Title: Machine Learning-Based Prediction for Axial Pain following Expansive Unilateral Open-Door Laminoplasty: A Retrospective Cohort Study

Authors: Kelun Huang, MD[1, *], Sheng Li, MD[2, *], Yile Dai, MD[3, *], Dingjun Xu, MD[1], Guoliang Chen, MD[1], Yaozhi He, MD[1], Ziwei Fan, MD[1, #], Honglin Teng, MD[1, #]

1. Department of Orthopedics (Spine Surgery), the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325035, Zhejiang, China.
2. Department of Orthopedics, People's Hospital of Ruian City, Wenzhou 325200, Zhejiang, China.
3. Department of Ultrasonic Diagnosis, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325035, Zhejiang, China.

*These authors contributed equally: Kelun Huang, Sheng Li, and Yile Dai.

#Ziwei Fan and Honglin Teng are the co-corresponding authors for this manuscript.

#Corresponding author: Ziwei Fan, Department of Orthopedics (Spine Surgery), the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325035, Zhejiang, China. Email: fzw_formal@163.com

Honglin Teng, Department of Orthopedics (Spine Surgery), the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325035, Zhejiang, China. Email: tenghonglin@wzhospital.cn

Supplementary Figures

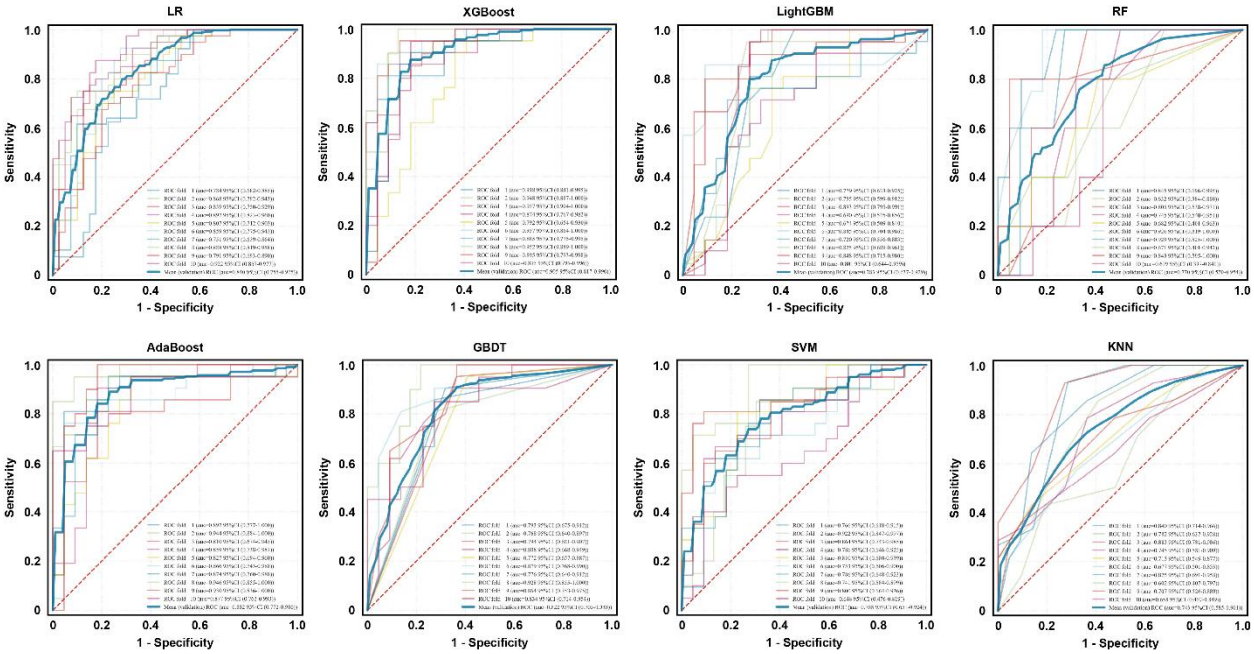

Figure S1. The ROC curves (10-fold cross-validation) for baseline ML model. Abbreviations: ML, machine learning; ROC, receiver operating characteristics; AUC, the area under curve; CI, confidence interval; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

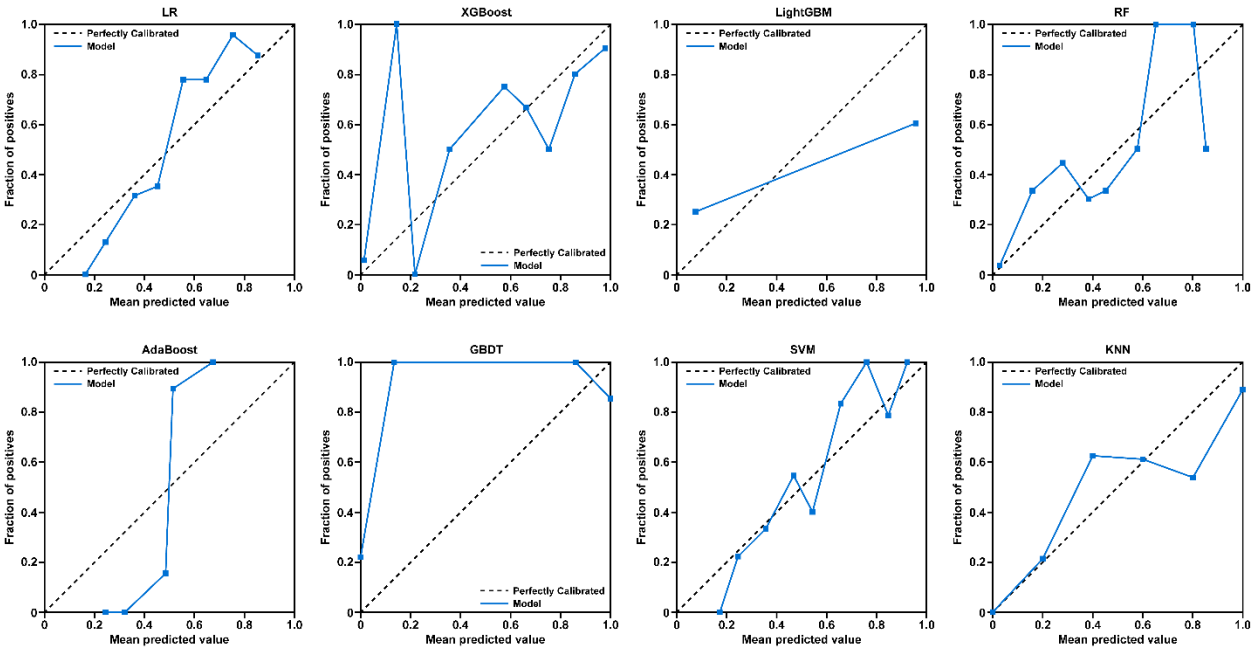

Figure S2. The calibration curves of ML models for baseline prediction. The ideal calibration was represented by a black dotted line, which indicated the perfect matching between the model-predicted and the actual observed probabilities. And the calibration curves of each ML model were shown as a blue broken line. Abbreviations: ML, machine learning; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

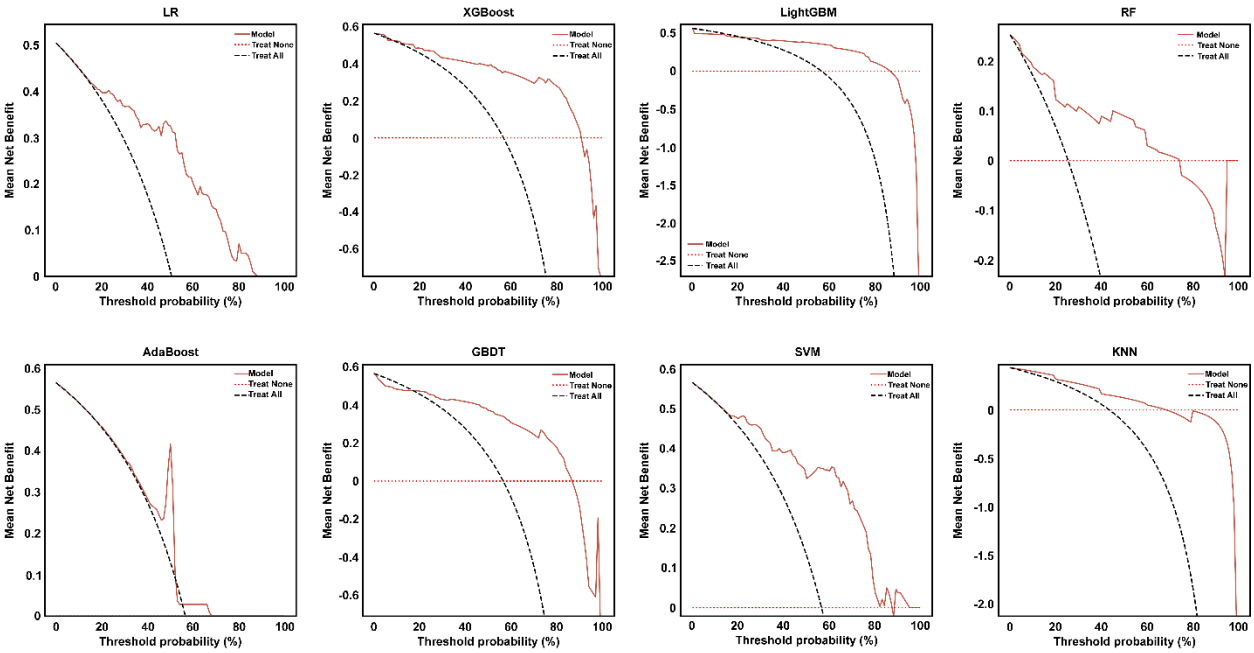

Figure S3. The decision curves of ML models for baseline prediction. The treated-for-none scheme is indicated by the red dotted line, while the treated-for-all scheme is indicated by the black dotted line. Abbreviations: ML, machine learning; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

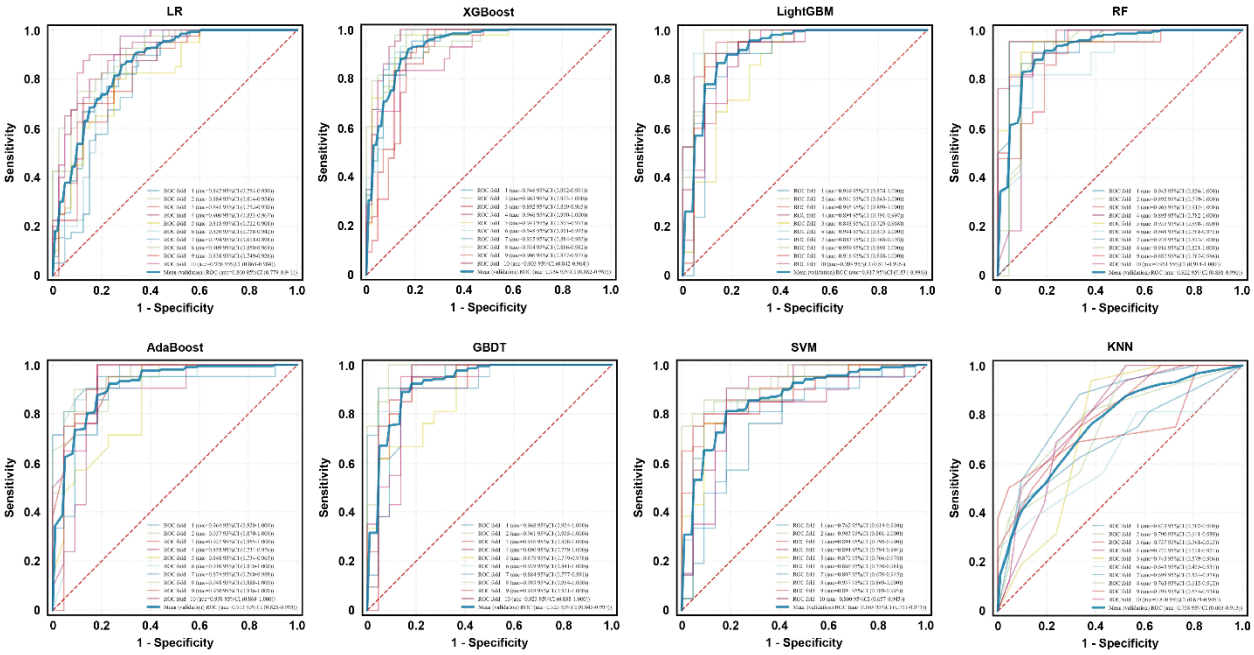

Figure S4. The ROC curves (10-fold cross-validation) for hyperparameter-optimized ML models. Abbreviations: ML, machine learning; ROC, receiver operating characteristics; AUC, the area under curve; CI, confidence interval; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

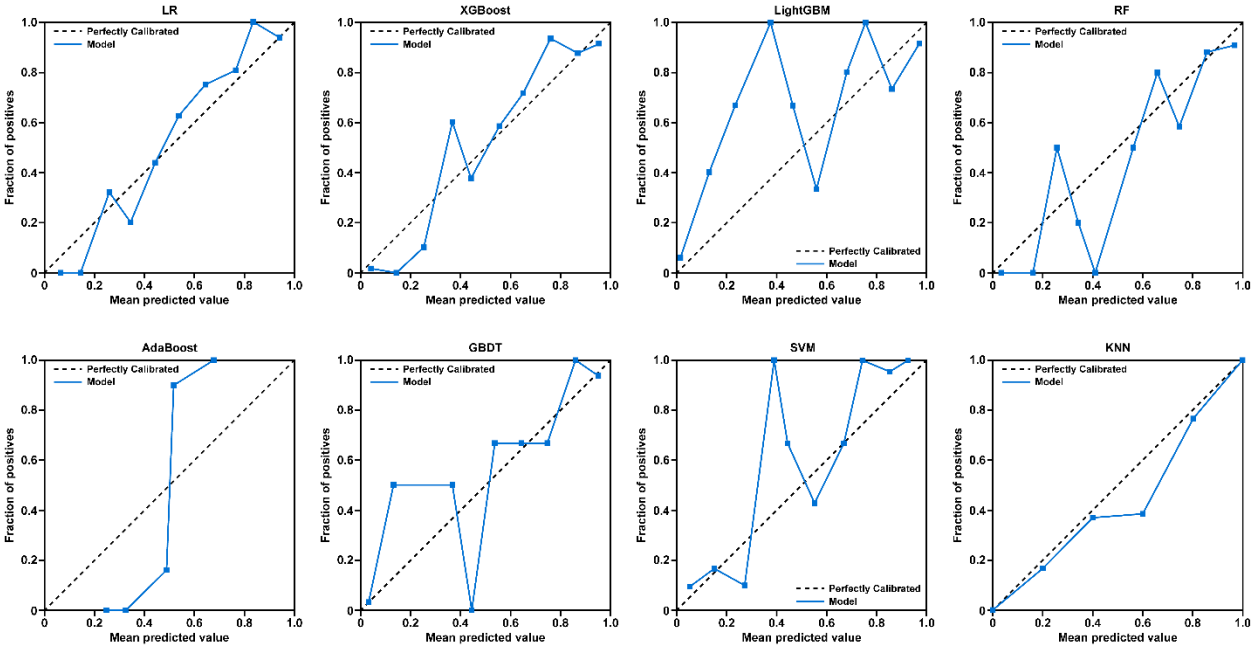

Figure S5. The calibration curves of ML models with the optimized hyperparameters. The ideal calibration was represented by a black dotted line, which indicated the perfect matching between the model-predicted and the actual observed probabilities. And the calibration curves of each ML model were shown as a blue broken line. Abbreviations: ML, machine learning; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

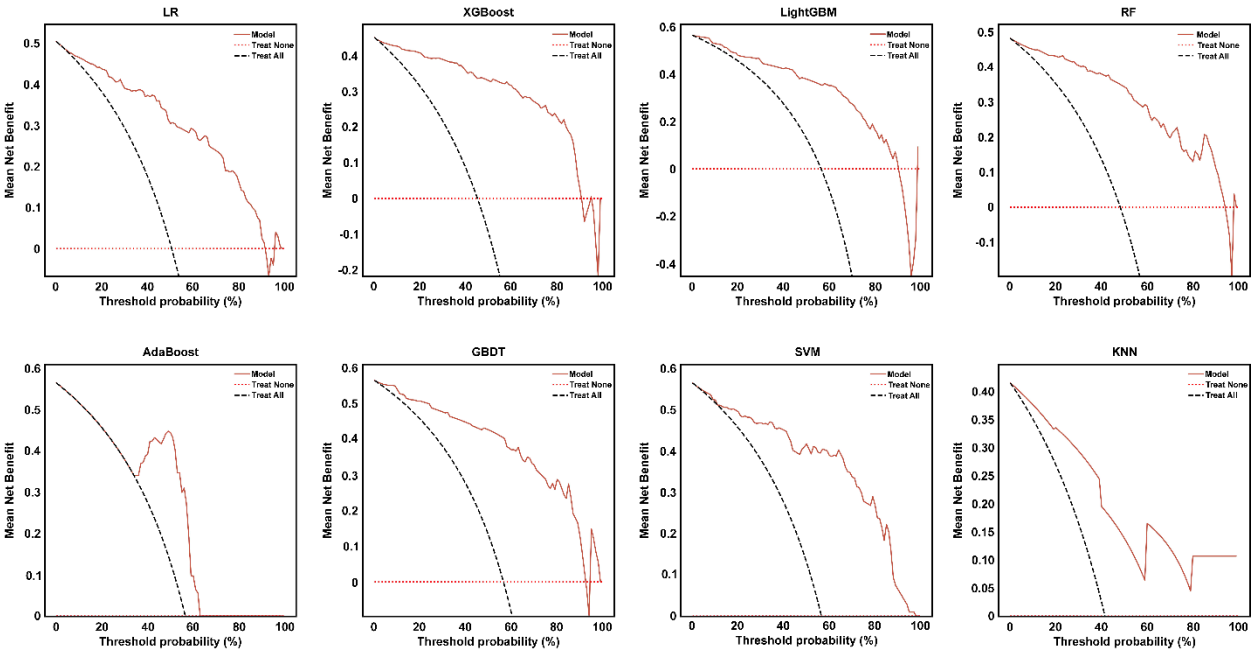

Figure S6. The decision curves of ML models with the optimized hyperparameters. The treated-for-none scheme is indicated by the red dotted line, while the treated-for-all scheme is indicated by the black dotted line. Abbreviations: ML, machine learning; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

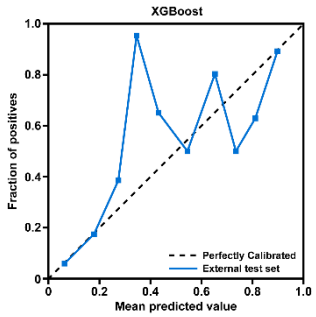

Figure S7. The calibration curves of XGBoost model with the optimized hyperparameters in the temporal validation set. The ideal calibration was represented by a black dotted line, which indicated the perfect matching between the model-predicted and the actual observed probabilities. And the calibration curve of the XGBoost model was shown as a blue broken line. Abbreviations: XGBoost, Extreme Gradient Boost.

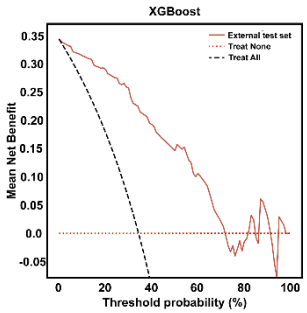

Figure S8. The decision curves of XGBoost model with the optimized hyperparameters in the temporal validation set. The treated-for-none scheme is indicated by the red dotted line, while the treated-for-all scheme is indicated by the black dotted line. Abbreviations: XGBoost, Extreme Gradient Boost.

Supplemental Tables

Table S1. Analyses of intra-rater and inter-rater reliabilities in radiological measurements and clinical-characteristic assessments.

Variable	Intra-rater reliability		Inter-rater reliability	
	PCC or W	*P*	PCC or W	*P*
Cervical ROM	0.895	< 0.001[a]	0.833	< 0.001[a]
Cervical eROM	0.867	< 0.001[a]	0.835	< 0.001[a]
Cervical fROM	0.920	< 0.001[a]	0.913	< 0.001[a]
TIA	0.935	< 0.001[a]	0.908	< 0.001[a]
NT	0.952	< 0.001[a]	0.901	< 0.001[a]
T1S	0.930	< 0.001[a]	0.924	< 0.001[a]
C2-7 SVA	0.913	< 0.001[a]	0.905	< 0.001[a]
SCA	0.904	< 0.001[a]	0.854	< 0.001[a]
CL	0.977	< 0.001[a]	0.969	< 0.001[a]
Cervical instability	0.981	< 0.001[b]	0.970	< 0.001[b]
Cervical kyphosis	0.972	< 0.001[b]	0.967	< 0.001[b]
Occupying rate of cervical spinal canal	0.918	< 0.001[a]	0.888	< 0.001[a]
ONL	0.978	< 0.001[b]	0.964	< 0.001[b]
Quantity-based surgical segment classification	1.00	< 0.001[b]	1.00	< 0.001[b]
C6 laminoplasty	1.00	< 0.001[b]	1.00	< 0.001[b]
C7 laminoplasty	1.00	< 0.001[b]	1.00	< 0.001[b]
Open-door side	1.00	< 0.001[b]	1.00	< 0.001[b]
Angle of lamina open-door	0.954	< 0.001[a]	0.924	< 0.001[a]
JOA	0.951	< 0.001[a]	0.946	< 0.001[a]
Axial pain	0.992	< 0.001[b]	0.989	< 0.001[b]

PCC is used in continuous variables and W is used in categorical variables. The statistical methods indicated by the superscripted letters were as follows: a. Pearson's correlation analysis, b. Kendall's W test. Abbreviations: PCC,

Pearson correlation coefficient; ROM, range of motion; TIA, thoracic inlet angle; NT, neck tilt; T1S, T1 slope; cSVA, cervical sagittal vertical axis; SCA, spine cranial angle; CL, cervical lordosis; ONL, ossification of the nuchal ligament; JOA, Japanese Orthopaedic Association.

Table S2. Lasso regression coefficients of each variable.

Variable	The minimum mean square error	The minimum distance standard error
(Intercept)	-1.147	-0.512
Age	-0.002	0.0
Sex	-0.190	0.0
Weight	0.0	0.0
Height	0.0	0.0
BMI	0.0	0.0
Smoking	0.0	0.0
Drinking	0.0	0.0
Hypertension	0.168	0.0
Diabetes	0.0	0.0
Cervical ROM	0.0	0.0
Cervical eROM	0.0	0.0
Cervical fROM	0.0	0.0
TIA	0.0	0.0
NT	0.0	0.0
T1S	0.0	0.0
C2-7 SVA	0.0	0.0
SCA	0.0	0.0
CL	-0.016	-0.008
Cervical instability	0.0	0.0
Cervical kyphosis	0.504	0.241
Occupying rate of cervical spinal canal	1.588	0.114
ONL	0.0	0.0
Quantity-based surgical segment classification	-1.082	-0.808

C6 laminoplasty	-0.147	0.0
C7 laminoplasty	1.307	0.870
Open-door side	0.0	0.0
Angle of lamina open-door	-0.052	-0.023
Operative time	0.0	0.0
Intraoperative blood loss	0.0	0.0
JOA score	0.013	0.0
VAS score	0.0	0.0

Based on the λ.1-s.e. (standard error) criterion (λ = 0.055), six variables (C7 laminoplasty, quantity-based surgical segment classification, cervical kyphosis, occupying rate of cervical spinal canal, angle of lamina open-door, and CL) were selected as predictors using Lasso regression to subsequent ML model training and testing. Abbreviations: BMI, body mass index; ROM, range of motion; TIA, thoracic inlet angle; NT, neck tilt; T1S, T1 slope; cSVA, cervical sagittal vertical axis; SCA, spine cranial angle; CL, cervical lordosis; ONL, ossification of the nuchal ligament; JOA, Japanese Orthopaedic Association; VAS, visual analogue scale.

Table S3. The hyperparameters for each ML model respectively.

Model	Hyperparameters	Default	Optimized
LR	C	1.0	0.0293
	Max_iter	100	700
	Penalty	L2	L2
	tol	0.0001	0.0001
XGBoost	Objective	Binary:logistic	Binary:logistic
	Learning_rate	None	0.0642
	Max_depth	None	19
	Min_child_weight	None	2
	Reg_lambda	None	0.5
LightGBM	Boosting_type	gbdt	dart
	Learning_rate	0.1	2
	Max_depth	-1.0	20
	N_estimators	100	5
	Num_leaves	31	5
RF	Criterion	gini	gini
	Max_depth	None	None
	Min_impurity_decrease	0.0	0.0
	N_estimators	20	100
AdaBoost	Learning_rate	1.0	0.1
	N_estimators	50	50
GBDT	N_estimators	100	100
	Min_samples_split	2	50
	Min_samples_leaf	50	50
	Max_depth	20	1
	Loss	Exponential	Log_loss

	Learning_rate	2	0.001
SVM	C	1.0	4.1978
	Kernel	rbf	rbf
	tol	0.001	0.0030
KNN	N_neighbors	5	2
	Weights	uniform	distance

Abbreviations: LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

Table S4. Comparison of AUC-ROC, accuracy, sensitivity, specificity, PPV, NPV, and F1-score of the baseline prediction performance for ML models subjected to 10-fold cross-validation.

Model	AUC-ROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1-score (95% CI)
LR	0.840(0.755-0.925)	0.769(0.745-0.793)	0.737(0.704-0.769)	0.800(0.759-0.841)	0.789(0.755-0.823)	0.754(0.733-0.776)	0.760(0.735-0.786)
XGBoost	0.905(0.817-0.990)	0.820(0.782-0.858)	0.774(0.699-0.848)	0.865(0.814-0.915)	0.848(0.803-0.894)	0.811(0.756-0.866)	0.803(0.759-0.847)
LightGBM	0.783(0.637-0.929)	0.769(0.728-0.810)	0.846(0.771-0.922)	0.697(0.631-0.762)	0.728(0.684-0.772)	0.843(0.777-0.908)	0.778(0.733-0.822)
RF	0.770(0.570-0.954)	0.786(0.724-0.848)	0.350(0.203-0.497)	0.885(0.823-0.948)	0.991(0.982-1.000)	0.857(0.825-0.889)	0.996(0.991-1.000)
AdaBoost	0.882(0.772-0.986)	0.816(0.790-0.841)	0.827(0.780-0.875)	0.806(0.752-0.859)	0.806(0.763-0.849)	0.837(0.799-0.875)	0.812(0.787-0.838)
GBDT	0.822(0.705-0.938)	0.773(0.742-0.805)	0.808(0.706-0.909)	0.742(0.673-0.811)	0.756(0.712-0.801)	0.829(0.755-0.903)	0.769(0.726-0.812)
SVM	0.788(0.651-0.924)	0.733(0.685-0.781)	0.749(0.694-0.803)	0.720(0.629-0.810)	0.726(0.670-0.783)	0.753(0.706-0.800)	0.732(0.692-0.771)
KNN	0.743(0.585-0.901)	0.671(0.615-0.728)	0.615(0.523-0.707)	0.706(0.618-0.793)	0.594(0.523-0.665)	0.743(0.689-0.796)	0.594(0.532-0.656)

Abbreviations: AUC-ROC, the area under the receiver operating characteristics curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.

Table S5. Comparison of AUC-ROC, accuracy, sensitivity, specificity, PPV, NPV, and F1-score of hyperparameter-optimized ML models subjected to 10-fold cross-validation.

Model	AUC-ROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1-score (95% CI)
LR	0.860(0.779-0.941)	0.788(0.770-0.805)	0.857(0.828-0.887)	0.718(0.686-0.750)	0.753(0.733-0.773)	0.837(0.809-0.866)	0.801(0.784-0.818)
XGBoost	0.934(0.882-0.985)	0.867(0.856-0.879)	0.869(0.844-0.894)	0.866(0.841-0.891)	0.867(0.846-0.887)	0.872(0.852-0.891)	0.867(0.855-0.879)
LightGBM	0.917(0.831-0.996)	0.830(0.802-0.857)	0.827(0.763-0.891)	0.833(0.783-0.883)	0.828(0.789-0.867)	0.846(0.796-0.895)	0.823(0.789-0.856)
RF	0.922(0.838-0.996)	0.850(0.816-0.884)	0.814(0.751-0.877)	0.886(0.857-0.915)	0.881(0.853-0.909)	0.830(0.784-0.877)	0.843(0.802-0.884)
AdaBoost	0.913(0.828-0.993)	0.832(0.803-0.860)	0.865(0.798-0.932)	0.802(0.744-0.859)	0.810(0.772-0.847)	0.876(0.825-0.927)	0.831(0.797-0.864)
GBDT	0.925(0.845-0.997)	0.848(0.812-0.884)	0.856(0.797-0.915)	0.842(0.779-0.906)	0.843(0.792-0.894)	0.869(0.820-0.918)	0.845(0.808-0.882)
SVM	0.863(0.751-0.973)	0.790(0.754-0.826)	0.831(0.782-0.880)	0.751(0.651-0.851)	0.776(0.716-0.836)	0.833(0.802-0.864)	0.795(0.770-0.820)
KNN	0.758(0.603-0.913)	0.678(0.640-0.716)	0.657(0.585-0.729)	0.693(0.645-0.740)	0.622(0.577-0.666)	0.732(0.687-0.777)	0.635(0.587-0.683)

Abbreviations: AUC-ROC, the area under the receiver operating characteristics curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; LR, Logistic Regression; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; RF, Random Forest; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine; KNN, K Nearest Neighbor.