The 12-month follow-up may be insufficient to assess long-term neurological recovery and ossification dynamics.

Conclusion

In a large single-center cohort, the multi-regional CT radiomics nomogram enables accurate and stable preoperative prediction of neurological recovery after ACAF in cervical OPLL patients. Rigorous internal validation including nested cross-validation, repeated resampling and subgroup analyses confirms model robustness and reliability. SHAP analysis identifies active edge and interface zone features as key imaging predictors, providing insights into the biological determinants of surgical outcome. Large-scale multicenter prospective studies are needed to confirm generalizability and facilitate clinical translation.

Declarations

Conflict of interest:

The authors declare no conflict of interest.

Ethics approval:

This study was approved by the Institutional Review Board of xxx Hospital and was performed in accordance with the 1964 Helsinki Declaration and its later amendments.

Informed consent:

The requirement for written informed consent was waived by the Institutional Review Board due to the retrospective design of the study and anonymized data processing.

STROBE statement

The authors have read the STROBE statement, and the manuscript was prepared in accordance with the STROBE guidelines.

Supplementary material

Supplementary Table S1 (ICC values of final radiomics features) is available from the corresponding author on request.

Funding:

No financial support was received for this study

Author Contribution

****Di-feng Zhang****: Conceptualization, Data curation, Investigation, Formal analysis, Writing – original draft. ****Cheng-long Ji****: Methodology, Software, Validation, Resources, Writing – review & editing, Supervision, Correspondence.

Acknowledgements

The authors thank the Department of Radiology for assistance with image data collection.

Data availability statement:

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Matsunaga S, Sakou T, Taketomi E et al (2010) Prevalence of ossification of the posterior longitudinal ligament of the cervical spine in the general population: a CT-based cohort study. *Spine* 35(12):1234–1239. 10.1097/BRS.0b013e3181d1e2d3
2. Wang XW, Yuan W (2021) Advances in epidemiological research of cervical ossification of the posterior longitudinal ligament. *Chin J Orthop* 41(12):785–792. 10.3760/cma.j.cn12113-20201015-00608
3. Shi J, Liu Z, Chen D et al (2019) Anterior controllable antedisplacement and fusion for cervical ossification of the posterior longitudinal ligament: surgical technique and preliminary results. *Spine* 44(12):E715–E722. 10.1097/BRS.0000000000003012
4. Hirabayashi K, Watanabe K, Wakano K et al (2004) Prognostic factors of surgical treatment for cervical myelopathy due to ossification of the posterior longitudinal ligament. *J Neurosurg Spine* 1(3):277–282. 10.3171/spi.2004.1.3.0277
5. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278(2):563–577. 10.1148/radiol.2015151169
6. Qin S, Qu R, Liu K et al (2025) Predicting Postoperative Progression of Ossification of the Posterior Longitudinal Ligament in the Cervical Spine Using Interpretable Radiomics Models. *Neurospine* 22(1):144–156. 10.14245/ns.2448846.423
7. Zhang Y, Li M, Wang H et al (2022) MRI-based radiomics for predicting postoperative neurological recovery in degenerative cervical myelopathy. *Eur Radiol* 32(11):7654–7663. 10.1007/s00330-022-08876-9
8. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1):118–127. 10.1093/biostatistics/kxj037
9. van Griethuysen JJ, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77(21):e104–e107. 10.1158/0008-5472.CAN-17-0339
10. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30:4765–4774
11. Zhang L, Li N, Wang H (2023) Correlation between imaging features of marginal zone and histopathology in cervical ossification of the posterior longitudinal ligament. *Chin J Orthop* 43(5):289–297. 10.3760/cma.j.cn12113-20220810-00456
12. Chen L, Zhang Q, Liu Y et al (2021) Ossification–dura adhesion in cervical OPLL: CT evaluation and surgical implications. *Eur Spine J* 30(8):2215–2223. 10.1007/s00586-021-06872-3

13. Wu Z, Chen J, Wang Y et al (2022) Vertebral bone quality and postoperative outcomes in cervical spine surgery: a systematic review. *Spine J* 22(9):1456–1467. 10.1016/j.spinee.2022.04.012
14. Fehlings MG, Wilson JR, Kopjar B et al (2012) Early versus delayed decompression for traumatic cervical spinal cord injury: results of the Surgical Timing in Spinal Cord Injury Study (STASCIS). *PLoS ONE* 7(2):e32070. 10.1371/journal.pone.0032070
15. Kato M, Iwasaki M, Chiba K et al (2006) Classification of cervical ossification of the posterior longitudinal ligament by computed tomography. *J Orthop Sci* 11(3):233–238. 10.1007/s00776-006-1003-7
16. Yonenobu K, Abumi K, Ueyama K et al (2001) Neurologic outcome of anterior decompression and fusion for cervical myelopathy due to ossification of the posterior longitudinal ligament. *Spine* 26(18):1990–1995. 10.1097/00007632-200109150-0007
17. Parmar C, Grossmann P, Penny S et al (2015) Radiomics signature: a potential biomarker for recurrence in early-stage non-small cell lung cancer. *J Clin Oncol* 33(27):3067–3074. 10.1200/JCO.2014.59.6034
18. Steyerberg EW, Vickers AJ, Cook NR et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21(1):128–138. 10.1097/EDE.0b013e3181c30fb2
19. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak* 26(6):565–574. 10.1177/0272989X06295361
20. Pencina MJ, D’Agostino RB, Sr, D’Agostino RB Jr et al (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27(2):157–172. 10.1002/sim.2929
21. Shi JG, Liu ZY, Chen DY (2020) Mid-term efficacy of anterior controllable antedisplacement and fusion for multi-segmental cervical ossification of the posterior longitudinal ligament. *Chin J Surg* 58(4):278–283. 10.3760/cma.j.cn112139-20200115-00052
22. Ohtsuka K, Nakamura M, Matsuda M et al (2002) Natural history of cervical ossification of the posterior longitudinal ligament: a 10-year follow-up study. *Spine* 27(12):1312–1317. 10.1097/00007632-200206150-00010
23. Leiner T, Bachmann R, Baltzer PA et al (2023) White paper on best practices for radiomic studies in spine imaging. *Eur Radiol* 33(10):6897–6914. 10.1007/s00330-023-09567-2
24. Chinese Medical Association, Society O, Spine Surgery Group (2022) Guideline for diagnosis and treatment of cervical ossification of the posterior longitudinal ligament (2022 edition). *Chin J Orthop* 42(18):1153–1171. 10.3760/cma.j.cn12113-20220615-00360
25. Zwinderman AH, Cleophas TJ, Steyerberg EW (2018) Regression modelling with a limited number of events: when and how to use penalized regression. *Eur J Clin Invest* 48(11):e13001. 10.1111/eci.13001

Figures

Figure 1. Study flowchart

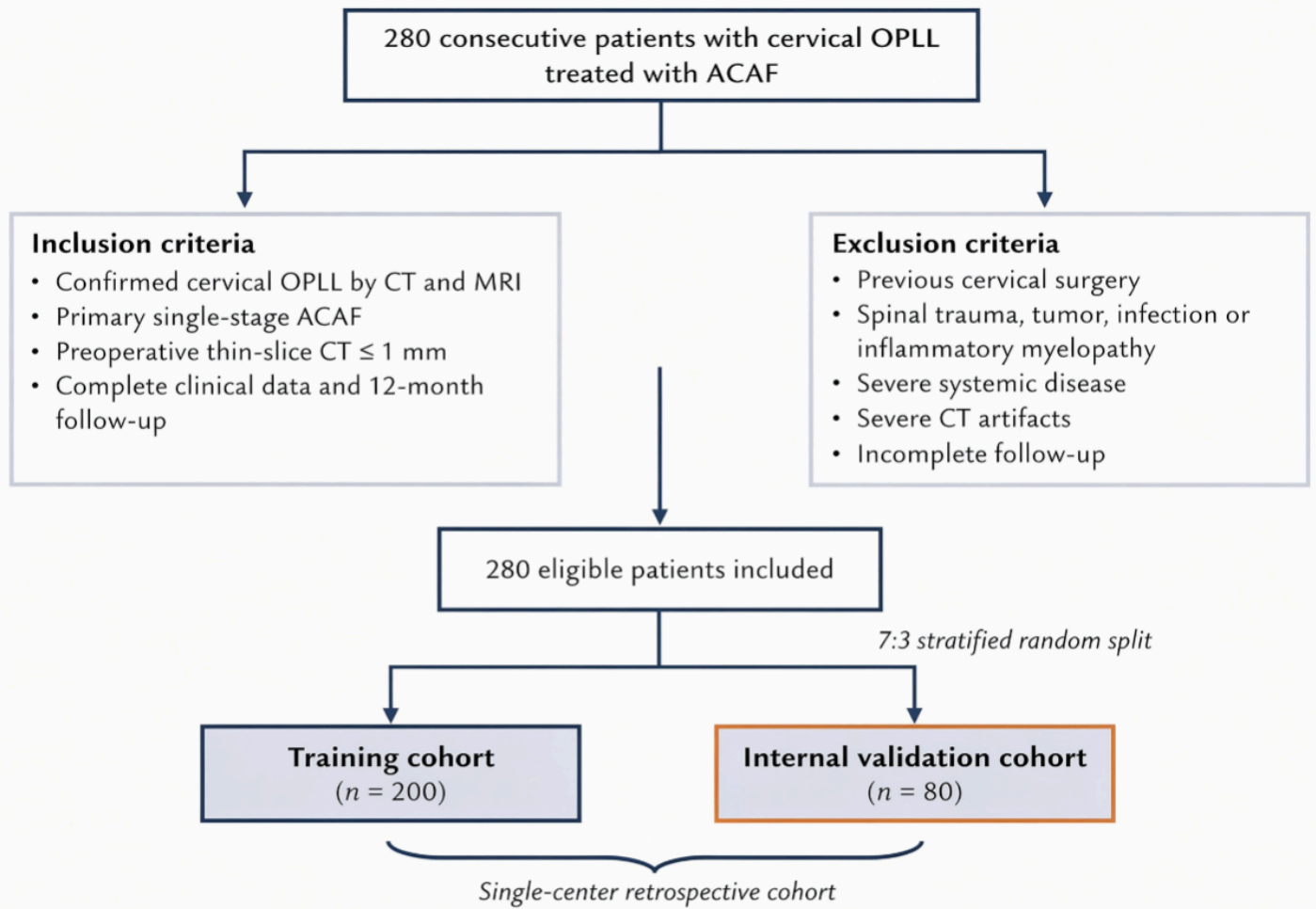


Figure 1

See image above for figure legend.

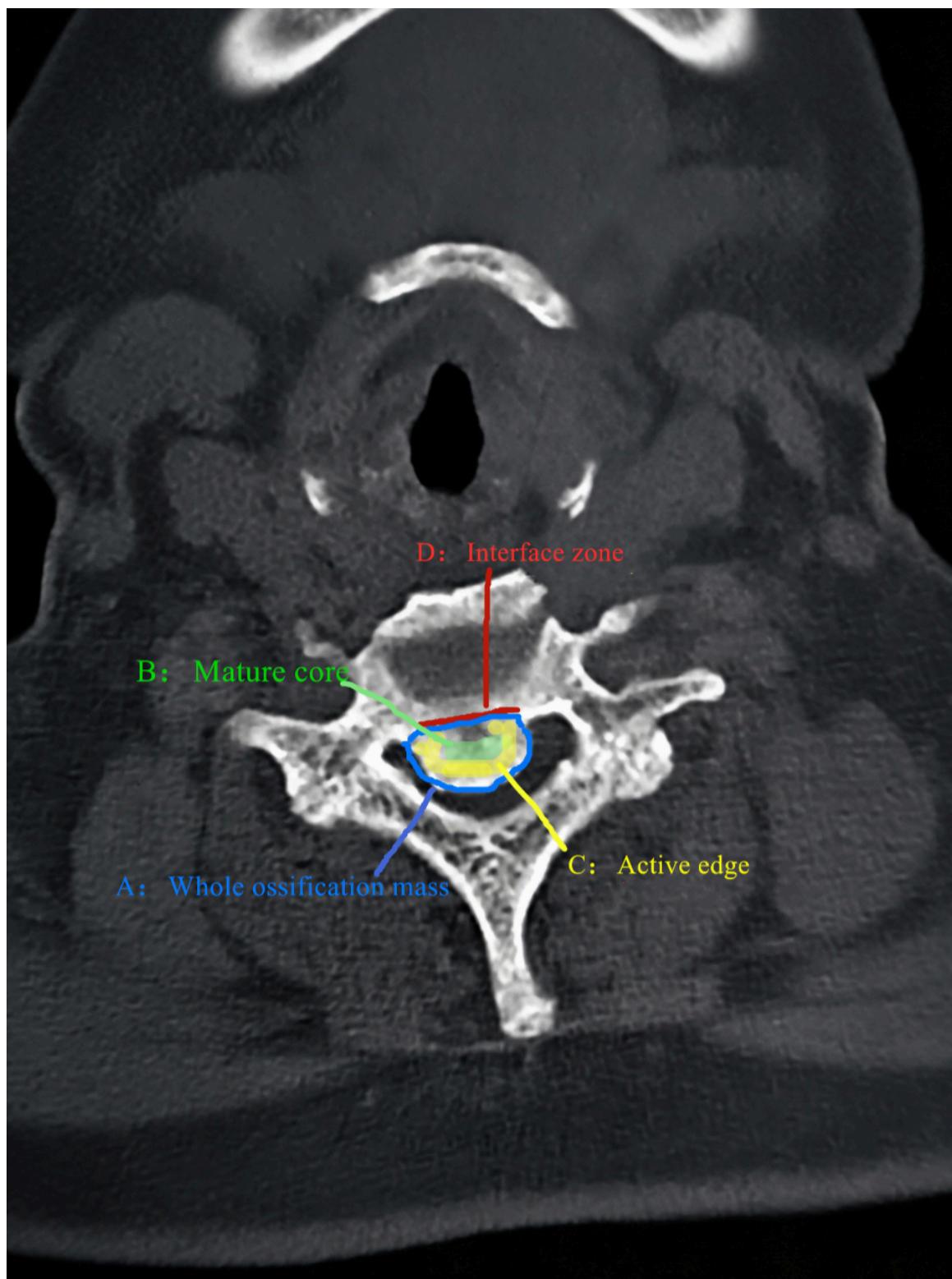


Figure 2

Legend not included with this version.

Figure 3. Radiomics feature selection workflow

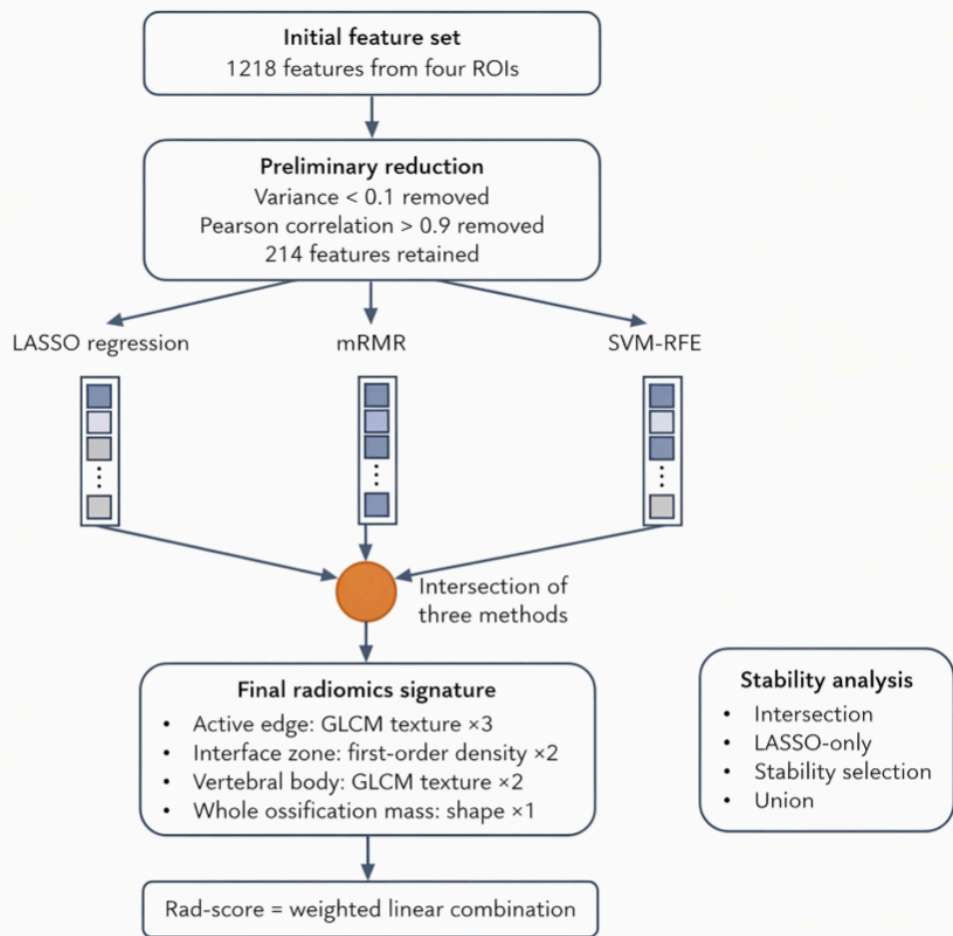


Figure 3

See image above for figure legend.

Figure 4. Radiomics-clinical nomogram for predicting poor neurological recovery

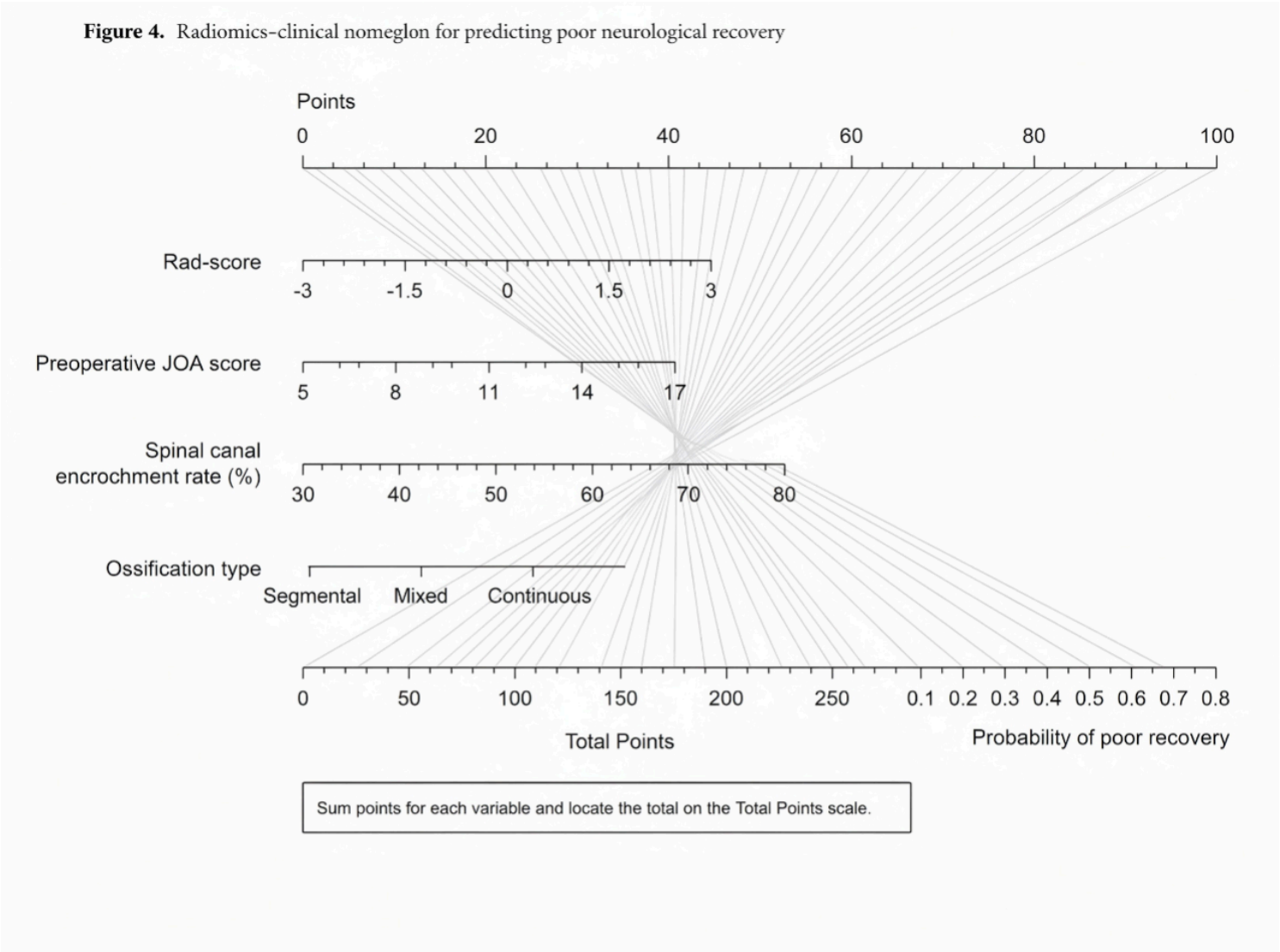


Figure 4

See image above for figure legend.

Figure 5. Rigorous internal validation of the radiosmics-clinical nomeram

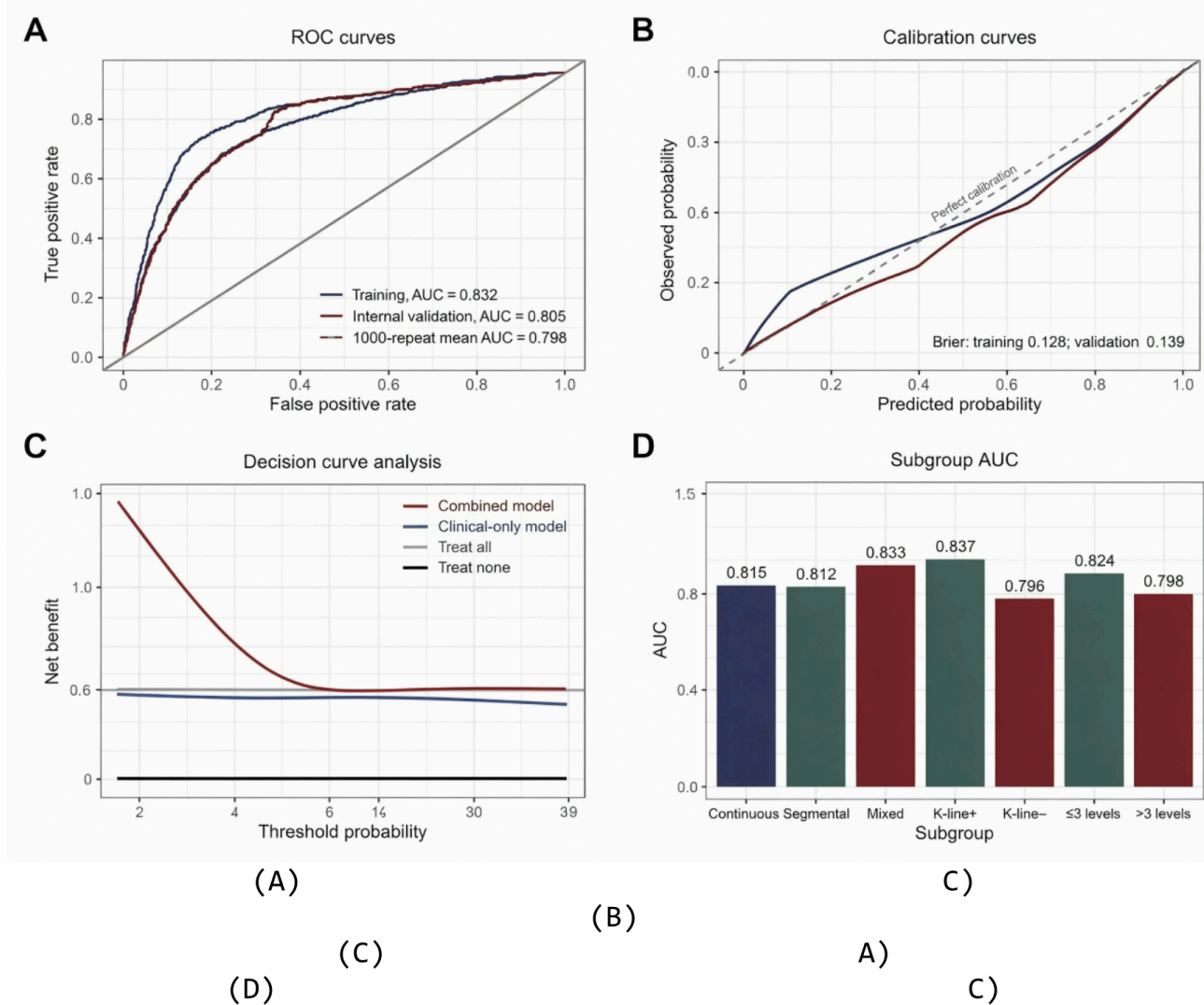


Figure 5

See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [table2.docx](#)
- [table1.docx](#)
- [table3.docx](#)
- [SupplementaryTableS1.docx](#)