Original Article

DOI: 10.5582/bst.2023.01327

Automated machine learning-based model for the prediction of pedicle screw loosening after degenerative lumbar fusion surgery

Feng Jiang[1], Xinxin Li[1], Lei Liu[2], Zhiyang Xie[2], Xiaotao Wu[1,2], Yuntao Wang[1,2,*]

[1] Southeast University Medical College, Nanjing, Jiangsu, China;
[2] Department of Spine Surgery, Southeast University ZhongDa Hospital, Nanjing, Jiangsu, China.

SUMMARY The adequacy of screw anchorage is a critical factor in achieving successful spinal fusion. This study aimed to use machine learning algorithms to identify critical variables and predict pedicle screw loosening after degenerative lumbar fusion surgery. A total of 552 patients who underwent primary transpedicular lumbar fixation for lumbar degenerative disease were included. The LASSO method identified key features associated with pedicle screw loosening. Patient clinical characteristics, intraoperative variables, and radiographic parameters were collected and used to construct eight machine learning models, including a training set (80% of participants) and a test set (20% of participants). The XGBoost model exhibited the best performance, with an AUC of 0.884 (95% CI: 0.825–0.944) in the test set, along with the lowest Brier score. Ten crucial variables, including age, disease diagnosis: degenerative scoliosis, number of fused levels, fixation to S1, HU value, preoperative PT, preoperative PI-LL, postoperative LL, postoperative PT, and postoperative PI-LL were selected. In the prospective cohort, the XGBoost model demonstrated substantial performance with an accuracy of 83.32%. This study identified crucial variables associated with pedicle screw loosening after degenerative lumbar fusion surgery and successfully developed a machine learning model to predict pedicle screw loosening. The findings of this study may provide valuable information for clinical decision-making.

Keywords CT Hounsfield units, osteoporosis, lumbar degenerative disease, screw loosening, explainable machine learning

1. Introduction

Pedicle screw fixation is a commonly utilized surgical technique for thoracolumbar disease treatment, which can stabilize the spine before solid fusion and restore spinal balance (*1*). However, screw loosening is one of the common complications associated with this treatment (*2,3*) and may lead to fixation failure, chronic low back pain, non-union and pseudarthrosis, and in severe cases may even require revision surgery (*4-7*), affecting the patient's quality of life. Thus, it is crucial to prevent screw loosening.

Osteoporosis has been identified as the predominant risk factor for screw loosening. In the osteoporotic spine, the bone-screw interface tends to be unstable, resulting in diminished pullout force and cutout force. Clinical studies indicated a pedicle screw loosening rate of less than 15% in non-osteoporotic patients, whereas it could escalate to as much as 60% in osteoporotic patients (*6,8,9*). Dual-energy X-ray absorptiometry (DXA) is currently considered the gold standard for assessing bone mineral density (BMD), with osteoporosis defined by the lowest T-score≤-2.5 (*10,11*). To prevent screw loosening, most spine surgeons have opted for target patients with a T-score of ≤ -2.5 for the application of pedicle screw augmentation techniques (*2, 12-14*). However, the lumbar degenerative changes in patients with lumbar degenerative disease (LDD) can result in an overestimation of T-scores, leading to potential false negative results (*15,16*). As a consequence, DXA outcomes may misguide spine surgeons in their preoperative surgical planning. In recent years, preoperative computed tomography (CT) measurements of vertebral body Hounsfield unit (HU) values have been widely used for the prediction of screw loosening. HU values are measured in the vertebral body, at the midsagittal plane, central transverse plane, and transverse planes close to the superior and inferior endplates separately (*17,18*). In this process, the region of interest (ROI) is expanded as much as possible within the cancellous bone but excluding other bony structures, such as cortical, bony endplates, and osteophytes. The confusion caused by pathological bone formations can be eliminated (*18*), and the specific BMD of cancellous bone can be measured (*17*). Clinical research showed that it was a better predictor of postoperative complications than the T-score (*10,11,15*), and its predictive performance was superior to that of DXA (*10,15,16*). In addition, it was reported that gender, age, number of fused segments, and imaging parameters (fixation to S1, sagittal imbalance) were associated with pedicle screw loosening (*19-21*). Nevertheless, previous studies on predicting screw loosening have predominantly relied on a single statistical approach, potentially limiting their predictive performance (*10,19*). A study by Da *et al*. reported an AUC of only 0.666, below 0.75, for predicting pedicle screw loosening using Hounsfield units in patients with LDD (*10*). Thus the predictive performance is insufficient to meet the clinical needs in the existing models.

Recently, there have been increasing reports that applying machine learning techniques to develop various disease prediction models could improve their predictive performance (*22,23*). Machine learning is an advanced predictive modeling technique founded in computer science, utilizing artificial intelligence to create algorithms trained on data to perform diverse tasks. By employing validation measures, machine learning enhances model robustness, enabling predictions beyond the scope of traditional inferential statistics (*24,25*). Machine learning has been applied to spinal deformity and tumor patients for enhancing the clinical decision-making process (*26,27*). Currently, there are no studies on machine learning models and postoperative pedicle screw loosening in degenerative lumbar fusion surgery. Therefore, in this study, we used multiple artificial intelligence algorithms to construct predictive models for screw loosening and compared these models to finalize the model with the best predictive performance to support clinical decision making.

2. Materials and Methods

2.1. Patient data collection

This study was approved by the Ethics Committee of Southeast University ZhongDa Hospital and conforms to the provisions of the Declaration of Helsinki (as

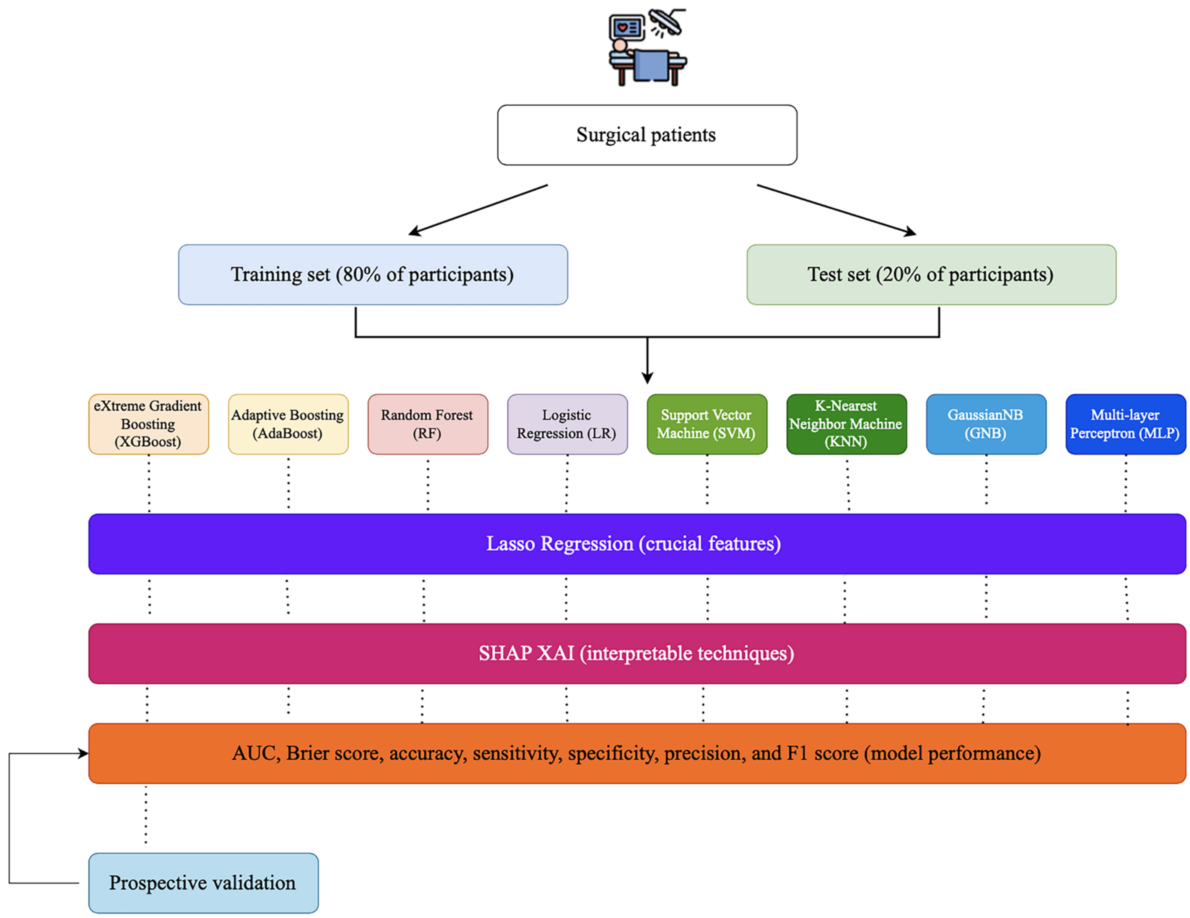

Figure 1. Flow chart of the study design. The figure shows the relevant data collected from patients undergoing surgery for degenerative lumbar fusion in hospitals, including demographic characteristics, radiological measurement parameters, and surgical information. A total of 32 variables were collected, out of which 10 non-zero features were selected through LASSO regression for building machine learning models. Subsequently, the model's performance was evaluated to determine the optimal predictive model. The data of 45 patients were prospectively collected for further validation. Finally, SHAP interpretability analysis was conducted based on the best predictive model.

revised in 2013). Informed consent was waived for this retrospective study. The workflow of our study design and its corresponding analyses are depicted in Figure 1. We retrospectively analyzed the clinical data and radiographic data of 552 patients who underwent primary transpedicular lumbar fixation for LDD at the Spine Surgery Center of Southeast University ZhongDa Hospital from January 2018 to December 2021. The inclusion criteria were as follows: *i*) patient's age over 50 years; *ii*) patients who underwent primary pedicle screw fixation for LDD, including lumbar disc herniation, degenerative lumbar spondylolisthesis, degenerative lumbar spinal stenosis, and degenerative lumbar scoliosis; *iii*) the number of fused levels ≤4 segments; *iv*) patients who underwent lumbar X-ray, CT, and DXA within 1 month prior to surgery at our institution; and *v*) patients were followed up for 3-12 months after surgery, and the follow-up data were complete. The exclusion criteria were as follows: *i*) patients with congenital spinal deformities, spinal trauma, spinal tumors, spinal tuberculosis, spine infection, ankylosing spondylitis, or a history of previous spinal surgery; *ii*) the presence of metabolic bone disease or long-term use of drugs such as corticosteroids that affect bone density; and *iii*) patients with screw loosening due to surgical site infection.

To further validate the model's accuracy, we prospectively collected data from patients who underwent primary transpedicular lumbar fixation for LDD at the Spine Surgery Center of Southeast University ZhongDa Hospital from January 2022 to April 2022.

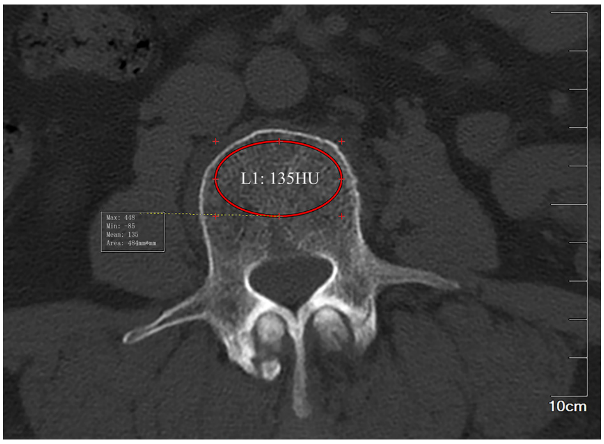

Figure 2. The measurement of HU value: the HU value of L1 was 135.

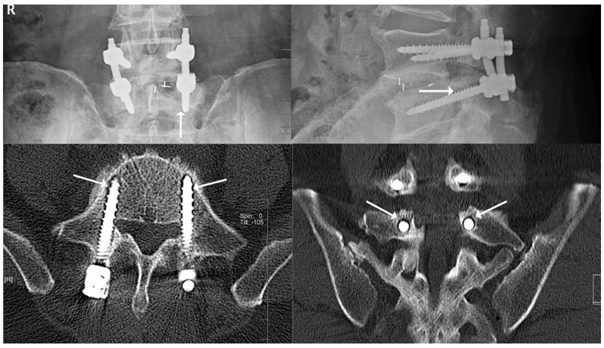

Figure 3. Postoperative follow-up radiographs and CT scans shows screw loosening.

2.2. Evaluation of BMD and screw loosening

All patients received DXA scanning and lumbar CT with three-dimensional reconstruction examination at our radiology center one month prior to the surgery. The tube voltage for the CT scan was 120 kV. The HU values for L1 – L4 were independently measured for each patient by two authors (FJ and XXL), adhering to the methodology outlined in prior studies (*16,28*). This involved placing an elliptical region of interest (ROI) on the central cross-sectional CT image of the vertebral body, with the inclusion of trabecular bone within the ROI and the exclusion of cortical bone, osteophytes, bone endplates, and the posterior venous plexus (Figure 2). Subsequently, the HU value was automatically calculated by the picture archiving and communication system (PACS). The mean HU values of L1 to L4 represented the lumbar BMD. In addition, DXA scans were conducted at the lumbar vertebrae (L1 – L4), as well as the total hips and femoral necks, and the lowest lumbar BMD and the lowest T-score were documented for subsequent analysis.

Patients were followed up with a lumbar X-ray at 3–12 months postoperatively. Lumbar CT was not routinely conducted throughout the follow-up duration; consequently, if abnormalities were detected on the lumbar X-ray, supplementary lumbar CT scans were conducted to confirm the presence of screw loosening (Figure 3). In the current study, screw loosening was defined as the presence of a radiolucent zone with a minimum width of 1 mm around the pedicle screw on radiographs taken during the 3–12 month follow-up period (*29,30*). Patients were categorized into two groups based on the presence or absence of screw loosening at the 12-month follow-up examination: the loosening group and the non-loosening group.

In order to assess reliability, a random selection of 30 patients was made to evaluate the measurement of HU values and the judgment of screw loosening. Two authors independently measured the HU values of LI – L4 for all patients and judged screw loosening for each patient. Two weeks later, the HU values of these 30 patients were measured again and screw loosening was reevaluated. Throughout the process of HU value measurement and screw loosening assessment, two authors were kept blinded to both the DXA results of the patients and the measurements recorded by the other author.

2.3. Lumbar X-ray assessment

Patients underwent lumbar X-ray examination one month before surgery and prior to discharge. The lumbar lordosis (LL), pelvic incidence (PI), pelvic tilt (PT), sacral slope (SS), and the difference between pelvic incidence and lumbar lordosis (PI-LL) were measured and recorded.

2.4. Model input features and model development

We collected 32 potential characteristics, including basic patient characteristics: age, gender, height, weight, BMI, hypertension, diabetes, history of smoking, and history of alcoholism; surgery-related information: duration of surgery, intraoperative blood loss, number of fused levels, fixation to S1, hospitalization time; and preoperative and postoperative radiographic parameters. To identify the crucial factors attributed to screw loosening, the least absolute shrinkage and selection operator (LASSO) technique was employed for feature selection (*31,32*).

In order to maximize predictive performance, we developed eight machine learning models: the eXtreme Gradient Boosting (XGBoost) algorithm, Adaptive Boosting (AdaBoost), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbor Machine (KNN), GaussianNB (GNB), and Multi-layer Perceptron (MLP).

2.5. Sample size and statistical analysis

For the binary prediction model, the sample size calculation formula is obtained according to the previous study (*33*), which is:

$$N = \exp\left(\frac{-0.508+0.259\ln(\varphi)+0.504\ln(P)-\ln(MAPE)}{0.544}\right)$$

Here, φ denotes the ratio of positive events, P denotes the number of model input features, and *MAPE* denotes the mean absolute percentage error between the observed and actual outcome probability. Based on the above formula, the minimum sample size was estimated to be 406. Thus, we performed a random partition of the complete dataset (n = 552) into a training set (n = 442) and a test set (n = 110) using an 8:2 ratio.

In this study, all analyses were performed using Python version 3.9.0 (*34*). Interobserver and intraobserver reliability of the HU values were assessed using the Intraclass correlation coefficient (ICC). Excellent reliability was defined as ICC $\geq$ 0.8. The agreement in determining screw loosening was evaluated using a kappa statistics test. The Shapiro-Wilk test was used to test the normality of the distribution of continuous variables. Continuous variables that conformed to a normal distribution were reported as mean ± standard deviation (SD) and compared using independent-samples *t*-test. Continuous variables that were not normally distributed were expressed as the median and interquartile range (IQR) and analyzed using the Mann-Whitney *U* test. Categorical variables were described as frequencies and percentages, and compared using chi-square tests or Fisher's exact probability tests. Lastly, crucial features were selected through LASSO regression analysis, and based on these features, eight models were developed.

For the selection of model hyperparameters, ten-fold cross-validation was performed on the training datasets. The approach for handling missing data was as follows: missing values were imputed using the random forest regression method if the percentage of missing values was less than 20%; otherwise, the missing cases were excluded from the analysis. The predictive performance of the model was assessed through discrimination and calibration. Discrimination was quantified using the AUROC and Brier score, and model performance was assessed by accuracy, sensitivity, specificity, precision, and F1 score. The Brier score, representing the average squared difference between predicted probabilities and true labels, served as an indicator of model performance, with lower scores indicating higher accuracy. Following the identification of the optimal model, the Python-based SHAP package was utilized to illustrate the significance of individual features (*35*). At last, the selected model was employed to visualize prospective validations.

3. Results

3.1. Patient characteristics and pedicle screw loosening rates

A total of 552 patients were included in this study. Patients were divided into the loosening group (n = 128) and the non-loosening group (n = 424) based on the presence or absence of screw loosening within 12 months of postoperative follow-up. Table 1 shows the demographic characteristics and surgical information of the study participants who underwent transpedicular lumbar fixation surgery for LDD. The radiographic data for the loosening group and the non-loosening group are shown in Table 2. The incidence rate of pedicle screw loosening was approximately 23.19%. The reliability of interobserver and intraobserver measurements of HU value was deemed excellent, as indicated by ICC values of 0.88 and 0.86, respectively. The determination of screw loosening demonstrated high agreement, with kappa values of 0.79 and 0.76, respectively. There were few missing values for the study variables. No statistically significant difference was found in the patient characteristics between the training and test datasets.

3.2. Crucial features

The optimal parameter (lambda) for the LASSO model selection was determined using ten-fold cross-validation. With the optimal lambda, ten features demonstrated non-zero coefficients (Figure 4), encompassing age, disease diagnosis: degenerative scoliosis, number of fused levels, fixation to S1, HU value, preoperative PT,

Table 1. Demographic characteristics and clinical information of the study participants who underwent transpedicular lumbar fixation surgery

Variables	All (n = 552)	Loosening group (n = 128)	Non-loosening group (n = 424)	p-Value
Age, median (Q1, Q3)	61 (55, 70)	69 (64, 74)	58 (54, 66)	< 0.001
Sex, n %				0.105
Female	315 (57.065)	81 (63.281)	234 (55.189)	
Male	237 (42.935)	47 (36.719)	190 (44.811)	
Height, median (Q1, Q3)	162 (158,170)	161 (158, 170)	163 (158, 170)	0.107
Weight, median (Q1, Q3)	67.5 (60, 75)	65 (60, 72.5)	68 (60, 75)	0.110
BMI, median (Q1, Q3)	24.65 (22.86, 26.74)	24.615 (22.823, 26.725)	24.770 (22.883, 26.823)	0.584
Hypertension, n %	224 (40.580)	59 (46.094)	165 (38.915)	0.147
Diabetes, n %	98 (17.754)	29 (22.656)	69 (16.274)	0.098
Alcohol, n %	94 (17.029)	22 (17.188)	72 (16.981)	0.957
Smoking, n %	35 (6.341)	7 (5.469)	28 (6.604)	0.644
Primary diagnosis				
Lumbar disc herniation, n %	185 (33.514)	36 (28.125)	149 (35.142)	0.141
Lumbar degenerative spondylolisthesis, n %	145 (26.268)	30 (23.438)	115 (27.123)	0.406
Lumbar spinal stenosis, n %	191 (34.601)	47 (36.719)	144 (33.962)	0.566
Degenerative scoliosis, n %	31 (5.616)	15 (11.719)	16 (3.774)	< 0.001
Number of fused levels, n %				< 0.001
1	199 (36.051)	27 (21.094) 13.57	172 (40.566)	
2	188 (34.058)	45 (35.156) 23.94	143 (33.726)	
3	103 (18.659)	28 (21.875) 27.18	75 (17.689)	
4	62 (11.232)	28 (21.875) 45.16	34 (8.019)	
Intraoperative blood loss, median (Q1, Q3)	250.0 (150.0, 400.0)	250.0 (150.0, 462.5)	250.0 (100.0, 350.0)	0.103
Duration of surgery, median (Q1, Q3)	165.0 (133.75, 200.0)	180.0 (148.75, 206.25)	162.5 (130.0, 200.0)	0.052
Fixation to S1, n %	268 (48.551)	79 (61.719)	189 (44.575)	< 0.001
Hospitalization time, median (Q1, Q3)	11 (9, 13)	11.5 (9, 14)	11 (9, 13)	0.128

Table 2. The preoperative and postoperative radiographic data for the loosening group and the non-loosening group

Variables	All (n = 552)	Loosening group (n = 128)	Non-loosening group (n = 424)	p-Value
The lowest lumbar BMD, median (Q1, Q3)	1.045 (0.947, 1.189)	1.030 (1.028, 1.056)	1.045 (0.929, 1.321)	0.044
The lowest T-score, median (Q1, Q3)	-1.9 (-2.7, -0.6)	-2.45 (-3.2, -0.2)	-1.8 (-2.7, -1)	0.081
HU value, median (Q1, Q3)	138.25 (113.5, 164.0)	96.875 (76, 114.563)	148.250 (126, 169.688)	< 0.001
lumbar instability, n %	187 (33.877)	41 (32.031)	146 (34.434)	0.615
Preoperative LL, median (Q1, Q3)	42 (33, 50)	39 (31, 51.25)	42 (35, 49.25)	0.106
Preoperative PI, median (Q1, Q3)	48 (41, 54)	51.7(44, 60.25)	50.6 (44, 63)	0.763
Preoperative PT, median (Q1, Q3)	17 (13, 23)	23 (16, 30)	16 (13, 21)	< 0.001
Preoperative SS, median (Q1, Q3)	30 (25, 36)	30 (23.75, 37)	30 (25, 36)	0.832
Preoperative PI-LL, median (Q1, Q3)	6.0 (3.0, 9.0)	10.5 (3, 22)	5 (2, 8)	< 0.001
Postoperative LL, median (Q1, Q3)	43.0 (36, 50)	46 (38, 53)	42 (36, 49)	< 0.001
Postoperative PI, median (Q1, Q3)	48 (42, 56)	52 (45, 60)	51 (44.75, 58)	0.778
Postoperative PT, median (Q1, Q3)	13 (9, 16)	15 (10, 20)	12 (9, 15)	< 0.001
Postoperative SS, median (Q1, Q3)	36 (30, 41)	31.6 (26.6, 37)	33 (27, 37)	0.384
Postoperative PI-LL, median (Q1, Q3)	6.0 (2.0, 9.0)	10 (2.5, 17)	5 (2, 9)	< 0.001

Abbreviations: BMD, bone mineral density; HU, Hounsfield unit; LL, lumbar lordosis; PI, pelvic incidence; PT, pelvic tilt; SS, sacral slope.

preoperative PI-LL, postoperative LL, postoperative PT, and postoperative PI-LL.

3.3. Model performance

Eight machine learning algorithms were used to construct prediction models for screw loosening, and the predictive performance of each model was evaluated by calculating Brier scores and AUROC. In comparison to the other models, XGBoost demonstrated the lowest Brier score, as illustrated in Figure 5A, which presented the calibration plots for all eight models. Furthermore, the XGBoost model outperformed the others with a higher AUROC, as shown in Figure 5B. Based on the AUROC values of the eight models, a forest plot illustrating the AUC scores for multiple models was generated (Figure 5C). Through ten-fold cross-validation, the XGBoost model achieved a smaller standard deviation of 0.036 for

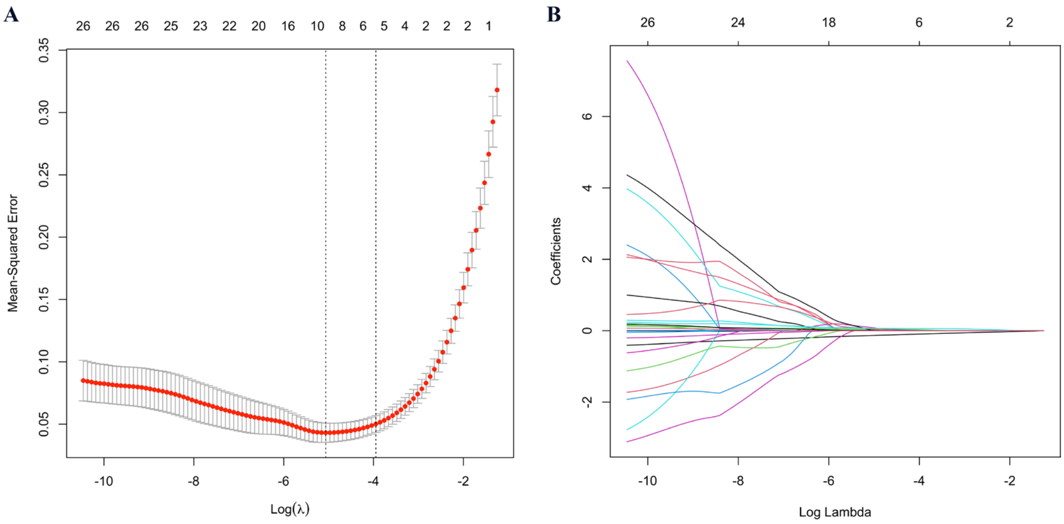

Figure 4. Clinical and radiographic feature selection using the LASSO regression. (A) LASSO coefficient profiles of 32 features. **(B)** Feature selection for the predictive model. Turning parameter (λ) selection used tenfold cross-validation. The vertical axis shows the model misclassification rate, and the horizontal axis shows the log(λ). The two vertical dashed lines represent the minimum value and one standard deviation on one side from the minimum value.

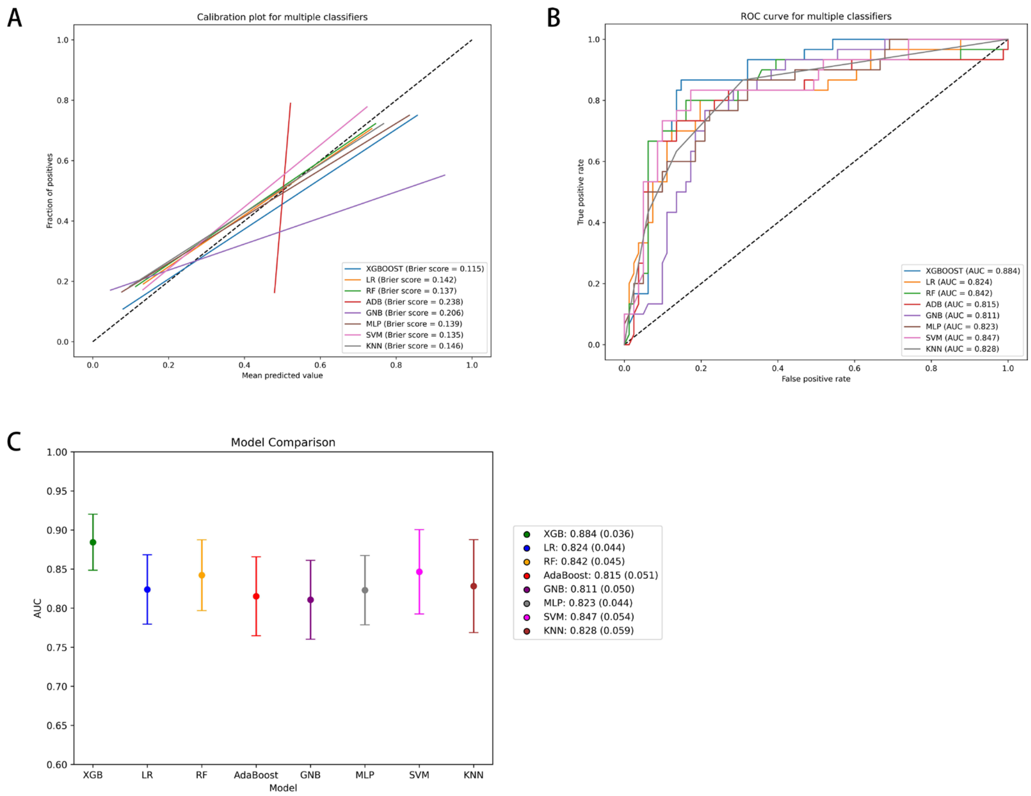

Figure 5. Model performance. (A) Calibration plots of the eight models. **(B)** Receiver-operating characteristic curves for the eight models. **(C)** Forest plot of the AUC score and 95 CI% of the eight models.

its AUC score. This outcome suggested that the XGBoost model exhibited the most stable performance compared to the other seven models. The performance metrics of the eight models in the test dataset are presented in Table 3.

In Figure 6, the XGBoost model was analyzed using the SHAP method. This figure provided a clear understanding of the contribution of each feature to the model output. Additionally, the bar chart illustrated the magnitude of the impact that the feature importance had on the model predictions.

3.4. Application of the model

The SHAP waterfall and force plots for the XGBoost model are shown in Figure 7. Inputting the clinical information of a typical patient into the model, for example, in Figure 7A, the true outcome of the

Table 3. Performance metrics for eight models in the test dataset

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score
XGBoost	0.847	0.600	0.938	0.783	0.679
LR	0.793	0.767	0.802	0.706	0.667
RF	0.829	0.800	0.840	0.714	0.716
AdaBoost	0.782	0.732	0.801	0.729	0.612
GNB	0.775	0.633	0.827	0.576	0.638
MLP	0.802	0.600	0.877	0.643	0.621
SVM	0.829	0.733	0.864	0.667	0.698
KNN	0.802	0.633	0.864	0.633	0.633

Abbreviations: XGBoost, eXtreme Gradient Boosting; LR, Logistic Regression; RF, Random Forest; AdaBoost, Adaptive Boosting; GNB, Gaussian Naive Bayes; MLP, Multi-layer Perceptron; SVM, Support Vector Machine; KNN, K-Nearest Neighbor Machine.

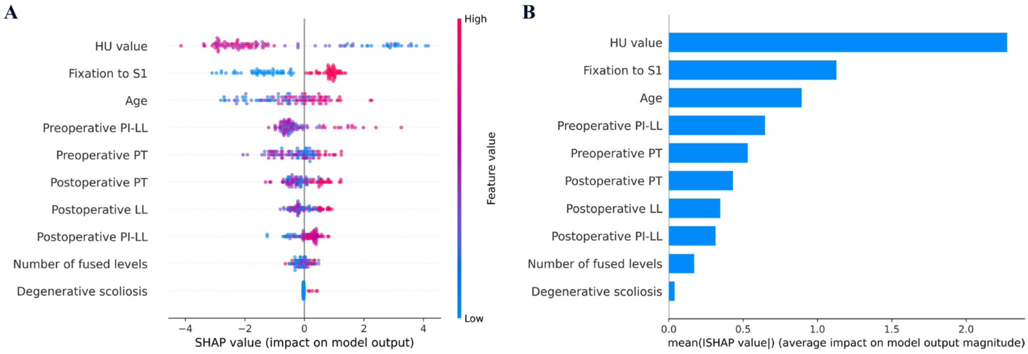

Figure 6. SHAP analysis of the XGBoost model. (A) SHAP summary plot of clinical features, with red indicating higher values and blue indicating lower values. **(B)** Importance matrix plot of the XGBoost model, indicating the importance of each variable in predicting postoperative screw loosening.

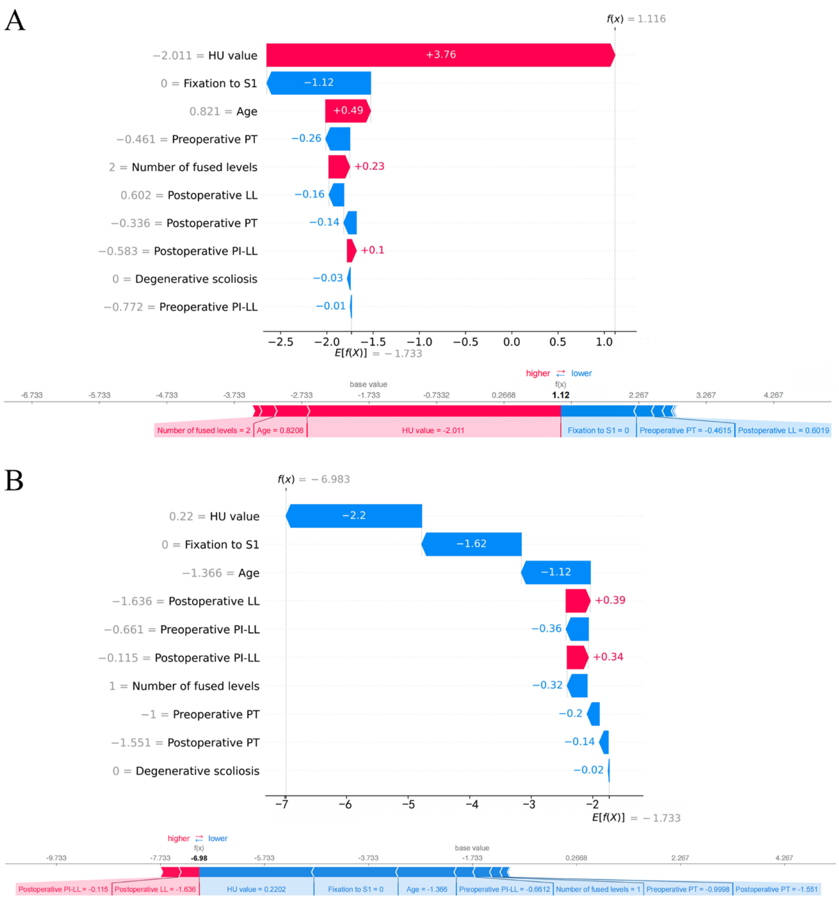

Figure 7. SHAP force plot for patients in the dataset at (A) high or (B) low risk of postoperative screw loosening.

patient was screw loosening, and the predicted value of the model was 1.12, with a predicted probability of 75.40% for screw loosening; in Figure 7B, the true outcome of the patient was no screw loosening, and the predicted value of the model was -6.98, with a predicted probability of 0.09% for screw loosening.

3.5. Prospective validation

A total of 45 patients were enrolled in this study for prospective validation, with 20.0% (9/45) of them experiencing screw loosening. The proposed model achieved an accuracy of 83.32% when tested on the prospective dataset, with a respective sensitivity of 0.543 and specificity of 0.940.

4. Discussion

Few models are available to predict screw loosening after degenerative lumbar fusion surgery in patients with LDD. In this study, we applied machine learning methods to identify risk factors for pedicle screw loosening after degenerative lumbar fusion surgery and to develop risk prediction models for screw loosening. The performance of eight machine learning models was compared. The results showed that the XGBoost model had the highest AUC (88.4%). The calibration of the models was quantitatively compared using Brier scores. The calibration of the XGBoost model showed good agreement between the prediction outcome and the actual observed outcome. The standard deviation of the AUC score of the XGBoost model obtained after using ten cross-validations was 0.036, which was smaller than the other seven models, suggesting that the XGBoost model has the most stable performance. Based on the above aspects, it can be concluded that the XGBoost model exhibited superior performance compared to seven other machine learning models. The SHAP method further explained the predictors and model prediction performance. It provided a simple and robust method for individualized prediction of pedicle screw loosening after degenerative lumbar fusion surgery, which can provide important information for medical decision support.

The rate of screw loosening in the current study was observed to be 23.19%. In a study conducted by Tokuhashi *et al*. (*36*), they reported a screw loosening rate of 26.8% at the 12-month follow-up in patients with LDD who underwent pedicle screw fixation in the lumbar spine. Additionally, Shin *et al*. (*19*) found a screw loosening rate of 22.5%. The occurrence of screw loosening is caused by a variety of factors (*3,7,19,37*). With a total of ten variables included in the XGBoost model analysis, we found that potential risk factors for screw loosening were associated with low BMD, older age, fixation to S1, multi-segment fixation, and sagittal imbalance.

Osteoporosis is the most commonly discussed cause of screw loosening. This is because in patients with osteoporosis, the screw-bone interface has a lower ability to bind the screw, leading to reduced screw pullout strength. A biomechanical study demonstrated that decreased bone density resulted in a decline in screw pullout force, ultimately leading to the occurrence of screw loosening (*38*). Osteoporosis is typically diagnosed using the standard technique of DXA. Previous studies have shown a difference in DXA between patients undergoing lumbar fusion surgery with and without screw loosening (*20,39*), but Kim *et al*. reported no difference in DXA between the two groups (*37*). These contradictory findings can be attributed to the inaccuracies of DXA in evaluating BMD. Degenerative changes in the lumbar spine may lead to overestimation of BMD, particularly in patients with severe lumbar degeneration (*28*). Hence, in this study, we used lumbar CT to measure the HU value of the vertebral body and recorded the T-score and lumbar BMD results of DXA. The results of the study suggested that the screw loosening rate was higher in patients who possessed a low HU value than in those who had a high HU value, but that the T-score and lumbar BMD value of the DXA performed poorly in recognizing screw loosening. Furthermore, since BMD decreases significantly with age, our findings also suggested that screw loosening was more common in older patients. Therefore, we should not concentrate only on the DXA results when making a surgical strategy for lumbar fixation in elderly patients with LDD. We recommend routinely measuring the HU value for surgical planning in LDD patients.

Some studies have illustrated the significance of S1 fixation in the occurrence of screw loosening. Serving as a critical connection between the spine and pelvis, S1 exhibits a greater susceptibility to loosening due to its longer lever arm and larger physiological arc under fixed stress (*20,40*). The anatomical attributes of the S1 pedicle, characterized by a larger diameter and shorter length compared to lumbar pedicles, and the presence of predominantly cancellous bone within the sacrum, collectively point towards a heightened susceptibility of S1 screw loosening (*41*). These anatomical factors likely contribute to the increased incidence of screw loosening in the S1 region.

Multiple-segment screw fixation have consistently identified as a notable risk factor associated with screw loosening (*20,36,42*). According to the study by Zou *et al*. (*43*), the rate of screw loosening in single-level procedures was found to be 4.1%, while it increased to 33.3%, 53.3%, and 78.8% in two-level, three-level, and four-level procedures, respectively. In our own investigation, we observed a similar trend; specifically, an escalating rate of screw loosening was observed with an increasing number of screw fixation levels. This rise in the incidence of screw loosening with the more segments for screw fixation can be attributed to the amplified cantilever bending moment exerted on the surgical construct (*3,7,20,42*). Notably, screw loosening frequently occurred at the distal end of the screw instrumentation in patients undergoing multi-level fixation (*44*).

The presence of sagittal imbalance can contribute to an elevated risk of screw loosening. In our study, we identified postoperative LL as a predictive factor for screw loosening. Livshits *et al*. (*45*) demonstrated that restoring postoperative LL was associated with a reduced incidence of screw loosening. Kuo *et al*. (*46*) found that, even among patients undergoing dynamic stabilization, the loss of LL postoperatively increased the rate of screw loosening. In this study, the loosening group also exhibited higher preoperative and postoperative PT compared to the non-loosening group. The posterior tilting of the pelvis (increased PT) could be a compensatory response to sagittal malalignment (*47*). Furthermore, studies have shown that PI-LL mismatch is an important indicator of sagittal balance and is associated with adjacent segment degeneration, screw loosening, and disability and quality of life scores (*48,49*). In our study, we revealed that preoperative and postoperative PI-LL mismatch was a significant predictive feature of screw loosening.

In this study, we used prospective data to validate the predictive performance of the model, but there are still some limitations of this study. First, given the small sample size of this study, further research with a larger sample is needed to validate the predictive model. Second, the data in this study was collected from a single large academic medical center. Consequently, the generalizability of this model to other medical institutions may be limited. It is highly probable that recalibration of the model would be essential when implementing it in another institution, as the relative weights of the features may necessitate adjustments. Last, an independent dataset is indispensable to assess the model's extrapolation and generalization. To address this need, our future research endeavors will focus on acquiring an ample number of external validation datasets to further refine and enhance the performance of this model.

5. Conclusion

In this study, we developed eight different prediction models for postoperative screw loosening, among which the XGBoost model demonstrated good discrimination and overall performance. In addition, based on interpretable techniques, this model enables individualized prediction of postoperative screw loosening. We believe that this model is an important tool for identifying the postoperative occurrence of pedicle screw loosening in patients requiring degenerative lumbar fusion surgery.

Funding: None.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Boos N, Webb JK. Pedicle screw fixation in spinal disorders: a European view. Eur Spine J. 1997; 6:2-18.
2. Hoppe S, Keel MJ. Pedicle screw augmentation in osteoporotic spine: indications, limitations and technical aspects. Eur J Trauma Emerg Surg. 2017; 43:3-8.
3. Galbusera F, Volkheimer D, Reitmaier S, Berger-Roscher N, Kienle A, Wilke HJ. Pedicle screw loosening: a clinically relevant complication? Eur Spine J. 2015; 24:1005-1016.
4. Alanay A, Vyas R, Shamie AN, Sciocia T, Randolph G, Wang JC. Safety and efficacy of implant removal for patients with recurrent back pain after a failed degenerative lumbar spine surgery. J Spinal Disord Tech. 2007; 20:271-277.
5. Berjano P, Bassani R, Casero G, Sinigaglia A, Cecchinato R, Lamartina C. Failures and revisions in surgery for sagittal imbalance: analysis of factors influencing failure. Eur Spine J. 2013; 22 Suppl 6:S853-858.
6. Bredow J, Boese CK, Werner CM, Siewe J, Löhrer L, Zarghooni K, Eysel P, Scheyerer MJ. Predictive validity of preoperative CT scans and the risk of pedicle screw loosening in spinal surgery. Arch Orthop Trauma Surg. 2016; 136:1063-1067.
7. Röllinghoff M, Schlüter-Brust K, Groos D, Sobottke R, Michael JW, Eysel P, Delank KS. Mid-range outcomes in 64 consecutive cases of multilevel fusion for degenerative diseases of the lumbar spine. Orthop Rev (Pavia). 2010; 2:e3.
8. Reis MT, Reyes PM, Bse, Altun I, Newcomb AG, Singh V, Chang SW, Kelly BP, Crawford NR. Biomechanical evaluation of lateral lumbar interbody fusion with secondary augmentation. J Neurosurg Spine. 2016; 25:720-726.
9. Bokov A, Bulkin A, Aleynik A, Kutlaeva M, Mlyavykh S. Pedicle Screws Loosening in Patients With Degenerative Diseases of the Lumbar Spine: Potential Risk Factors and Relative Contribution. Global Spine J. 2019; 9:55-61.
10. Zou D, Sun Z, Zhou S, Zhong W, Li W. Hounsfield units value is a better predictor of pedicle screw loosening than the T-score of DXA in patients with lumbar degenerative diseases. Eur Spine J. 2020; 29:1105-1111.
11. Mikula AL, Puffer RC, Jeor JDS, Bernatz JT, Fogelson JL, Larson AN, Nassr A, Sebastian AS, Freedman BA, Currier BL, Bydon M, Yaszemski MJ, Anderson PA, Elder BD. Teriparatide treatment increases Hounsfield units in the lumbar spine out of proportion to DEXA changes. J Neurosurg Spine. 2019; 1-6.
12. Xie Y, Fu Q, Chen ZQ, Shi ZC, Zhu XD, Wang CF, Li M. Comparison between two pedicle screw augmentation instrumentations in adult degenerative scoliosis with osteoporosis. BMC Musculoskelet Disord. 2011; 12:286.
13. Piñera AR, Duran C, Lopez B, Saez I, Correia E, Alvarez L. Instrumented lumbar arthrodesis in elderly patients: prospective study using cannulated cemented pedicle screw instrumentation. Eur Spine J. 2011; 20 Suppl 3:408-414.
14. Dai F, Liu Y, Zhang F, Sun D, Luo F, Zhang Z, Xu J.

Surgical treatment of the osteoporotic spine with bone cement-injectable cannulated pedicle screw fixation: technical description and preliminary application in 43 patients. Clinics. 2015; 70:114-119.
15. Blake GM, Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. Postgrad Med J. 2007; 83:509-517.
16. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. Ann Intern Med. 2013; 158:588-595.
17. Gausden EB, Nwachukwu BU, Schreiber JJ, Lorich DG, Lane JM. Opportunistic Use of CT Imaging for Osteoporosis Screening and Bone Density Assessment: A Qualitative Systematic Review. J Bone Joint Surg Am. 2017; 99:1580-1590.
18. Pompe E, de Jong PA, de Jong WU, Takx RA, Eikendal AL, Willemink MJ, Oudkerk M, Budde RP, Lammers JW, Mohamed Hoesein FA. Inter-observer and inter-examination variability of manual vertebral bone attenuation measurements on computed tomography. Eur Radiol. 2016; 26:3046-3053.
19. Shin HK, Koo HW, Kim KH, Yoon SW, Sohn MJ, Lee BJ. The Usefulness of Trabecular CT Attenuation Measurement at L4 Level to Predict Screw Loosening After Degenerative Lumbar Fusion Surgery: Consider Number of Fused Levels and Postoperative Sagittal Balance. Spine. 2022; 47:745-753.
20. Kim JB, Park SW, Lee YS, Nam TK, Park YS, Kim YB. The Effects of Spinopelvic Parameters and Paraspinal Muscle Degeneration on S1 Screw Loosening. J Korean Neurosurg Soc. 2015; 58:357-362.
21. Sakai Y, Takenaka S, Matsuo Y, Fujiwara H, Honda H, Makino T, Kaito T. Hounsfield unit of screw trajectory as a predictor of pedicle screw loosening after single level lumbar interbody fusion. J Orthop Sci. 2018; 23:734-738.
22. Erickson BJ. Basic Artificial Intelligence Techniques: Machine Learning and Deep Learning. Radiol Clin North Am. 2021; 59:933-940.
23. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med. 2018; 1:5.
24. Rudisill SS, Hornung AL, Barajas JN, *et al*. Artificial intelligence in predicting early-onset adjacent segment degeneration following anterior cervical discectomy and fusion. Eur Spine J. 2022; 31:2104-2114.
25. Chang M, Canseco JA, Nicholson KJ, Patel N, Vaccaro AR. The Role of Machine Learning in Spine Surgery: The Future Is Now. Front Surg. 2020; 7:54.
26. Karhade AV, Thio Q, Ogink PT, Bono CM, Ferrone ML, Oh KS, Saylor PJ, Schoenfeld AJ, Shin JH, Harris MB, Schwab JH. Predicting 90-Day and 1-Year Mortality in Spinal Metastatic Disease: Development and Internal Validation. Neurosurgery. 2019; 85:E671-e681.
27. Joshi RS, Haddad AF, Lau D, Ames CP. Artificial Intelligence for Adult Spinal Deformity. Neurospine. 2019; 16:686-694.
28. Zou D, Li W, Deng C, Du G, Xu N. The use of CT Hounsfield unit values to identify the undiagnosed spinal osteoporosis in patients with lumbar degenerative diseases. Eur Spine J. 2019; 28:1758-1766.
29. Sandén B, Olerud C, Petrén-Mallmin M, Johansson C, Larsson S. The significance of radiolucent zones surrounding pedicle screws. Definition of screw loosening in spinal instrumentation. J Bone Joint Surg Br. 2004; 86:457-461.
30. Spirig JM, Sutter R, Götschi T, Farshad-Amacker NA, Farshad M. Value of standard radiographs, computed tomography, and magnetic resonance imaging of the lumbar spine in detection of intraoperatively confirmed pedicle screw loosening-a prospective clinical trial. Spine J. 2019; 19:461-468.
31. Tibshirani R. Regression shrinkage and selection *via* the lasso. Journal of the Royal Statistical Society, Series B. 1996; 58.
32. Okada A, Hashimoto Y, Goto T, Yamaguchi S, Ono S, Ikeda Kurakawa K, Nangaku M, Yamauchi T, Yasunaga H, Kadowaki T. A Machine Learning-Based Predictive Model to Identify Patients Who Failed to Attend a Follow-up Visit for Diabetes Care After Recommendations From a National Screening Program. Diabetes Care. 2022; 45:1346-1354.
33. Riley RD, Ensor J, Snell KIE, Harrell FE, Jr., Martin GP, Reitsma JB, Moons KGM, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. Bmj. 2020; 368:m441.
34. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVylder J, Walter M, Berrouiguet S, Lemey C. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. J Med Internet Res. 2021; 23:e15708.
35. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020; 2:56-67.
36. Tokuhashi Y, Matsuzaki H, Oda H, Uei H. Clinical course and significance of the clear zone around the pedicle screws in the lumbar degenerative disease. Spine. 2008; 33:903-908.
37. Kim HJ, Kim SG, Lee HM, Kim HS, Moon ES, Park JO, Seol NH, Moon SH. Risk factors associated with the halo phenomenon after lumbar fusion surgery and its clinical significance. Asian Spine J. 2008; 2:22-26.
38. Zindrick MR, Wiltse LL, Widell EH, Thomas JC, Holland WR, Field BT, Spencer CW. A biomechanical study of intrapeduncular screw fixation in the lumbosacral spine. Clin Orthop Relat Res. 1986; 99-112.
39. Okuyama K, Abe E, Suzuki T, Tamura Y, Chiba M, Sato K. Influence of bone mineral density on pedicle screw fixation: a study of pedicle screw fixation augmenting posterior lumbar interbody fusion in elderly patients. Spine J. 2001; 1:402-407.
40. Luk KD, Chen L, Lu WW. A stronger bicortical sacral pedicle screw fixation through the s1 endplate: an *in vitro* cyclic loading and pull-out force evaluation. Spine. 2005; 30:525-529.
41. McLachlin SD, Al Saleh K, Gurr KR, Bailey SI, Bailey CS, Dunning CE. Comparative assessment of sacral screw loosening augmented with PMMA versus a calcium triglyceride bone cement. Spine. 2011; 36:E699-704.
42. Wu X, Shi J, Wu J, Cheng Y, Peng K, Chen J, Jiang H. Pedicle screw loosening: the value of radiological imagings and the identification of risk factors assessed by extraction torque during screw removal surgery. J Orthop Surg Res. 2019; 14:6.
43. Zou D, Muheremu A, Sun Z, Zhong W, Jiang S, Li W. Computed tomography Hounsfield unit-based prediction of pedicle screw loosening after surgery for degenerative lumbar spine disease. J Neurosurg Spine. 2020; 32:716-721.
44. Ohba T, Ebata S, Oba H, Koyama K, Haro H. Risk Factors

for Clinically Relevant Loosening of Percutaneous Pedicle Screws. Spine Surg Relat Res. 2019; 3:79-85.
45. Livshits G, Ermakov S, Popham M, Macgregor AJ, Sambrook PN, Spector TD, Williams FM. Evidence that bone mineral density plays a role in degenerative disc disease: the UK Twin Spine study. Ann Rheum Dis. 2010; 69:2102-2106.
46. Kuo CH, Chang PY, Tu TH, Fay LY, Chang HK, Wu JC, Huang WC, Cheng H. The Effect of Lumbar Lordosis on Screw Loosening in Dynesys Dynamic Stabilization: Four-Year Follow-Up with Computed Tomography. Biomed Res Int. 2015; 2015:152435.
47. Ochtman AEA, Kruyt MC, Jacobs WCH, Kersten R, le Huec JC, Öner FC, van Gaalen SM. Surgical Restoration of Sagittal Alignment of the Spine: Correlation with Improved Patient-Reported Outcomes: A Systematic Review and Meta-Analysis. JBJS Rev. 2020; 8:e1900100.
48. Banno T, Hasegawa T, Yamato Y, Kobayashi S, Togawa D, Oe S, Mihara Y, Matsuyama Y. Prevalence and Risk Factors of Iliac Screw Loosening After Adult Spinal Deformity Surgery. Spine. 2017; 42:E1024-e1030.
49. Schwab FJ, Blondel B, Bess S, *et al*. Radiographical spinopelvic parameters and disability in the setting of adult spinal deformity: a prospective multicenter analysis. Spine. 2013; 38:E803-812.

Received December 28, 2023; Revised February 14, 2024; Accepted February 21, 2024.

*Address correspondence to:
Yuntao Wang, Department of Spine Surgery, Southeast University ZhongDa Hospital, Nanjing 210009, Jiangsu, China.
E-mail:220183560@seu.edu.cn; wangyttod@seu.edu.cn

Released online in J-STAGE as advance publication February 27, 2024.